

CACIC 2018

congreso argentino de ciencias de la computación

Tandil - Buenos Aires

LIBRO DE ACTAS

8 AL 12 DE OCTUBRE DE 2018

RedUNCI

FACULTAD DE CIENCIAS EXACTAS

UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES



XXIV CONGRESO ARGENTINO DE CIENCIAS DE LA COMPUTACION (CACIC 2018)

Tandil (Buenos Aires), 8 al 12 de octubre 2018

Organizadores

Red de Universidades con Carreras en Informática – RedUNCI.

Facultad de Ciencias Exactas - Universidad Nacional del Centro de la Provincia de Buenos Aires

Libro de Actas XXIV Congreso Argentino de Ciencias de la Computación CACIC 2018: .

Pesado, Patricia; Aciti, Claudio; . - 1a ed.

Tandil: Universidad Nacional del Centro de la Prov. de Buenos Aires. 2019.

Libro digital, PDF

Archivo Digital: descarga y online

ISBN 978-950-658-472-6

1. Informática. 2. Computación. 3. Ciencias Exactas. I. Aciti, Claudio, ed. II. Pesado, Patricia, ed.

CDD 004



Comité Científico

Abásolo, María José (Argentina)
Aciti, Claudio (Argentina)
Alfonso, Hugo (Argentina)
Ardenghi, Jorge (Argentina)
Arroyo, Marcelo (Argentina)
Astudillo Hernán (Chile)
Baldasarri, Sandra (España)
Balladini, Javier (Argentina)
Barbosa, Luis (Portugal)
Bertone, Rodolfo (Argentina)
Bría, Oscar (Argentina)
Brisaboa, Nieves (España)
Buckle, Carlos (Argentina)
Cañas, Alberto (EE.UU)
Casali, Ana (Argentina)
Castro, Silvia (Argentina)
Cechich, Alejandra (Argentina)
Chavez, Edgar (México)
Coello Coello, Carlos (México)
Cuckierman, Uriel (Argentina)
De Giusti, Armando (Argentina)
De Giusti, Laura (Argentina)
De Vincenzi, Marcelo (Argentina)
Deco, Claudia (Argentina)
Depetris, Beatriz (Argentina)
Diaz, Javier (Argentina)
Dix, Juerguen (Alemania)
Doallo, Ramón (España)
Docampo, Domingo (España)
Dujmovic Jozo (USA)
Estayno, Marcelo (Argentina)
Estevez, Elsa (Argentina)
Falappa, Marcelo (Argentina)
Fillottrani, Pablo (Argentina)
Finocchieto, Jorge (Argentina)
Frati, Emmanuel (Argentina)
Fridlender, Daniel (Argentina)
García Garino, Carlos (Argentina)
García Villalba, Javier (España)
Género, Marcela (España)
Gomez, Sergio (Argentina)
Gröller, Eduard (Austria)
Guerrero, Roberto (Argentina)
Ierache, Jorge (Argentina)
Janowski, Tomasz (Naciones Unidas)
Kuna Horacio (Argentina)
Lanzarini, Laura (Argentina)
Leguizamón, Guillermo (Argentina)
Lopez Gil, Fernando (España)
Loui, Ronald Prescott (EEUU)
Luque, Emilio (España)
Madoz, Cristina (Argentina)
Malberti, Alejandra (Argentina)
Malbran, Maria (Argentina)
Manresa Yee, Cristina (España)
Marco Javier (España)
Marín, Mauricio (Chile)
Mas Sansó, Ramón (España)
Micolini, Orlando (Argentina)
Mon, Alicia (Argentina)
Motz, Regina (Uruguay)
Naiouf, Marcelo (Argentina)
Navarro Martín, Antonio (España)
Olivas Varela, José Angel (España)
Pasini, Ariel (Argentina)
Pesado, Patricia (Argentina)



Piccoli, María Fabiana (Argentina)
Printista, Marcela (Argentina)
Pardo, Álvaro (Uruguay)
Piattini, Mario (España)
Puppo, Enrico (Italia)
Ramón, Hugo (Argentina)
Rexachs, Dolores (España)
Reyes, Nora (Argentina)
Roig Vila, Rosabel (España)
Rossi, Gustavo (Argentina)
Rosso, Paolo (España)
Rueda, Sonia (Argentina)
Ruiz Francisco (España)
Russo, Claudia (Argentina)
Salto, Carolina (Argentina)
Sanz, Cecilia (Argentina)
Simari, Guillermo (Argentina)
Sposito, Osvaldo (Argentina)
Steinmetz, Ralf (Alemania)
Suppi, Remo (España)
Tarouco, Liane (Brasil)
Tirado, Francisco (España)
Velho, Luiz (Brasil)
Vénere, Marcelo (Argentina)
Vendrell, Eduardo (España)
Villagarcía, Horacio (Argentina)
Zanarini, Dante (Argentina)



Comité Académico

UBA Cs. Exactas - Garbervetsky, Diego
UBA Ingeniería - Echeverria, Adriana
UN La Plata - Pesado, Patricia
UN Sur - Rueda, Sonia
UN San Luis - Piccoli, Fabiana
UNCPBA - Aciti, Claudio
UN Comahue - Vaucheret, Claudio
UN La Matanza - Spositto, Osvaldo
UN La Pampa - Alfonso, Hugo
UN Lomas de Zamora - Estayno, Marcelo
UNTierra del Fuego - Feierherd, Guillermo
UN Salta - Gil, Gustavo
UN Patagonia Austral - Lasso, Marta
UN SanJuan - Rodriguez, Nelson
UADER - Vivas, Patricia
UN Patagonia SJB - Buckle, Carlos
UN Entre Ríos - Tugnarelli, Mónica
UN Nordeste - Dapozo, Gladys
UN Rosario - Kantor Raul
UN Misiones - Kuna, Horacio
UNNOBA - Russo, Claudia
UN Chilecito - Carmona, Fernanda
UN Lanús - Rodriguez, Darío
UN Santiago del Estero - Duran, Elena
Esc. Sup. Ejército Arroyo - Arzubi, Alejandro
UN Litoral - Loyarte, Horacio
UN RioIV - Arroyo, Marcelo
UN Córdoba - Fridlender, Daniel
UN Jujuy Herrera - Cognetta, Analía
UN Rio Negro - Vivas, Luis
UN Villa María - Prato, Laura
UN Lujan - Panessi, Wálter
UN Catamarca - Poliche Maria Valeria
UN La Rioja - Campazzo Eduardo
UNTres de Febrero - Oliveros, Alejandro
UN Tucumán - Luccioni, Griselda María
UNAJ - Morales, Martín
UN Chaco Austral - Zachman Patricia
UN del Oeste - Foti, Antonio
UN de Cuyo - Forradelas, Raymundo
UN de Mar del Plata - Doumecq, Julio Cesar
UN de Quilmes - a definir
U Morón - Padovani Hugo
UAI - De Vincenzi, Marcelo
U Belgrano - Guerci, Alberto
U kennedy - Panizzi, Marisa
U Adventista del Plata - Bournissen Juan
UCAECE - Finocchieto, Jorge
U Palermo - Alvarez Adriana
UCA Rosario - Grieco, Sebastián
U Salvador - Zanitti, Marcelo
U Aconcagua - Giménez, Rosa
U GastónDachary - Beyersdorf, Carlos
UADE - a definir
UCEMA - Guglianone, Ariadna
U Austral - Cassol, Ignacio
U Atlántida Argentina - Rathmann, Liliana
UCA La Plata - Bertone, Rodolfo
ITBA - Mon, Alicia
U Champagnat - Pinciroli, Fernando
UN Hurlingham - Medrano Gustavo



Comité Organizador

El XXIV Congreso Argentino de Ciencias de la Computación es organizado por la Facultad de Ciencias Exactas de la UNICEN en nombre de la Red de Universidades con Carreras de Informática (RedUNCI).

Responsable

Claudio Aciti

Sub - Responsables

José Marone
Elías Todorovich

Colaboradores

Gustavo Illescas
María Jesús Martín
José Massa
Matias Presso
Gustavo Tripodi
Laura Felice
Claudia Marcos
Silvia Schiaffino
Viviana Ferraggine
Martín Menchón
Virginia Mauco
Carmen Leonardi
Mauricio Poncio
Lucas Picchi
Luis Berdún
Paula Siracusa
Florencia Rodriguez
Juan D' Amato
Diego Dalponte



Autoridades RedUNCI

Coordinadora Titular

Patricia Pesado (UNLP) 2016-2018

Coordinador Alterno

Marcelo Estayno (UNLZ) 2016-2018

Junta Directiva

Fabiana Piccoli (UNSL) 2016-2018
Guillermo Feierherd (UNTDF) 2016-2018
Fernanda Carmona (UNdeC) 2016-2018
Hugo Padovani (U Moron) 2016-2018
Daniel Fridlender (UN Córdoba) 2017-2019
Osvaldo Spositto (UN LaMatanza) 2017-2019
Claudia Russo (UNNOBA) 2017-2019
Alicia Mon (ITBA) 2017-2019

Secretarías

Secretaría Administrativa: Tugnarelli Mónica (UNER)
Secretaría Académica: Lasso Marta (UNPA)
Secretaría de Ciencia y Técnica: Balladini Javier (UNCOMA)
Secretaría de Asuntos Reglamentarios: De Vincenzi Marcelo (UAI)
Secretaría de Vinculación Tecnológica y Profesional: Gil Gustavo (UNSa)
Secretaría de Congresos, Publicaciones y Difusión: Thomas Pablo (UNLP)



Indice

XIX Workshop agentes y sistemas inteligentes (WASI) **1**

Un enfoque de Recomendación multi-dominio basado en Embeddings y redes semánticas 2

SA to optimize the Multi-Period Water Distribution Network Design 12

Filtrado basado en contenido para artículos académicos en repositorios institucionales 22

ToM-Dyna-Q: on the integration of reinforcement learning and machine Theory of Mind 32

Spectogram Prediction with Neural Networks 42

Transferencia de Aprendizaje para Clasificación de Peatones 52

XIX Workshop Procesamiento Distribuido y Paralelo (WPDP) **62**

Factores que afectan el consumo energético de operaciones de checkpoint y restart en clusters 63

Rendimiento del algoritmo AES sobre arquitecturas de memoria compartida 73

CLOUD ROBOTICs: Navegación de un vehículo autónomo en un entorno con obstáculos 83

A Cloud Powered Relaxed Heterogeneous Distributed Shared Memory System 93

Análisis de consumo energético en Cluster de GPU y MultiGPU en un problema de Alta Demanda Computacional 103

Modelado estadístico de potencia usando contadores de rendimiento sobre Raspberry Pi 113

Análisis de las topologías Edge y Fog para la continuidad del extremo al Cloud 124

Análisis del Uso de un Cluster de Raspberry Pi para Cómputo de Alto Rendimiento 134

XVII Workshop Tecnología Informática Aplicada en Educación (WTIAE) **145**

Validación de contenido de instrumentos para medir el nivel de integración tecnológica en el aula y el nivel de impacto en los estudiantes 146

Relato de experiencia de operación a distancia de un Reactor Nuclear para capacitación y entrenamiento 155

Use Case: OpenSource Virtualization for Educational Computer Laboratories 163

Diseño de una plataforma remota para desarrollo de prácticas de laboratorio 172

Descubrimiento de las áreas de investigación seleccionadas por los tesistas de las carreras de informática de la UM mediante árboles de decisión 183

Avances en la Personalización y Adaptación de Pruebas en un Ambiente Virtual de Aprendizaje 193

Normalización de la evaluación de calidad de producto software para el ámbito educativo 203

Análisis de Interacciones Sociales en Facebook 213

Diseño, desarrollo, publicación y evaluación de Objetos de Aprendizaje. Un estudio de casos 223

Análisis objetual de la interfaz gráfica de usuario de una plataforma de EaD. El caso MIeL (Materias Interactivas en Línea) 233

Desarrollo de Personajes Virtuales: estrategias y desafios para la construccion del aula virtual 243

Una Herramienta de Autor Web para Generar Aplicaciones Educativas Basadas en Posicionamiento con Interacción Física 253

Tecnologías aplicadas en el aula universitaria - Caso de aplicación en la enseñanza de programación 263

CadiDF una Herramienta de Soporte para la enseñanza aprendizaje 273

A Nomadic Testbed for Teaching Computer Architecture 284

Análisis de uso de un juego de realidad aumentada en la creación del hábito de la caminata en niños con necesidades especiales 294

Los e-Portfolios como Metodología de Evaluación y Seguimiento de Aprendizajes en Ámbitos Educativos 304

Propuesta pedagógica para la comprensión del espacio tridimensional utilizando Realidad Aumentada 310

XVI Workshop Computación Gráfica, Imágenes y Visualización (WCGIV) **320**

Inspección Automática de Defectos de Superficie en Baldosas Cerámicas 321

Virtual and Augmented Reality Therapy Framework for Phobia Treatment 331

Método de estimación de head pose para la obtención del punto de atención en base a una cámara web estándar 341

Simulador de tiro de tanques: NeoNahuel II 351

Reality Computing: An end-to-end process for Herpetological Heritage 361

Segmentación multi-atlas de imágenes médicas con selección de atlas inteligente y control de calidad automático 371

Mejora de imágenes a color utilizando un enfoque morfológico multiescala 381

Calibración de un Algoritmo de Detección de Anomalías Marítimas basado en la Fusión de Datos Satelitales 391

Recognition of Surface Irregularities on Roads: a machine learning approach on 3D models 401

Video Surveillance for Road Traffic Monitoring 412

XV Workshop Bases de Datos y Minería de Datos (WBDDM) **422**

Data Sources Integrator for Management by Indicators 423

Descubriendo reglas de asociación en bases de datos del sector retail usando R 432

Proceso para el Descubrimiento de Asociaciones Espaciales Mediante Subgrafos Frecuentes 442

D3CAS: un Algoritmo de Clustering para el Procesamiento de Flujos de Datos en Spark 452

Replicacion de Bases de Datos en Modalidad maestro-esclavo, caso de estudio: Firebird SQL Server 462

Segmentación de Mercado Usando Técnicas de Minería de Datos en Redes Sociales 471

Text pre-processing tool to increase the exactness of experimental results in summarization solutions 481

Aceleración en la Recuperación de Información utilizando Algoritmos de Minería de Datos de R 491

Análisis Preliminar del Rendimiento de Algoritmos para el Procesos de Descubrimiento de Reglas de Pertenencia a Grupos 501

Clasificación de autores para un proceso de recomendación integrado a un metabuscador científico 511

Incremento de Semántica en Fases Iniciales de Transformación de Modelos de Datos 521

XV Workshop Ingeniería de Software (WIS) 531

White-Box Testing Framework for Object-Oriented Programming based on Message Sequence Specification 532

Método de inclusión de Hacking ético en el proceso de Testing de software 542

Software para Decodificación y Análisis de Archivos Binarios en Computadoras Navales 552

Propuesta de un Modelo de Proceso de Implantación de Sistemas Informáticos (MoProIMP) 561

Enfoque Holístico de Calidad: Aplicación de una Estrategia de Selección de Alternativas 571

Green BPM: ciclo de vida de procesos de negocio incorporando aspectos ambientales 581

Problemas que afectan a la Calidad de Software en Entrega Continua y Pruebas Continuas 592

Exploring Specification Pattern based Behavioral Synthesis with Scenario Clauses	602
TG Project Management: Seguimiento y documentación de actividades orientada al usuario desarrollador	612
Análisis de consumo de energía en aplicaciones 3D para dispositivos móviles	622
Mobile Application Development Approaches: A Comparative Analysis on the Use of Storage Space	631
Usabilidad de Framework Web: identificación de problemas y propuesta de evaluación	642
SGR: Sistema para la Gestión de Requerimientos	652
Hacia una mejora de calidad en Scrum. Integrando documentación requerida por IRAM-ISO 9001:2015	662
XIII Workshop Arquitectura, Redes y Sistemas Operativos (WARSO)	**672**
Hub Of Things: Concentrador para Internet de las Cosas	673
Redes 5G y Virtualización de funciones de red en Argentina	683
Variante Heurística para el Problema del Viajante. Caso de Aplicación: Circuito de Pesca Deportiva	695
Interacción de elementos de networking con contenedores de software (docker)	705
Mejoras en maquinaria industrial con IoT. Hacia la Industria 4.0	715
Un análisis de comportamiento entre distintos mecanismos de control de congestión ensayados sobre una topología mixta	725
A Multi-objective Evolutionary Algorithms Study applied to Routing and Spectrum Assignment in EON networks	735
Análisis de rendimiento de protocolos de Publicación/Suscripción en comunicación con una Red de Sensores Inalámbricos Zigbee	745
Gestión de Calidad Móvil: Contratos, QoS y QoE	755

Sub-graph based Multicast Protection in WDM Networks, A Multi/Many-Objective Evolutionary Algorithms approaches 765

OBD-II vehicle data capture and monitoring system prototype 777

Análisis para Despliegue de una Red de Sensores Heterogénea 787

X Workshop Innovación en Sistemas de Software (WISS) **797**

Propuesta de Proceso para Selección de Arquitectura de un Modelo Predictivo Inteligente 798

An Open Source Solution for Money Bill Recognition for the Visually Impaired User using Smartphones 808

SIP (Sistema de Identificación de Personal), una solución informática al servicio de la seguridad de nuestros adultos mayores 818

Interoperabilidad e Integración entre Sistemas de Información en Salud 827

Framework para la Generación de Templates en Sistema de Catálogos de Realidad Aumentada 837

Desarrollo multiplataforma de Aplicaciones Móviles combinadas con el uso de Beacons 847

Towards a Framework for Ontology-Based Data Access: Materialization of OWL Ontologies From Relational Databases 857

A GraphicalWeb Tool with DL-based Reasoning Support over Orthogonal Variability Models 867

Registro emocional de personas interactuando en contextos de entornos virtuales 877

Análisis de las Tecnologías de la Información y la Comunicación y su innovación en la industria 887

Técnicas de control de concurrencia en base de datos: implementación en un sistema de gestión 897

Prototipo de dispositivo de ubicación a través de puntos estratégicos para personas no videntes o con discapacidad visual 907

Towards a Visual SPARQL-DL Query Builder 917

IX Workshop Procesamiento de Señales y Sistemas de Tiempo Real (WPSTR) **927**

Detección de carriles en caminos rurales no pavimentados 928

Software de segmento terreno de próxima generación 938

Interleaving Scheduling Algorithm for SLM Transactions in Mode S Surveillance Radar 948

Reconstrucción topográfica a partir de iluminación estructurada. Su aplicación en Radioterapia Externa 957

Control Activo de Ruido Basado en DSP para un Soplante Industrial 967

Digitalización de Imágenes de ECG para la Detección del Síndrome de Bayés 979

Prototipo Avanzado de un Simulador de Planificación de Tareas de Tiempo Real en Sistemas Heterogéneos 989

VII Workshop Innovación en Educación en Informática (WIEI) **999**

Adelantos en el desarrollo de un Protocolo de Interpretación Gráfica para no videntes y herramienta de implementación 1000

Checkers Game, un agente de Software para jugar a las Damas 1010

Propuestas de aplicaciones móviles desde la cátedra Sistemas de Información 1019

Construyendo FrankLab: Una Plataforma Web de Rob´otica Educativa 1030

Ciencias de la Computación y escuelas ¿una didáctica específica? 1040

VII Workshop Seguridad Informática (WSI) **1050**

Application of Keystroke Dynamics Modelling Techniques to Strengthen the User Identification in the Context of E-commerce 1051

An Analysis of Convolutional Neural Networks for detecting DGA 1060

Análisis, desarrollo e implementación de sistema que permite detectar páginas fraudulentas 1070

Track Gobierno Digital y Ciudades Inteligentes **1081**

Modelos de evaluación de gobiernos abiertos, aplicado a los municipios de la provincia de Buenos Aires	1082
Gestor de contenidos orientado a portales de organismos gubernamentales	1092
Caso de estudio sobre GDPR aplicado en Sistemas de Gestión Académica	1103
Monitoreo de unidades de transporte público de Ciudad del Este	1113
Sistema de medición de energía multiusuario de alta precisión	1122
Detección de Averías Viales Mediante IoMT Aplicada a Smart Cities	1133
Análisis del diseño adaptativo en sitios web gubernamentales	1142
Comparison of services for the recognition of flora images. Uses in augmented reality and tourism	1152
Value enhancement of the Artistic Heritage of Tierra del Fuego using Augmented Reality	1160

XIX Workshop Agentes y Sistemas Inteligentes (WASI)

Coordinadores

Sergio A. Gómez (UNS) - Guillermo Leguizamón (UNSL) - Carolina Salto (UNLP)

Temas

Approximate Reasoning
Bio-Inspired Metaheuristics
Constrains, Satisfiability and Search
Intelligent Data Mining
Intelligent Robotics
Intelligent System Performance Measurements
Intelligent Systems
Knowledge Representation
Machine Learning
Metaheuristics Based on Collective Intelligence
Multi-Agent Systems
Natural Language Processing
Planning and Scheduling
Reasoning and Logic
Semantic Web

Un enfoque de Recomendación multi-dominio basado en Embeddings y redes semánticas

Ignacio Gatti, Silvia Schiaffino, and Andrés Diaz Pace

ISISTAN (CONICET-UNCPBA),Campus Universitario, Tandil
{ignacio.gatti,silvia.schiaffino,andres.diazpace}@isistan.unicen.edu.com

Abstract. Actualmente, la generación de información en la web crece exponencialmente. Los Sistemas de Recomendación se valen de ésta para ofrecer a los usuarios potenciales ítems que pueden ser valiosos para ellos. Sin embargo, la mayoría de estos sistemas se mueven dentro de un sólo dominio (por ej., películas), desaprovechando información de otros contextos que pueden ser valiosa para la construcción del perfil del usuario, especialmente en escenarios de *cold-start*. El objetivo de este trabajo es aprovechar la información de un dominio puntual (películas) para generar recomendaciones más precisas en uno nuevo (libros). Básicamente, se propone un enfoque híbrido que, por un lado, utiliza un grafo semántico que relaciona artistas; y, por el otro, utiliza la técnica de *word embeddings* para encontrar relaciones de similitud entre las sinopsis de ambos ítems. Los resultados de una evaluación inicial muestra que el enfoque constituye una alternativa interesante para problema del *cold-start*.

Keywords: Recomendación multidominio, *word embedding*, redes semánticas

1 Introducción

En esta época donde el concepto de producción se da en su máxima expresión, desde la generación de datos e información hasta el desarrollo de nuevos productos; donde los usuarios son asfixiados por el peso de ésta, los Sistemas de Recomendación (SR) proponen una alternativa para que ellos puedan "filtrar" de forma automatizada de qué cosas quieren ser informadas. Básicamente, un SR es una pieza de software cuyo propósito es sugerirle al usuario potenciales ítems que consumirá en el futuro [14].
En sintonía con lo anterior, la popularidad de las redes sociales y los comercios electrónicos han permitido la recolección de infinidad de datos sobre el comportamiento de los usuarios a la hora de recorrer dichos lugares, dejando rastros de aquellos productos que les llamó la atención, de aquellos que compraron y los que no; de las películas que les gustaron. Estas huellas sirven de insumo para la construcción de un perfil personalizado. Asimismo, estas redes han permitido el desarrollo de diversas comunidades en la Web donde se deposita información sobre distintos tópicos, desde científicos a entretenimiento. Así, de manera análoga a los perfiles de usuario, es posible generar una descripción compleja de los ítems a recomendar.

Sin embargo, la construcción de dichos perfiles, presenta un desafío en el área de los SR. En la mayoría de los SR desarrollados, se observa que los perfiles generados se restringen al dominio en el que el SR se mueve. Básicamente, al trabajar sobre un único dominio, se presupone que el interés de un usuario sobre algunos de sus elementos está formado únicamente en aspectos propios de éste. Esta suposición coloca varias restricciones para hacer frente a dos problemas clásicos en SR: el arranque en frío (conocido como *cold-start*) y la escasez de ratings. El primero, es un caso particular del segundo, que se da cuando se agrega un nuevo usuario o ítem al sistema, para el cual no se posee ratings que permitan generar recomendaciones personalizadas. En el segundo, los ratings recolectados son tan pocos que repercute negativamente en la calidad de las recomendaciones. En el seno de estas problemáticas, se colocan los sistemas de recomendación multi-dominio. Básicamente, el objetivo que se proponen es poder extraer conocimiento de un dominio origen (D_s) para poder aplicarlo en un dominio objetivo (D_t), de manera tal de poder mejorar las recomendaciones en este último [2].

El presente trabajo propone un enfoque novedoso para atacar la problemática presentada, materializada en los dominios de películas y libros (Fig. 1). En líneas generales, la cuestión radicó en la definición de un espacio de representación donde los ítems de ambos dominios resultasen comparables. Para ellos, se extrajo la información textual que estuviera avalada por la comunidad de ambos dominios (Wikipedia[1], IMDB[2], GoodReads[3]). En la literatura, la forma clásica para la representación de documentos es el enfoque TF-IDF, donde cada ítem es mapeado a un vector de palabras de acuerdo a la frecuencia de aparición de cada una de ellas en éste [5]. Sin embargo, aquí se utilizó el concepto de *word embedding* para su representación. Coloquialmente, éstos son "descripciones" de palabras en forma de características o categorías compartidas a través éstas [9]. Finalmente, para vincular a los directores con los escritores se definió una red de influencia que permite calcular la distancia que existe entre cada par. Aprovechando la información semántica que ofrece Wikipedia sobre los artistas, es posible construir una red a partir de las etiquetas *"Influenciado por"* y *"Influencia a"*. Si bien los resultados son preliminares, se puede avizorar que el enfoque desarrollado es una alternativa para atacar el problema del *cold-start*.

El resto del trabajo se organiza de la siguiente manera. En la sección 2 se analizan los trabajos relacionados que sustentan el enfoque. La sección 3 presenta el enfoque desarrollado. En la 4 muestran los resultados, y se discuten pros y constras del enfoque. En la sección 5 se presentan las conclusiones y se esbozan lineas de trabajo futuro.

[1] https://www.wikipedia.org/
[2] https://www.imdb.com/
[3] https://www.goodreads.com/



Fig. 1: Esquema: SR multi-dominio

2 Trabajos Relacionados

A la hora de analizar SR multi-dominio, Cantador[2] propone una clasificación tomando como criterio la forma en que "explotan el conocimiento", definiendo dos ramas: *agregación de conocimiento* y *transferencia de conocimiento*. El enfoque presentado en este trabajo se encuentra en la segunda rama, con lo cual a continuación se analizarán algunos de los trabajos en el área.
En este tipo de técnicas, el foco está puesto en tomar al D_s como soporte para extraer conocimiento en forma de patrones para aplicarlo en las recomendaciones a realizar sobre el D_t. En esta dirección, el trabajo de [4] propone un framework basado en grafos semánticos (extraído de *DBpedia Ontology*[4]), para tomar conocimiento de diversos dominios y aplicarlos en la recomendación de ítems en D_t. Puntualmente, ellos aprovechan dicho grafo para encontrar relaciones entre un lugar de interés (POI) que visita el usuario y la música más apropiada para escuchar allí. Para esto, recorren la estructura buscando músicos relacionados con el POI. Por otro lado, en [12] se propone un enfoque de *Transferencia por Factorización Integrativa*, donde buscan atacar el problema de la escasez de datos en D_s (puntualmente películas), tomando una matriz auxiliar del mismo dominio pero de otro sistema, que posee una probabilidad de distribución de ratings para algunos ítems. Básicamente, dicha probabilidad define una restricción a la hora de predecir un ítem en D_t. Según ellos, lo interesante de este enfoque es que dicha distribución de probabilidad puede obtenerse de manera implícita, sin tener que pedirle a los usuarios que voten determinados ítems. Otro trabajo interesante en esta rama es [6], donde la transferencia de conocimiento se realiza a través de la definición de un codebook. Éste se define a partir de patrones extraídos en D_s, generalmente aplicando alguna técnica de factorización sobre la matriz del sistema (Usuarios x Ítems), para luego llevar a D_t, donde se reconstruye la matriz del sistema del dominio objetivo aplicando el codebook.
Los *word embeddings* irrumpieron en el mundo del Procesamiento del Lenguaje Natural (NLP), provocando una revolución en tareas como el reconocimiento de

[4] https://wiki.dbpedia.org/services-resources/ontology

entidades o análisis sintáctico. Desde entonces, el área de Recuperación de la Información (RI) lo ha adoptado como una nueva estrategia para la recuperación de documentos, teniendo resultados alentadores. Vale la pena resaltar el trabajo [7], donde propone definir cada documento y la consulta como un promedio de los vectores de *word embeddings* de sus palabras. De esta forma los lleva a un espacio vectorial donde puede aplicar similitud de coseno para encontrar documentos similares. En la misma dirección, se puede ubicar [11], que utiliza una implementación de *word embeddings*, llamada *Word2Vec*, en el modelado de sus ítems. Básicamente, éste se centra en la recomendaciones de puntos de interés en una ciudad para un visitante, representando cada lugar a partir de los "check-ins" que hace el usuario y mapeando estos a un espacio vectorial utilizando los *word embeddings*.

3 Enfoque propuesto

El presente enfoque propone una estrategia para poder extraer conocimiento de D_s -en este caso películas-, para aplicarlo en D_t -libros-. Puntualmente, el objetivo del trabajo es poder establecer un método que utilice todo el conocimiento que se tiene de D_s -matriz de ratings, información de los ítems aportada por la comunidad-, para poder ofrecerle recomendaciones más precisas y apropiadas a los usuarios de D_s en el D_t -donde solo se posee información de los ítems-.
Desde el punto de vista de los SR tradicionales, Ricci[14] marca dos enfoques para atacar el problema de la recomendación: *filtrado colaborativo* y *basado en contenido*. El primero, solo usa la matriz ratings para calcular las recomendaciones, buscando usuarios similares (es decir que hayan votado un mismo subconjunto de ítems de forma análoga) para recomendar aquellos ítems que un usuario aún no ha consumido y que otro con gustos parecido sí experimentó positivamente. Mientras que el segundo, define un espacio de representación al que son mapeados ítems y usuarios con el fin de encontrar aquellos ítems que son similares para un usuario determinado. Aquí se optó por este último, dado que es conocido su potencialidad para tratar el *cold-start*, como lo muestra [8] en su trabajo de análisis comparativo de ambos enfoques.
Como se dijo anteriormente, uno de los problemas centrales que se presentaron fue la representación de los ítems. Porque si bien ambos pueden entrar en las categorías de "entretenimiento" o "arte", la naturaleza de ellos es distinta: las películas son objetos audiovisuales, mientras que los libros objetos escritos. Por eso, de la diversas fuentes de datos que existen de ambos (por ej., sinopsis, imágenes, guión), sólo se escogieron aquellas que estuvisten en formato texto, es decir, se tomaron las sinopsis, los géneros, los nombres de los directores y escritores. De esta forma, se logró una homogeneización de los datos para luego poder procesarlos y representar a todos los ítems en un mismo espacio.
De esta selección de datos, un subconjunto de ellos poseen una correlación directa -las sinopsis y los géneros-, mientras que entre directores y escritores existen una serie de relaciones implícitas. Por el lado de los primeros, ambos pueden ser visto como documentos que expresan en palabras el contenido de la obra, ya sea

a través de una reseña o a través de un conjunto de palabras claves (como lo son los géneros); con lo cual, sobre ellos se puede aplicar alguna técnica conocida de RI para modelarlos. Por el otro lado, directores y escritores son casos particulares, porque entre ellos se tejen relaciones de tipo social e intelectual; de manera tal que el conocimiento existente allí se desprende de esa red, siendo necesario algún análisis de grafo.
Por eso, en los siguientes apartados se detallan las dos estrategias definidas para mapear ambos dominios a un mismo espacio: red de influencia y *embedding*.

3.1 Red de influencia

Hoy en día, Wikipedia es una de las mayores fuentes de conocimiento existente en Internet, con una comunidad de más de 307.000 usuarios que participan activamente. Más allá de la revolución que ésta significa por su edición de manera colaborativa, lo que la hace aún más potente es la ontología (*DBpedia Ontology*) que se ha ido desarrollando trás ella. Ésta ha sido generada manualmente (siguiendo el estándar RDF[5]), a partir de los infobox de cada entrada de Wikipedia, cubriendo actualmente 685 clases relacionadas de forma jerárquica, con 2.795 propiedades. Asimismo, ésta puede ser consultada programáticamente, utilizando el lenguaje SPARQL[6], ofreciendo la posibilidad de hacer diversos recorridos sobre el grafo subyacente.
En este trabajo, se observó que una de las propiedades que poseían los artistas registrados en Wikipedia eran las etiquetas: *"Influido por"* o *"Influye a"*. Con lo cual, resultaba factible poder extraer de dicha ontología un "grafo de influencias" (Fig. 2), donde se materializaba las relaciones intelectuales entre directores-directores, escritores-escritores y directores-escritores. Cuanto más cerca estuvieran los artistas en el grafo, mayor sería su afinidad en sus obras, pues "todo artista siempre es una amalgama entre la tradición que mamó y su pulsión por romper con ella" [10]. A partir de esta premisa, se definió la métrica de influencia 1 para poder ir del espacio de directores al de escritores.

$$Influencia(director, escritor) = 1/camino_mas_corto(director, escritor) \quad (1)$$

De esta forma, es posible determinar cuánta influencia de cada escritor tiene una película a partir de su director (lo mismo vale para el caso director y libro), permitiendo representar una película en el espacio de los libros y viceversa. Así, es factible establecer un espacio de representación común donde películas y libros se vuelven comparables.

3.2 Word Embedding

Con la irrupción de las redes neuronales, en NLP ha surgido un cambio de visión en la forma de representación del lenguaje, inclinándose por modelos vectoriales

[5] https://www.w3.org/TR/rdf-schema/
[6] http://dbpedia.org/sparql

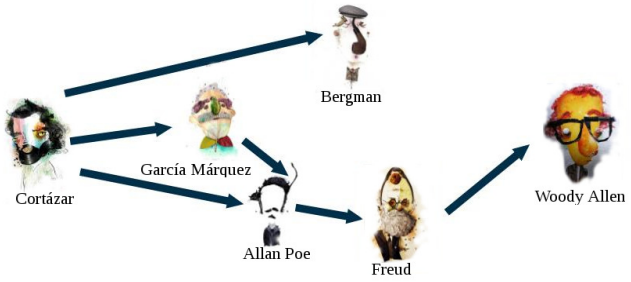


Fig. 2: Red de influencias

que permiten almacenar relaciones internas entre las palabras. Básicamente, estas redes son entrenadas con grandes datasets, obteniendo así modelos que aprenden, por ejemplo, relaciones gramaticales o probabilidad de coocurrencia entre palabras. Entre ellos,*word embeddings* es uno de los modelos que posee mayor popularidad por su buen desempeño en tareas como análisis sintáctico[15], análisis de sentimiento[16] o reconocimiento de entidades[3] .
En palabras de Mikolov[9], los *word embeddings* son vectores n-dimensionales en un espacio vectorial continuo donde cada dimensión representa conceptos o categorías compartidas a través de palabras. Pero se trata de una "categorización interna", que no se traduce a aquellas definidas por el intelecto humano. Por ejemplo, aunque "rey" y "reina" no son sinónimos, hiperónimos ni comparten ninguna relación gramatical, es sabido que están relacionados y, de hecho, sus representaciones están cerca en un espacio generado por los *embeddings* (Fig.3).
Si bien los detalles de los tipos de redes neuronales que se utilizan para obtener

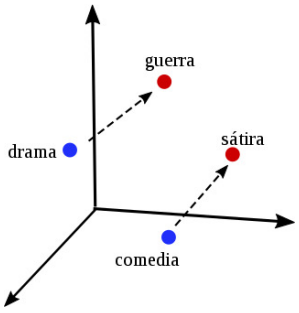


Fig. 3: Word embeddings: espacio vectorial

estos modelos exceden a este trabajo, resulta importante destacar los dos enfoques: *predictivo* y *por conteo*. El primero, parte de la idea que las palabras se ven afectadas inherentemente por el contexto en el que se utilizan. De esta forma, [9] propone generar un modelo que pueda predecir la aparición de una palabra

objetivo en un contexto determinado en función de las características que tiene la *palabra de objetivo*. Si se minimiza el error en este proceso para cada palabra en el vocabulario, se obtendrá efectivamente buenas representaciones para las palabras, porque la predicción de una palabra en este escenario implicará que se hayan modelado correctamente las características de las palabras individuales. Este enfoque también se llama *Skip-Gram* y es el que se utiliza para generar el modelo *Word2Vec*[7]. Por otro lado, [13] propone el otro enfoque, obteniendo el modelo *GLoVe*[8]. En líneas generales, éste consta de dos pasos: primero calcula la probabilidad de co-ocurrencia entre las palabras a lo largo del corpus, determinando que una palabra se relaciona con otra palabra si tienen una alta probabilidad conjunta. Luego, para crear los *embeddings* se realiza un paso de "aprendizaje" donde éstos se refinan para que sean representaciones efectivas de las probabilidades conjuntas, por ejemplo, usando regresión logística. En otras palabras, la idea es crear una matriz que pueda "explicar" la matriz original minimizando la pérdida de reconstrucción, que es la cantidad de información que se pierde cuando se reconstruye la matriz.
En este trabajo se utilizaron dos de los corpus pre-entrenados más utilizados por su robustez: *GLoVe* y *Word2Vec*. Ambos, fueron entrenados con un vocabulario de un par de millones de palabras, generando un vector de 300 dimensiones para cada una de ellas.
A partir de allí, se propuso representar a cada película y libro en el espacio dimensional definido por los *word-embeddings*. Para ello, siguiendo la estrategia definida en [7], se obtuvo el vector promedio de la suma de vectores de embeddings de cada palabra de la sinopsis y los géneros.

$$mean_embedding(document) = 1/n * \sum_{d_i \in document} word_embedding(d_i) \quad (2)$$

3.3 Recomendación multi-dominio

Para realizar la recomendación de ítems en D_t -en este caso libros-, se siguió el enfoque clásico basado en contenido definido en [1]. Básicamente, una vez mapeados los ítems de ambos dominios a un mismo espacio vectorial, sólo queda llevar al usuario a dicho espacio, de forma tal de buscar los libros que sean más similares a sus preferencias.
Para ello, se definió el vector de un usuario como el promedio de la suma de los vectores que éste consumió en Ds, ponderado por el rating que le dio.

$$vector_u = 1/n * \sum_{i \in R_u} mean_embedding(i) * r_{ui} \quad (3)$$

Ya en el espacio vectorial, resulta sencillo encontrar vectores similares utilizando similitud del coseno.

$$sim_{cos}(vector_u, embedding_i) = \frac{vector_u \cdot embedding_i}{|vector_u| * |embedding_i|} \quad (4)$$

[7] https://github.com/mmihaltz/word2vec-GoogleNews-vectors
[8] https://nlp.stanford.edu/projects/glove/

4 Evaluación

En esta sección se presentan los resultados de la evaluación del enfoque.

4.1 Dataset

La evaluación del enfoque se realizó sobre el recorte de un dataset de Amazon[9]. Puntualmente, de aquí se extrajeron, por un lado, 6354 títulos de libros y 326 películas, que luego fueron enriquecidos con datos extraídos de la comunidad de Goodreads y IMDB respectivamente. Por el otro, se obtuvieron dos matrices de ratings, una de películas con 67.193 entradas y otra de libros con 223.695.

4.2 Resultados

Para evaluar la propuesta de este trabajo, se tomó como indicador el error en la predicción -tanto el Error Absoluto (MAE) como el Error Cuadrático Medio (MSE)-. Tal como destaca Cantador en [2], en el área de RS multi-dominio, uno de los objetivos es poder lograr predicciones más precisas en D_t utilizando sólo el conocimiento existente en D_s. Por eso, suele utilizarse como dataset de entrenamiento todo lo referente a D_s, dejando para la prueba los ratings de D_t.
A partir del enfoque propuesto, se implementaron una serie de variantes que exploran las posibilidades ofrecidas. Por un lado, se definió toda una rama que mapea los ítems al dominio de las películas (*To movie space*) y el otro que los lleva al de libros (*To book space*). En ambos caso, la variación se encuentra en la forma que se recorre el grafo: de los escritores a los directores (determinando cuánta influencia de cada director tiene un libro) o viceversa (determinando cuánta influencia de cada escritor tiene una película) . A su vez, se posee un grupo de implementaciones cuyo espacio vectorial fue calculado con *GLoVe* y otro con *Word2Vec (W2B)*. La combinación de ambos genera una hibridización, que es manejada con la importancia que se la da a cada aspecto (*Red de Influencias*, *word embeddings*).
Asimismo, se definieron dos enfoques que se tomaron como base para realizar posteriormente un análisis comparativo. El primero, se trata de uno basado en popularidad, que suele ser uno de los más utilizados en la industria para resolver el problema del *cold-start*[5]. El segundo, es uno que calcula el espacio vectorial utilizando TF-IDF, por ser la alternativa clásica que propone en [1].
El estudio realizado consistió en predecir, para aquellos usuarios que poseían ratings en ambos dominios, los ratings en D_t. Con dichas predicciones se calcularon, los dos indicadores clásicos de error: MAE y MSE. Como se puede observar en la tabla, todas las variaciones del enfoque propuesto mejoran el MAE, mientras que el MSE sólo es mejorado por aquellas variaciones que dan la misma importancia a los *Embeddings* y a la *Red de Influencias*.
Si bien se tratan de resultados preliminares (Tabla 1), en todos los casos se

[9] http://jmcauley.ucsd.edu/data/amazon/links.html

mejora la estimación obtenida con el enfoque basado en popularidad. Sin embargo, al analizar el comportamiento respecto de aquel que usa TFI-DF es posible observar que para un subconjunto de predicciones, los embeddings producen estimaciones que suelen distar mucho del rating definido por el usuario, de allí que el MSE sea un poco mayor. Mientras que si se pone el foco en el MAE, es posible afirmar que las predicciones son más precisas.

Enfoque	**Peso Embedding**	**Peso Red Influencia**	**MAE**	**MSE**
Popularity baseline			0,8574	1,0773
To movie space	0,5	0,5	0,6837	0,8464
To book space	0,5	0,5	0,7717	0,8854
To movie space Blend (W2B)	0,6	0,4	0,6467	0,9095
To movie space Blend (GLoVe)	0,6	0,4	0,6446	0,9285
To book space Blend (W2B)	0,6	0,4	**0,6427**	**0,8748**
To book space Blend (GLoVe)	0,6	0,4	**0,6401**	**0,8998**
To movie space Blend (W2B)	0,5	0,5	**0,6517**	**0,8799**
To movie space Blend (GLoVe)	0,5	0,5	0,6455	0,8868
To book space Blend (W2B)	0,5	0,5	**0,6505**	**0,8455**
To book space Blend (GLoVe)	0,5	0,5	**0,6438**	**0,8587**

Table 1: Comparación de métodos

5 Conclusiones

En este trabajo se presentó un enfoque novedoso en el área de SR multi-dominio que utiliza, por un lado, las técnicas de *word embeddings* para representar en un mismo espacio n-dimensional a libros y películas; y, por el otro, usa *DBpedia Ontology* para extraer información del grafo de influencias que subyace entre los artistas. El objetivo fue definir una estrategia que lograra homogeneizar los ítems de D_s y D_t, de forma tal que sean comparables. Allí, es donde juega un rol fundamental la red de influencias, pues ofrece la posibilidad de ver las películas en relación con la influencia que tiene director respecto de los escritores (lo mismo se puede aplicar para el caso de libro con directores).
Por el lado de las técnicas de *word embedding*, éstas se presentan como una alternativa a la clásica TFI-DF, ofreciendo la posibilidad de detectar relaciones implícitas entre palabras que pueden ser provechosas para encontrar similitudes entre las sinopsis de los libros y las películas.
Si bien se trata de una prueba de concepto, los resultados obtenidos mostraron que el enfoque propuesto mejora dos enfoques clásicos que se utilzan ante esta problemática. Por eso, queda como trabajo futuro seguir explorando el enfoque desde otros indicadores como precisión. También se trabajará sobre otras formas de modelar documentos utilizando *word embeddings*.

Bibliografía

1. Balabanović, M., Shoham, Y.: Fab: content-based, collaborative recommendation. Communications of the ACM 40(3), 66–72 (1997)
2. Cantador, I., Fernández-Tobías, I., Berkovsky, S., Cremonesi, P.: Recommender Systems Handbook, chap. Cross-Domain Recommender Systems. Springer US (2015)
3. Das, A., Ganguly, D., Garain, U.: Named entity recognition with word embeddings and wikipedia categories for a low-resource language. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 16(3), 18 (2017)
4. Fernández-Tobías, I., Cantador, I., Kaminskas, M., Ricci, F.: A generic semantic-based framework for cross-domain recommendation. In: Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems. pp. 25–32. ACM (2011)
5. de Gemmis, M., Lops, P., Musto, C., Narducci, F., Semeraro, G.: Recommender Systems Handbook, chap. Semantic-Aware Content-Based Recommder Systems. Springer US (2015)
6. Li, B., Yang, Q., Xue, X.: Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction. In: IJCAI. vol. 9, pp. 2052–2057 (2009)
7. Lizarralde, I., Rodriguez, J.M., Mateos, C., Zunino, A.: Word embeddings for improving rest services discoverability. In: Computer Conference (CLEI), 2017 XLIII Latin American. pp. 1–8. IEEE (2017)
8. Magnini, B., Strapparava, C.: Improving user modelling with content-based techniques. In: International Conference on User Modeling. pp. 74–83. Springer (2001)
9. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. pp. 3111–3119 (2013)
10. Nietzsche, F.W.: El nacimiento de la tragedia o Helenismo y Pesimismo. Espasa-Calpe (1943)
11. Ozsoy, M.G.: From word embeddings to item recommendation. arXiv preprint arXiv:1601.01356 (2016)
12. Pan, W., Xiang, E.W., Yang, Q.: Transfer learning in collaborative filtering with uncertain ratings. In: AAAI. vol. 12, pp. 662–668 (2012)
13. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 1532–1543 (2014)
14. Ricci, F., Rokach, L., Shapira, B.: Recommender Systems Handbook, chap. Introduction. Springer US (2015)
15. Socher, R., Bauer, J., Manning, C.D., et al.: Parsing with compositional vector grammars. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 455–465 (2013)
16. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C.D., Ng, A., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 conference on empirical methods in natural language processing. pp. 1631–1642 (2013)

SA to optimize the Multi-Period Water Distribution Network Design

Carlos Bermudez[1], Carolina Salto[1,2], and Gabriela Minetti[1]

[1] Facultad de Ingeniería - Universidad Nacional de La Pampa
Calle 110 Nro. 390, General Pico, La Pampa, Argentina
[2] CONICET, Argentina
bermudezc@yahoo.com –saltoc,minettig@ing.unlpam.edu.ar

Abstract. The design optimization of the water distribution network is an important issue in modern cities. For this reason, we propose an optimization technique based on Simulated Annealing (SA) to solve the Water Distribution Network Design problem, considering multi-period restrictions with time varying demand patterns. The SA is improved with a local search procedure, yielding a hybrid SA, in order to obtain good quality networks designs. A broad experimentation using different test networks is carried out to test our proposal. Moreover, a comparison with an approach from the literature reveals the goodness of our proposal, the hybrid SA.

Keywords: Water Distribution Network Design, Optimization, Metaheuristic, Simulated Annealing

1 Introduction

In any modern city, a safe, adequate, and accessible supply of potable water is one of the basic necessities of any human being. In order to satisfy this vital requirement, enough volumes of water must be transported from a source of potable water, e.g., a treatment plant, to demand points (consumers) through a network of pipes. It is very important that the solution concerning the layout, design, and operation of the network of pipes result from good planning and management procedures. Consequently, this problem known as Water Distribution Network Design (WDND) requires to manage an important number of variables (pipes, pipe diameters, demand nodes, water pressure, reservoirs, etc.) and constraints (water velocity, pressure, etc.). This problem, even for simple networks, is very difficult to solve, in particular it is classified as NP-hard [1].

Early research works in the WDND optimization area were focused on the single-period, single-objective, gravity-fed design optimization problem. The first research works applied linear programming [2, 3], and non-linear programming [4, 5]. After that, the metaheuristics have been used to solve these problems, such as the trajectory-based ones: Simulated Annealing [6, 7] and Tabu Search [8]. Also population-based metaheuristics were applied, for example, Ant Colony Optimization [9], Ant Systems [10], Genetic Algorithms [11–13] Scatter Search [14], and Differential Evolution [15].

More recent works extends the single-period problem to a multi-period setting in which time varying demand patterns occur. In [16], the authors formulate the design problem as a multi-objective optimization problem and apply a multi-objective evolutionary algorithm. A Genetic Algorithm is used to solve six small instances considering velocity constraint on the water flowing through the distribution pipes in [12]. This constraint is also taken into account in [17], but the authors use mathematical programming on bigger, closer-to-reality instances. In [18], an Iterative Local Search is specifically-designed in order to consider that every demand node has 24 hrs water demand pattern and a new constraint, which imposes a limit on the maximal velocity of water through the pipes. Based on this last problem formulation, we propose a optimization technique in order to improve and optimize the distribution network design.

In this paper, we present a Simulated Annealing algorithm to optimize the optimal type of pipe connecting the supply, demand, and junction nodes in the distribution network. This SA incorporates a local search procedure in order to improve the layout of the network, arising the Hybrid Simulated Annealing (HSA) proposed. We test the performance of our proposal with a set of different networks with different sizes expressed by number of pipes and characteristics. The evaluation considers relevant aspects such as efficiency and internal behavior. Moreover, a comparison with state-of-the-art is carried out to outline the goodness of our proposal.

The rest of this article is organized as follows. In Section 2, we introduce the problem definition. Section 3 explains our algorithmic proposal, HSA, to solve the Water Distribution Network Design Optimization Problem. Section 4 describes the experimental analysis and the methodology used. Then, we analyze the results obtained by HSA and compare with the obtained by the ILS [18], in Section 5. Finally, we present our principal conclusions and future lines of research.

2 Multi-Period Water Distribution Network Design

This work focuses on obtaining the minimum cost in a water distribution network design. The problem can be characterized as: simple-objective, multi-period, and gravity-fed. Two restrictions are considered: the limit of water speed in each pipe and the demand pattern that varies in time.

The objective of the WDND problem is to minimize the total investment cost (TIC) of a water distribution network. The network can be modeled by a connected graph, which is described by a set of nodes $N = \{n_1, n_2, ...\}$, a set of pipes $P = \{p_1, p_2, ...\}$, a set of loops $L = \{l_1, l_2, ...\}$, and a set of commercially available pipe types $T = \{t_1, t_2, ...\}$. The TIC is obtained by the formula shown in Equation 1.

$$\min TIC = \sum_{p \in P} \sum_{t \in T} L_p IC_t x_{p,t} \quad (1)$$

where IC_t is the cost of a pipe p of type t, L_p is the length of the tube, and $x_{p,t}$ is the binary decision variable that determines whether the tube p is of type t or not. The objective function is limited by: physical laws of mass and energy conservation, minimum pressure demand in the nodes, and the maximum speed in the pipes, for each time $\tau \in \mathcal{T}$. These laws are explained in the following paragraphs.

Mass conservation law: It must be satisfied for each node N in each period of time τ. This law establishes that the volume of water flowing towards a node in a unit of time must be equal to the flow that leaves it (see Equation 2).

$$\sum_{n_1 \in N/n} Q_{(n_1,n),\tau} - \sum_{n_2 \in N/n} Q_{(n,n_2),\tau} = WD_{n,\tau} - WS_{n,\tau} \quad \forall n \in N \quad \forall \tau \in \mathcal{T} \quad (2)$$

where $Q_{(n_1,n),\tau}$ is the flow from node n_1 to node n at time τ, $WS_{n,\tau}$ is the external water supplied and $WD_{n,\tau}$ is the external water demanded.

Energy conservation law: It states that the sum of pressure drops in a closed circuit in an instant of time τ is zero. These drops can be approximated using the Hazen-Williams equations with the parameters used in EPANET 2.0 [19] (the hydraulic solver used in this paper), as indicated in Equation 3.

$$\sum_{p \in l} \left[\frac{10.6668 y_{p,\tau} Q_{p,\tau}^{1.852} L_p}{\sum_{t \in T} (x_{p,t} C_t^{1.852} D_t^{4.871}} \right] = 0 \quad \forall l \in L \quad \forall \tau \in \mathcal{T} \quad (3)$$

In Equation 3, $y_{p,\tau}$ is the sign of $Q_{p,\tau}$ that indicates changes in the water flow direction relative to the defined flow directions, $Q_{p,\tau}$ is the amount of water flowing through pipe p in time τ, L_p is the pipe length, C_t is the Hazen-Williams roughness coefficient of pipe type t, and D_t is the diameter of pipe type t.

Minimum pressure head requirements: for each node n in each period of time τ, it must be satisfied (see Equation 4).

$$H_{n,\tau}^{min} \leq H_{n,\tau} \quad \forall n \in N \quad \forall \tau \in \mathcal{T} \quad (4)$$

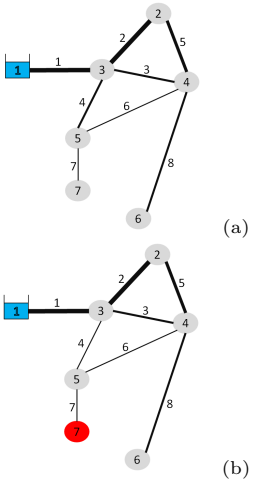
being H^{min} the minimum node pressure and $H_{n,\tau}$ the node's current pressure.

Maximum water velocity: The water velocity $v_{p,\tau}$ can not exceed the maximum stipulated speed $v_{p,\tau}^{max}$. Equation 5 shows this relationship.

$$v_{p,\tau} \leq v_{p,\tau}^{max} \quad \forall p \in P \quad \forall \tau \in \mathcal{T} \quad (5)$$

3 Our Proposal for the Multi-Period WDND Problem

Simulated Annealing (SA) is a simple trajectory-based metaheuristic [20], which is based on a thermal process for obtaining low energy states of a solid in a heat bath. At the beginning (with a high temperature), SA accepts solutions with high cost values under a certain probability in order to explore the search space and to escape from local optima. During the annealing process this probability



(a) (b)

Pipe ID	Length (m)
1	31
2	20
3	35
4	37
5	24
6	50
7	12
8	65

(c)

Diam. (mm)	Roughness unitless	Cost
60	130	22
80	130	35
100	130	38
150	130	50
200	130	61

(d)

Fig. 1. Different solutions or network designs. (a) Solution 1; (b) Solution 2; (c) Pipe lengths; (d) Available pipe types with their corresponding costs.

Table 1. Different solutions or network designs in vector representation.

Solution	Pipe ID	1	2	3	4	5	6	7	8	Feasibility
	Length (m)	**31**	**20**	**35**	**37**	**24**	**50**	**12**	**65**	**TIC**
1	diam. (mm)	150	150	80	80	100	60	60	80	feasible
	cost	1550	1000	1225	1295	912	1100	264	2275	9621
2	diam. (mm)	150	150	80	60	100	60	60	80	infeasible
	cost	1550	1000	1225	814	912	1100	264	2275	9140

decreases according to temperature cooling; intensifying the search and reducing the exploration in order to exploit a restricted area of a search space.

Simulated annealing evolves by a sequence of transitions between states and these transitions are generated by transition probabilities. Consequently, SA can be mathematically modeled by Markov chains, where a sequence of chains is generated by a transition probability, which is calculated involving the current temperature.

Our proposal consists in adapting and hybridizing the SA algorithm to solve the Multi-Period WDND optimization problem. In this way, the Hybrid Simulated Annealing (HSA) algorithm arises. A solution to this problem is a network, as shown in figures 1 (a) and (b). A network or a solution is represented by a vector, where each element is the diameter selected for that pipe, as can be seen in Table 1. In this table the vectors that represent the candidate solutions in figures 1(a) and (b) are shown. The total investment cost for each solution is calculated by the Equation 1, using the input data from tables (c) and (d) of Figure 1. The first solution is hydraulically feasible (satisfying all constraints mentioned in Section 2) and the second one is infeasible (violating the minimum pressure constraint in node 7).

Algorithm 1 HSA Algorithm to solve the WDND optimization Problem

```
1: t = 0;
2: initialize T and S_0; {temperature and initial solution}
3: evaluate S_0 in TIC_0;
4: repeat
5:   repeat
6:     t = t + 1;
7:     generate S_1 from S_0 applying the MP-GRASP Local Search;
8:     evaluate S_1 in TIC_1;
9:     if TIC_1 < TIC_0 then
10:      S_0 = S_1; TIC_0 = TIC_1
11:    end if
12:    generate S_2 from S_0 applying the perturbation operator;
13:    evaluate S_2 in TIC_2;
14:    if (TIC_2 < TIC_0) or (exp((TIC_2 − TIC_0)/T) > random(0, 1)) then
15:      S_0 = S_2; TIC_0 = TIC_2
16:    end if
17:  until (t mod Markov Chain Length) == 0
18:  update T;
19: until stop criterion is met
20: return S_0;
```

In Algorithm 1, we show a pseudo-code of the HSA algorithm to solve the WDND optimization problem. HSA uses the EPANET 2.0 toolkit [19] to solve the hydraulic equations, since this hydraulic solver is applied in most existing works. HSA generates a feasible initial solution S_0 applying both `HighCost` and `Lowcost` mechanisms proposed in [18] (line 2). After the evaluation of the initial solution (line 3), an iterative process starts (lines 4 to 19). As a first step in the iteration, the hybridization is carried out in order to intensify the search into the current region of the solution space. In this way a feasible solution, S_1, is obtained by applying the MP-GRASP local search [18] to S_0 (line 7), and then a greedy selection mechanism is performed (lines 9-11). As a consequence, S_0 can be replaced by S_1 if this is better than S_0. In the next step a perturbation operator is used to obtain a feasible neighbor, S_2, from S_0 (line 12), in order to explore another areas of the search space. This perturbation randomly changes some pipe diameters. If S_2 is worse than S_0, S_2 can be accepted under the Boltzmann probability (line 14, second condition). In this way, at high temperatures (T) the exploration of the search space is strengthened. In contrast, at low temperatures the algorithm only exploits a promising region of the solution space, intensifying the search. In order to update T, the proportional cooling process [20] is used (line 18) and it is applied after a certain number of iterations (t) given by the Markov Chain Length (MCL) (line 17). Finally, SA ends the search when the total evaluation number ($\sharp Eval.$) is reached.

4 Experimentation and Results

In this section, we introduce the experimental design used in this approach, the execution environment, and the result analysis. In order to evaluate HSA, the HydroGen instances of WDND optimization problem [21] are solved. These instances arise from 15 different distribution networks (HG-MP-i), which information is shown in Table 2. A set of 16 different pipe types is used and their

Table 2. Information on the HydroGen networks.

Network	Meshedness Coefficient	Pipes	Demand Nodes	Water Reservoirs	Network	Meshedness Coefficient	Pipes	Demand Nodes	Water Reservoirs
HG-MP-1	0.2	100	73	1	HG-MP-9	0.1	295	247	2
HG-MP-2	0.15	100	78	1	HG-MP-10	0.2	397	285	2
HG-MP-3	0.1	99	83	1	HG-MP-11	0.15	399	308	2
HG-MP-4	0.2	198	143	1	HG-MP-12	0.1	395	330	3
HG-MP-5	0.15	200	155	1	HG-MP-13	0.2	498	357	2
HG-MP-6	0.1	198	166	1	HG-MP-14	0.15	499	385	3
HG-MP-7	0.2	299	215	2	HG-MP-15	0.1	495	413	3
HG-MP-8	0.15	300	232	2					

Table 3. Available pipe types and their corresponding costs.

Number	Diam. (mm)	Roughness	Cost	Number	Diameter	Roughness	Cost
1	20	130	15	9	200	130	116
2	30	130	20	10	250	130	150
3	40	130	25	11	300	130	201
4	50	130	30	12	350	130	246
5	60	130	35	13	400	130	290
6	80	130	38	14	500	130	351
7	100	130	50	15	600	130	528
8	150	130	61	16	1,000	130	628

Table 4. Average of the best TIC values obtained by ILS and HSA for all instances grouped by the corresponding network. The best values are boldfaced.

Network	ILS	HSA	Network	ILS	HSA
HG-MP-1	339200	**335883**	HG-MP-9	**831200**	834821
HG-MP-2	303800	**298842**	HG-MP-10	790600	**788422**
HG-MP-3	389200	**387089**	HG-MP-11	**907000**	909038
HG-MP-4	694200	**690033**	HG-MP-12	1022800	1046426
HG-MP-5	728200	**722218**	HG-MP-13	1190400	**1179568**
HG-MP-6	750400	**741638**	HG-MP-14	**1070200**	1085195
HG-MP-7	818000	**817262**	HG-MP-15	**1122600**	1161200
HG-MP-8	**851600**	855479			

characteristics and costs can be found in Table 3. The demand nodes are divided into five categories (domestic, industrial, energy, public services, and commercial demand nodes), each one with a corresponding base load and demand pattern[3]. In this way, five different instances are considered for each HG-MP-i network, resulting 75 instances in total.

The computational environment used in this work to carry out the experimentation consists of computers with INTEL I7 3770K quad-core processors 3.5GHz, 8 GB RAM, and the Slackware Linux with 3.2.29 kernel version. Because of the stochastic nature of the algorithms, we performed 30 independent runs of each instance to gather meaningful experimental data and apply statistical confidence metrics to validate our results and conclusions. As a result, a total of 2,250 executions were carried out.

The methodology followed to analyze the results is described in the following. First, we study the HSA behavior, comparing the best cost values found by

[3] The base loads can be found in the EPANET input files of the instances

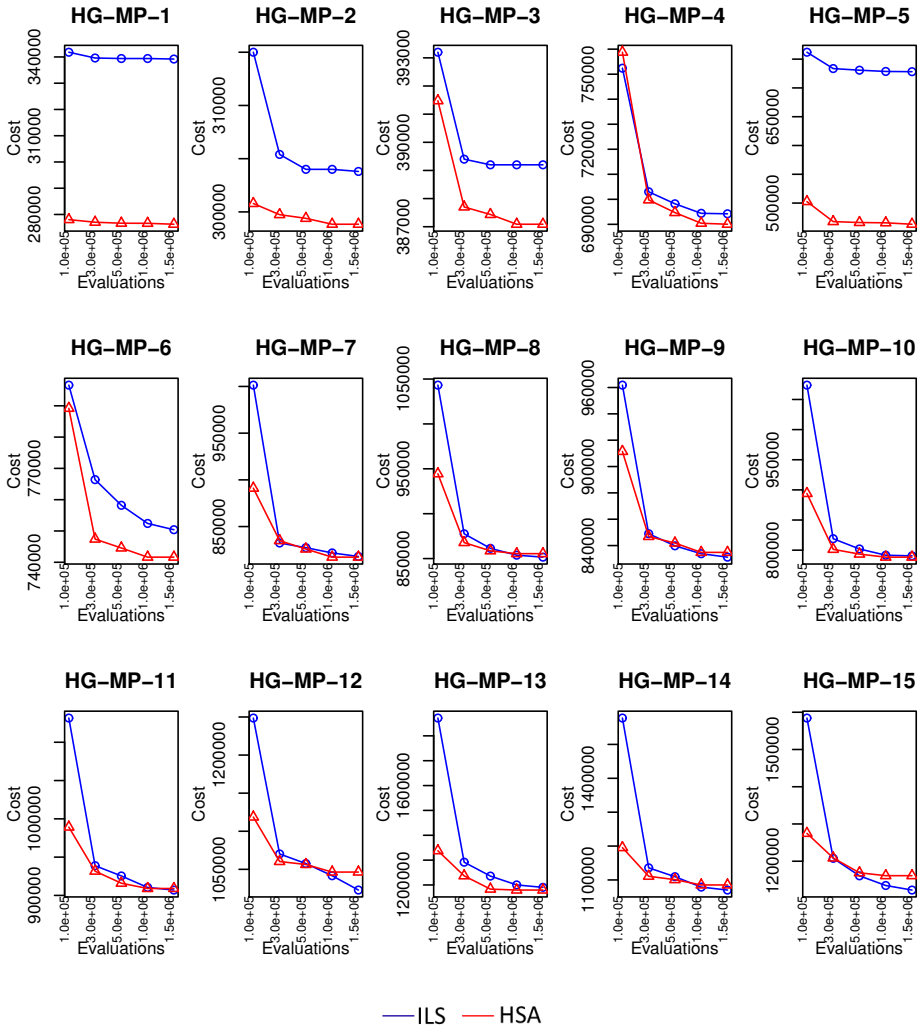


Fig. 2. Evolution of the TIC values during the search for all instances.

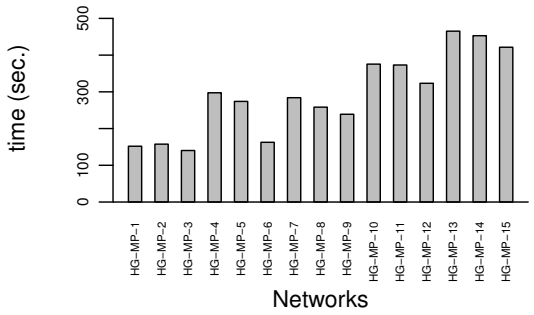


Fig. 3. Average of the total HSA's execution times for all instances grouped by the corresponding network.

our proposal and ILS [18] for the 75 instances grouped by their corresponding distribution network, as presented in Table 4. Secondly, we analyze the HSA convergence, in comparison with ILS, taking into account the cost values found at the 1e+05, 3e+05, 5e+05, 10e+05, and 15e+05 EPANET calls (evaluations), as shown in Figure 2. Besides, the HSA's execution times (in seconds) to carry out the maximum number of evaluations for each test case, grouped by network, are shown in Figure 3. Note that, we only present the HSA's total execution time for all test case, because no data about this metric are reported in [18].

Analyzing the Table 4, we detect that HSA found solutions with less TIC than the ones of ILS in nine networks. As a consequence, in 60% of the problem instances better cost values are found when they are solved by HSA. The HSA advantage arises out of the Boltzmann probability application to accept high TIC values, which allows diversify the search in order to escape from local optima.

From the convergence point of view, we observe that HSA found solutions with TIC near to the best ones in 80% of the instances, with only 1e+05 evaluations. Instead, ILS needed at least 3e+05 evaluations for that, besides this is achieved in only 66% of the test cases.

Evaluating both the Figure 3 and the Table 2 together, we notice that the HSA's execution time is affected by the number of pipes and demand nodes. In this way, five groups of three networks can be formed exhibiting similar execution times. These instances have consecutive numbers, e.g. the set of the HG-MP-1, 2, and 3 networks have similar number of pipes and demand nodes, and so on. Being the set formed by the HG-MP-13, 14, and 15 networks the most expensive cases to solve. Furthermore, analyzing what happened into each set of networks, the network with more demand nodes consumes less execution time than the other two, since more feasible solutions exists and HSA need less time to find one of them.

Summarizing, HSA outperformed ILS in the result quality of 60% of the problem instances. Moreover, a quick convergence to good solutions is also evidenced

by our proposal in most of the problem instances. Furthermore, the HSA's runtime is affected by the growing and combination of the number of pipes and demand nodes.

5 Conclusions

In this paper, we have proposed an optimization technique based on the Simulated Annealing algorithm that can find successful water distribution network designs, considering multi-period settings with time varying demand patterns. This new technique, called HSA, solves the hydraulic equations by using the EPANET 2.0 toolkit. HSA combines an WDND-adapted SA with the MP-GRASP local search [18]. For this study, we have tested 75 instances that come from 15 different HydroGen networks.

The HSA's results were compared with the obtained by the ILS proposed in [18] to solve this problem. As a consequence, we observed that HSA outperformed the results obtained by ILS in more than half (60%) of instances, since HSA achieved a better exploration than ILS by using the Boltzmann probability to accept new solutions. This advantage combined with the local search allowed HSA to converge quickly on the best solutions.

For future works, we will be tackling the multi-period WDND optimization problem improving the HSA by introducing a specific heuristic into the perturbation operator. We are also interested in testing larger dimension instances, as close as possible to real scenarios.

Acknowledgments

The authors acknowledge the support of Universidad Nacional de La Pampa and the Incentive Program from MINCyT. The second author is also funded by CONICET.

References

1. D. F. Yates, A. B. Templeman, and T. B. Boffey, "The computational complexity of the problem of determining least capital cost designs for water supply networks," *Engineering Optimization*, vol. 7, no. 2, pp. 143–155, 1984.
2. A. Alperovits and U. Shamir, "Design of optimal water distribution systems," *Water Resources Research*, vol. 13, no. 6, pp. 885–900, 1977.
3. A. Kessler and U. Shamir, "Analysis of the linear programming gradient method for optimal design of water supply networks," *Water Resources Research*, vol. 25, no. 7, pp. 1469–1480, 1989.
4. O. Fujiwara and D. Khang, "A two-phase decomposition method for optimal design of looped water distribution networks," *Water Resources Research*, vol. 26, no. 4, pp. 539–549, 1990.
5. N. Duan, L. W. Mays, and K. E. Lansey, "Optimal reliability-based design of pumping and distribution systems," *Journal of Hydraulic Engineering*, vol. 116, no. 2, pp. 249–268, 1990.

6. G. Loganathan, J. Greene, and T. Ahn, "Design heuristic for globally minimum cost water-distribution systems," *Journal of Water Resources Planning and Management*, vol. 121, no. 2, pp. 182–192, 1995.
7. M. d. C. Cunha and J. Sousa, "Hydraulic infrastructures design using simulated annealing," *Journal of Infrastructure Systems*, vol. 7, no. 1, pp. 32–39, 2001.
8. M. da Conceicao Cunha and L. Ribeiro, "Tabu search algorithms for water network optimization," *European Journal of Operational Research*, vol. 157, no. 3, pp. 746–758, 2004.
9. H. R. Maier, A. R. Simpson, A. C. Zecchin, W. K. Foong, K. Y. Phang, H. Y. Seah, and C. L. Tan, "Ant colony optimization for design of water distribution systems," *Journal of Water Resources Planning and Management*, vol. 129, no. 3, pp. 200–209, 2003.
10. A. C. Zecchin, A. R. Simpson, H. R. Maier, and J. B. Nixon, "Parametric study for an ant algorithm applied to water distribution system optimization," *IEEE Transactions on Evolutionary Computation*, vol. 9, no. 2, pp. 175–191, 2005.
11. G. C. Dandy, A. R. Simpson, and L. J. Murphy, "An improved genetic algorithm for pipe network optimization," *Water Resources Research*, vol. 32, no. 2, pp. 449–458, 1996.
12. I. Gupta, A. Gupta, and P. Khanna, "Genetic algorithm for optimization of water distribution systems," *Environmental Modelling & Software*, vol. 14, no. 5, pp. 437–446, 1999.
13. W. Bi, G. C. Dandy, and H. R. Maier, "Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge," *Environmental Modelling & Software*, vol. 69, pp. 370–381, 2015.
14. M.-D. Lin, Y.-H. Liu, G.-F. Liu, and C.-W. Chu, "Scatter search heuristic for least-cost design of water distribution networks," *Engineering Optimization*, vol. 39, no. 7, pp. 857–876, 2007.
15. A. Vasan and S. P. Simonovic, "Optimization of water distribution network design using differential evolution," *Journal of Water Resources Planning and Management*, vol. 136, no. 2, pp. 279–287, 2010.
16. R. Farmani, G. A. Walters, and D. A. Savic, "Trade-off between total cost and reliability for anytown water distribution network," *Journal of Water Resources Planning and Management*, vol. 131, no. 3, pp. 161–171, 2005.
17. C. Bragalli, C. D'Ambrosio, J. Lee, A. Lodi, and P. Toth, "On the optimal design of water distribution networks: a practical MINLP approach," *Optimization and Engineering*, vol. 13, no. 2, pp. 219–246, 2012.
18. A. De Corte and K. Sörensen, "An iterated local search algorithm for water distribution network design optimization," *Network*, vol. 67, no. 3, pp. 187–198, May 2016.
19. L. A. Rossman, *The EPANET Programmer's Toolkit for Analysis of Water Distribution Systems*, 1999.
20. S. Kirkpatrick, C. G. Jr, and M. Vecchi, "Optimization by simulated annealing," *Science*, no. 220, pp. 671–680, 1983.
21. A. De Corte and K. Sörensen, "Hydrogen," available online: http://antor.uantwerpen.be/hydrogen (accessed on 27 June 2018).

Filtrado basado en contenido para artículos académicos en repositorios institucionales

José Federico Medrano[1]

Facultad de Ingeniería - Universidad Nacional de Jujuy, Jujuy 4600, Argentina
jfedericomedrano@gmail.com

Resumen La mayor parte del tiempo, los investigadores deben filtrar varios documentos académicos para encontrar aquellos relevantes para su investigación. Este filtrado muchas veces es engorroso y requiere emplear una considerable cantidad de tiempo. En la búsqueda de este tipo de material resulta útil contar con un listado de objetos relacionados no sólo con la temática buscada, sino también material que pueda estar relacionado semánticamente con el objeto de la búsqueda. Sería deseable contar con este tipo de funcionalidad en los repositorios institucionales, por esta razón en este trabajo se realiza una comparación de técnicas de filtrado basadas en el contenido semántico (*Term Frequency – Inverse Document Frequency* (TF-IDF), *Latent Semantic Indexing/Latent Semantic Analysis* (LSI/LSA) y *Word Mover's Distance* (WMD)) de una búsqueda realizada, como una propuesta para un sistema de recomendación de material académico empleando como conjunto de datos los registros almacenados en un repositorio institucional. Los resultados propuestos por dichas técnicas fueron evaluados de forma manual y comparados contra los resultados arrojados por el propio motor de búsqueda del repositorio y por los resultados entregados por *Google Scholar*. El esquema planteado mejora en gran medida los resultados actuales.

Palabras Clave: filtrado basado en contenido, PLN, sistema de recomendación, similitud semántica

1. Introducción

Un sistema de recomendación se define como una estrategia de toma de decisiones para los usuarios en entornos de información complejos [27]. Este tipo de sistemas se presentan como técnicas y herramientas de software que proporcionan sugerencias de "artículos" que estiman serán de utilidad para un usuario [4,29]. Las sugerencias están relacionadas con varios procesos de toma de decisiones, tales como qué artículos comprar, qué música escuchar o qué noticias leer. Un "artículo" es el término general utilizado para indicar lo que el sistema recomienda a los usuarios. Los sistemas de recomendación se dirigen principalmente a personas que carecen de suficiente experiencia personal o competencia para evaluar el número potencialmente abrumador de artículos alternativos que por ejemplo un sitio web puede ofrecer [30].

Un tema que resulta de mucho interés son las recomendaciones que ofrecen los motores y bases de datos académicas de acceso libre como *Google Scholar*[1], *Microsoft Academic*[2] y *ResearchGate*[3]. En estas bases de datos los elementos sugeridos son artículos o material académico, basados mayormente en las relaciones que posee el contenido buscado con el material almacenado, aquí entran en juego cuestiones como la popularidad de los elementos, tal es el caso de la aplicación del algoritmo *PageRank* de *Google* [21] para los resultados de las búsquedas y posteriores recomendaciones, o la utilización de mecanismos como *word embedding* [18], el nuevo esquema utilizado por *Microsoft Academic* para ofrecer una lista de publicaciones relacionadas [17].

Existen otros elementos de difusión y preservación donde se aloja material académico, estos son los Repositorios Institucionales (RI), este tipo de bases de datos han sido estudiadas [16,20,15] por la relevancia que poseen como elementos para organizar, gestionar, difundir, preservar y ofrecer acceso libre a la producción científica de una comunidad. Los RI poseen un motor de búsqueda interno, el cual ofrece un conjunto de resultados a partir de una búsqueda por palabra o frase, autor o fecha; buscando dentro de todas las colecciones o una colección en particular. Para los RI que utilizan como software de base a *DSpace*, el motor de búsqueda, dependiendo de la versión, es *Lucene* o *SOLR* o combinaciones de ambos. Estos motores permiten un gran número de opciones al momento de realizar la búsqueda [25], pero las búsquedas sencillas, por palabra o frase, serán las analizadas en este trabajo.

Una vez ingresada una búsqueda en un RI, los resultados pueden estar organizados por relevancia, título, fecha de publicación entre otros criterios. Las palabras ingresadas en la opción de búsqueda son buscadas en los campos título, autores, resumen, series, sponsor e identificador de cada registro. Estas búsquedas se transforman en consultas a una base de datos donde los términos introducidos son cotejados con los campos de los metadatos almacenados, el motor devuelve un conjunto de resultados donde la palabra o frase buscada figura en algún o algunos de los campos mencionados. Más allá de la existencia de estos términos en los registros, los resultados no presentan algún tipo de relación más estrecha, por ejemplo relaciones semánticas de contenido, es más al ingresar a alguno de los resultados de esta búsqueda, el RI no ofrece un listado de registros relacionados.

En este sentido, el presente trabajo aborda la problemática de evaluar y proponer un mecanismo que permita a un RI ofrecer publicaciones relacionadas semánticamente en base a una búsqueda realizada, de este modo, el usuario podría seguir navegando entre los resultados asumiendo que las sugerencias entregadas por la aplicación están realmente relacionadas con el objeto de su búsqueda.

El trabajo presentado se estructura de la siguiente manera, la Sección **2** realiza una breve introducción a los tipos de sistemas de recomendación más comunes

[1] https://scholar.google.com
[2] https://academic.microsoft.com/
[3] https://www.researchgate.net/

y sus aplicaciones, la Sección **3** presenta la definición formal del problema que se intenta resolver, la Sección **4** expone una explicación detallada del método de recomendación basado en contenido que se propone, por su parte en la Sección **5** se presentan los distintos experimentos realizados con el modelo desarrollado, la Sección **6** ofrece una breve validación de los resultados ofrecidos y por último la Sección **7** expone las conclusiones alcanzadas.

2. Trabajos relacionados

Existen dos modelos muy populares de sistemas de recomendación, los *sistemas de filtrado basados en el contenido* y los *sistemas basados en filtrado colaborativo*. Los primeros son los más populares [30], donde el contenido desempeña un papel principal en el proceso de recomendación, en el que las calificaciones de los usuarios y las descripciones de los atributos de los elementos se aprovechan para hacer predicciones. La idea básica es que los intereses del usuario se puedan modelar sobre la base de las propiedades (o atributos) de los elementos que han calificado o accedido en el pasado. Los "elementos" suelen ser textuales, por ejemplo, correos electrónicos [22] o páginas web [32]. La "interacción" generalmente se establece mediante acciones, como descargar, comprar, crear o etiquetar un artículo. Los elementos están representados por un modelo de contenido que contiene las características de los elementos. Las características suelen estar basadas en palabras, es decir, palabras sueltas, frases o n-*grams* [3].

Los sistemas basados en *filtrado colaborativo* [30,31] se refieren al uso de calificaciones de múltiples usuarios de forma colaborativa para predecir las calificaciones faltantes. El funcionamiento de este tipo de sistemas radica en la teoría que a los usuarios les gusta lo que les gusta a los que piensan de forma similar, donde dos usuarios se consideraron con ideas afines cuando calificaron elementos por igual. Cuando se identificaron usuarios de ideas afines, los elementos que un usuario calificó positivamente se recomendaron al otro usuario, y viceversa. Comparado con los sistemas de filtrado basados en contenido, los sistemas de filtrado colaborativo ofrecen tres ventajas [3]. Primero, el filtrado colaborativo es independiente del contenido, es decir, no se requiere procesamiento de elementos que puede estar propenso a error [23,31]. En segundo lugar, debido a que los humanos hacen las clasificaciones, el filtrado colaborativo toma en cuenta las evaluaciones de calidad reales [7]. Finalmente, se supone que el filtrado colaborativo debe proporcionar recomendaciones fortuitas porque las recomendaciones no se basan en la similitud de los ítems, sino en la similitud del usuario [23,14]. Esto ilustra uno de los principales problemas del filtrado colaborativo: se requiere la participación del usuario, pero a menudo la motivación para participar es baja, esta situación se conoce como el problema de "arranque en frío" (*coldstart*).

Algunas de las aproximaciones relacionadas con sistemas de recomendación de contenido académico o relacionado a la educación están bien clasificadas y resumidas en [3], los autores mencionan algunos enfoques como sistemas de recomendación de libros, sistemas de recomendación educativa, servicios de alerta

académica, búsqueda de expertos, resumen automático de artículos académicos, recomendadores de conjuntos de datos académicos y detección de plagio. En el mismo sentido, otras de las aproximaciones es el enfoque de [33] el cual propone un sistema de recomendaciones para artículos académicos que combina el análisis de citas y el análisis de redes. Por otro lado, [9] plantea un sistema de recomendación basado en un esquema híbrido que emplea redes de coautoría y técnicas basadas en contenido. Por su parte [34] presenta un sistema de recomendación entre dominios con transferencia de información coherente (CIT). En [26] se presenta una técnica llamada AKR (*Author-specified Keywords based Retrieval* - recuperación basada en palabras claves especificadas por el autor), la misma proporciona un conjunto diverso de publicaciones como parte de la lista de lectura inicial en base a las palabras claves que el autor especifica.

3. Definición del problema

Se define el problema de la siguiente manera:

Dado como entrada un artículo p de un área específica del conocimiento A que posee un conjunto C de artículos, encontrar un conjunto P de artículos que sean los más relacionados a p. Aquí, cada artículo p_i perteneciente a cualquier C_j está representado por un conjunto de metadatos M.

Como se trabajará sobre un Repositorio Institucional, los metadatos de los registros almacenados serán los datos de entrada. Para este estudio, la relación que existe entre el artículo p y el conjunto P tiene que ver con el contenido semántico, es decir, estos registros estarán relacionados semánticamente con la búsqueda realizada.

4. Enfoque propuesto

Los RI de acceso abierto que implementan el protocolo OAI-PMH[11], respetan el formato de metadatos *Dublin Core*[4] de manera obligatoria, por esta razón cada registro posee 15 campos de metadatos y 3 de encabezado. Los campos utilizados para este estudio son los *títulos* (campo *title*) y las *descripciones* (campo *description*, empleado para almacenar el resumen del elemento digital) por ser los más representativos de cada registro. Dentro de los metadatos del registro no se encuentra el cuerpo del artículo sino un enlace al mismo, debido a esto queda fuera del alcance del esquema propuesto evaluar el contenido del cuerpo del artículo. Los datos sobre los cuales se trabajó fueron recolectados en un trabajo previo [16], por este motivo los datos almacenados respetan el formato antes mencionado.

Según el estudio de [13], las palabras que ocupan diferentes posiciones en un documento tienen diferentes poderes discriminatorios, por ejemplo, una palabra

[4] http://dublincore.org/

que aparece en el título suele ser más significativa que una palabra que aparece en el texto del cuerpo. Una aplicación de esta idea fue conducida por [19] aplicando una ponderación arbitraria tres veces más fuerte a los términos del título que a los términos del cuerpo del texto, y dos veces más fuerte para los términos del resumen. Mientras que [8] experimentó con diferentes pesos para el contenido de artículos y el contexto de citas. Descubrieron que un peso igual para ambos campos alcanzaba la mayor precisión (en este caso la medida de precisión estaba relacionada con la medida de coincidencia). Atendiendo a estas consideraciones, se decidió llevar a cabo esta investigación aplicando el mismo peso para ambos campos de la tupla {*titulo, descripción*}.

Para este trabajo se decidió utilizar la librería *gensim* [28] del lenguaje *Python*[5]. Los esquemas de evaluación seleccionados para el desarrollo de los experimentos fueron: *TF-IDF* [1], *Latent Semantic Indexing/Latent Semantic Analysis* (LSI/LSA) [12,6] y *Word Mover's Distance* (WMD) [10,18].

La solución planteada funciona de la siguiente manera: Inicialmente se descartan los registros que no poseen título y descripción, todos los registros poseen un campo/metadato *subject* el cual, por lo común alberga detalles del área de conocimiento del registro, razón por la cual este campo es utilizado para dividir el conjunto total en subconjuntos más manejables. Cada subconjunto posee un conjunto de tuplas {título, descripción}, estas tuplas son procesadas para eliminar signos de puntuación, pasar las palabras a minúsculas y eliminar las *stop words*, estas *stop words* fueron obtenidas a partir de la librería *nltk*[6]. Seguido a esto, los registros fueron *tokenizados* para iniciar la labor de construcción del diccionario de términos, el cual será utilizado por las técnicas de representación de documentos seleccionadas (TF-IDF, LSI/LSA y WMD) que se encargarán de evaluar la similitud del artículo seleccionado contra el corpus procesado. Cada modelo de evaluación entregará los 10 registros del corpus que posean las mejores evaluaciones de similitud, es decir, los registros más parecidos evaluando los términos del título y descripción del artículo seleccionado. En este punto el modelo de evaluación ya ha sido entrenado con los subconjuntos de datos (división por área del conocimiento) y está listo para realizar predicciones.

En el modo normal de funcionamiento, el usuario realizaría una búsqueda inicial con algún término o frase, el motor del repositorio entregaría los resultados de la búsqueda, y una vez que el usuario elige un registro del listado resultante (este registro es preprocesado y tokenizado al igual que el conjunto de registros), el esquema aquí propuesto, realizaría la búsqueda del material relacionado con el *titulo + descripción* del artículo elegido. Se espera obtener una mejora en las búsquedas ya que es más probable, que un registro esté relacionado semánticamente con otros registros por un número mayor de palabras que si el esquema estuviera limitado solo a los títulos de las publicaciones.

[5] https://www.python.org/
[6] https://www.nltk.org/

5. Experimentos realizados

Los datos sobre los cuales se basaron los experimentos son los registros del Servicio de Difusión de la Creación Intelectual (SEDICI) [5] que es el repositorio institucional de la Universidad Nacional de La Plata[7], el año pasado, este repositorio ocupaba la posición 77°en el ranking mundial de repositorios web, 5°en Latinoamérica y 1°en Argentina [24]. El repositorio en ese entonces superaba los 54.000 registros.

Se dividió el conjunto total de registros del repositorio en áreas temáticas para que el procesamiento y tratamiento de la información sea manejable debido al gran número de registros del repositorio. Para realizar los experimentos se recuperaron los registros del área "Ciencias Informáticas", es decir, los registros que poseen dicha frase dentro del campo *subject*, esto dio un total de 9482 registros cuyos campos *título* y *descripción* son no vacíos. Se eligió esta área de conocimiento debido a la facilidad de contar con expertos para la evaluación de los resultados. Este conjunto de registros fue procesado y tokenizado para crear el diccionario de términos que luego fue utilizado para generar cada uno de los tres modelos de recomendación según las técnicas empleadas (TF-IDF, LSI/LSA y WMD).

El experimento se inició seleccionando al azar 8 artículos del área *Ciencias Informáticas*, estos artículos no formaron parte del cuerpo total de 9482. Cada título completo de estos registros fue introducido en la opción de búsqueda del SEDICI para obtener los primeros 50 resultados ordenados por relevancia. Ocho profesionales informáticos graduados participaron del experimento del siguiente modo: (1) a cada sujeto se le asignó uno de los artículos elegidos al azar y el listado correspondiente de 50 resultados arrojados por el motor de búsqueda del repositorio, cada resultado, incluido el artículo de la búsqueda, fue provisto con el título y la descripción, (2) se le solicitó que evaluarán cada uno de los 50 registros como *relacionado* o *no relacionado* y conformaran una lista ordenada de elementos recomendados de mayor a menor relevancia, (3) de cada lista se obtuvieron los 10 primeros registros para conformar la lista dorada (*gold list*) de cada uno de los 8 artículos iniciales.

De igual manera, por cada uno de los 8 artículos, se realizó una búsqueda en *Google Scholar (GS)* primero con el título del artículo completo e indicando que los recuperara del sitio web del SEDICI ingresando el comodín: **site:sedici.unlp.edu.ar**, luego se realizaron consultas empleando la técnica n-*grams*. Con esta estrategia se generan consultas extrayendo subsecuencias de n palabras del título del artículo, para los experimentos se tomó $n = 2$. Antes de aplicar dicha estrategia se preprocesó el texto del título, removiendo caracteres especiales, removiendo *stop words* y pasando los términos a minúsculas. Para todos los casos, del listado de resultados y ordenados por relevancia, se tomaron los primeros 10 resultados para evaluarlos contra la lista dorada generada por los expertos.

[7] http://sedici.unlp.edu.ar/

Por último, se emplearon cada uno de los modelos generados con las técnicas TF-IDF, LSI/LSA y WMD, ingresando por un lado los títulos de los registros, en otros casos solo la descripción y por último la tupla {título, descripción} de cada uno de los 8 registros para computar la similitud semántica entre estos y el corpus de evaluación generado previamente. Las consultas generadas con la estrategia n-*grams* sobre los términos del título de los artículos que fueron aplicadas para recuperar los registros de GS, fueron igualmente aplicadas al motor de búsqueda del SEDICI y en los modelos de evaluación TF-IDF, LSI/LSA y WMD. La idea de utilizar este tipo de técnicas es útil para comprobar si puede o no existir algún sesgo al momento de consultar con el título completo o solo con algunas palabras.

6. Validación

Para evaluar los resultados de la solución propuesta contra los resultados entregados por el motor de búsqueda del SEDICI y los entregados por *Google Scholar*, se adoptó la métrica de evaluación *recall* [2]. La sensibilidad o exhaustividad (*recall*) es la fracción de instancias o elementos relevantes recuperados por una consulta:

$$Recall = \frac{Elementos\ Relevantes\ Recuperados}{Total\ de\ Elementos\ Recuperados} \tag{1}$$

A continuación se presentan y analizan los resultados obtenidos con los experimentos llevados a cabo. Para evaluar comparativamente las alternativas mencionadas, se decidió de forma arbitraria tomar los primeros 10 resultados como el total de elementos recuperados en cada consulta realizada tanto al motor de búsqueda del repositorio, a GS y con los modelos de evaluación generados. Dentro de los primeros 10 resultados se computarán los elementos relevantes recuperados por cada alternativa. Se toma este criterio principalmente porque el usuario común tiende a buscar contenido relacionado entre los primeros resultados de una búsqueda, por ello se considera que 10 es un número razonable, asumiendo además que el usuario solo leerá los títulos y resúmenes de dichos resultados y no el cuerpo completo del documento. La Tabla 1 reúne los resultados promedios de

Tabla 1. Valores relativos de *recall* para las distintas alternativas planteadas

	SEDICI	GS	TF-IDF	LSI/LSA	WMD
Título	0.35 (0.8)	0.41 (0.5)	0.46 (1.15)	0.31 (0.73)	0.48 (0.81)
Título (2-*grams*)	0.32 (0.8)	0.39 (0.5)	0.41 (1.10)	0.28 (0.83)	0.48 (0.93)
Descripción	–	–	0.61 (1.28)	0.38 (0.85)	0.55 (0.85)
Título + Descripción	–	–	0.65 (1.31)	0.41 (0.89)	0.57 (0.85)

exhaustividad (*recall*) calculados para cada una de las alternativas empleadas,

entre paréntesis se incluye el tiempo promedio empleado en segundos para cada experimento. Estos experimentos fueron llevados a cabo realizando consultas a los distinto motores (SEDICI y GS) y modelos generados (TF-IDF, LSI/LSA y WMD) consultando por el título completo del artículo (primera fila), empleando 2-*grams* con los términos del título (segunda fila), tomando los términos de la descripción del artículo (tercera fila) y tomando tantos los términos del título como los de la descripción (cuarta fila). A simple vista, el esquema que mejor resultados ofrece es el TF-IDF empleando en conjunto el título y la descripción con un valor de 0.65 de *recall*. El segundo esquema que ha entregado buenos resultados es el WMD con un valor de 0.57 de *recall*. En las filas tercera y cuarta, para el SEDICI y GS no existe un valor ya que para estos motores sólo se ha consultado por el título del artículo sin emplear los términos de la descripción.

Al emplear la estrategia n-*grams* no se obtuvieron diferencias sustanciales en relación a los otros esquemas, esto quiere decir que para realizar consultas empleando solamente palabras del título, basta ingresar dos palabras del mismo para obtener resultados aceptables en lugar de emplear el título completo. Además hay que destacar que el empleo de las *descripciones* de los registros, mejoró en todos los resultados en todos los casos, esto es un claro indicio de las relaciones semánticas entre los documentos recuperados ya que las descripciones representan un resumen del cuerpo del artículo.

7. Conclusiones

En este trabajo se presentó una solución preliminar a un problema específico y real, poder recomendar contenido académicos semánticamente relacionado en un repositorio institucional. Si bien, hasta el momento no se intentó resolver este problema en este tipo de herramientas, el modelo aquí presentado otorga una alternativa viable y una mejora sustancial a la funcionalidad ofrecida por el repositorio. La aplicación de los algoritmos seleccionados han mostrado resultados muy alentadores, pues las relaciones entre documentos aquí encontradas e inferidas, sería una tarea realmente compleja de resolver con simples consultas a una base de datos.

La aplicación de técnicas de Inteligencia Artificial, y más específicamente, técnicas de Procesamiento del Lenguaje Natural al problema presentado, demuestran el gran potencial de estas para resolver problemas de cualquier índole y la flexibilidad de adaptarse a nuevos entornos, tal es el caso del esquema TF-IDF que en este caso entregó los mejores resultados.

La búsqueda de contenidos académicos relacionados es un tema muy importante y un instrumento valioso para la comunidad científica, además, teniendo en cuenta la importancia y la relevancia que tiene el SEDICI como un repositorio referente a nivel nacional y latinoamericano, contar con herramientas y/o servicios que mejoren sus prestaciones le aportan un gran valor institucional.

El estudio aquí presentado deja abierto el panorama para un sin fin de aplicaciones y futuras líneas de investigación. La combinación de técnicas en conjunto con las aquí presentadas puede resultar en algo más fructífero, sin ir más lejos, el

empleo de palabras claves del mismo documento o la combinación de esquemas basados en la recuperación de documentos referenciados, o el empleo de redes de coautoría, o la evaluación del cuerpo del documento o empleando un esquema de ponderación para las palabras del título y descripción pueden llegar a otorgar una mejora considerable.

Referencias

1. Aizawa, A.: An information-theoretic perspective of tf–idf measures. Information Processing & Management **39**(1), 45–65 (2003)
2. Baeza-Yates, R., Ribeiro-Neto, B., et al.: Modern information retrieval, vol. 463. ACM press New York (1999)
3. Beel, J., Gipp, B., Langer, S., Breitinger, C.: Research-paper recommender systems : a literature survey. International Journal on Digital Libraries **17**(4), 305–338 (2016). https://doi.org/10.1007/s00799-015-0156-0
4. Burke, R.: The adaptive web. chap. Hybrid Web Recommender Systems, pp. 377–408. Springer-Verlag, Berlin, Heidelberg (2007)
5. De Giusti, M.R., Oviedo, N., Lira, A.J., Sobrado, A., Martínez, J.P., Pinto, A.V.: Sedici-desafíos y experiencias en la vida de un repositorio digital. In: Conferencia sobre Bibliotecas y Repositorios Digitales (Colombia, 2011). vol. 1 (2011)
6. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American society for information science **41**(6), 391 (1990)
7. Dong, R., Tokarchuk, L., Ma, A.: Digging friendship: paper recommendation in social network. In: Proceedings of Networking & Electronic Commerce Research Conference (NAEC 2009). pp. 21–28 (2009)
8. Huang, S., Xue, G.R., Zhang, B.Y., Chen, Z., Yu, Y., Ma, W.Y.: Tssp: A reinforcement algorithm to find related papers. In: Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence. pp. 117–123. IEEE Computer Society (2004)
9. Hwang, S.Y., Wei, C.P., Liao, Y.F.: Coauthorship networks and academic literature recommendation. Electronic Commerce Research and Applications **9**(4), 323 – 334 (2010), special Section: Social Networks and Web 2.0
10. Kusner, M., Sun, Y., Kolkin, N., Weinberger, K.: From word embeddings to document distances. In: International Conference on Machine Learning. pp. 957–966 (2015)
11. Lagoze, C., Van de Sompel, H.: The open archives initiative: Building a low-barrier interoperability framework. In: Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries. pp. 54–62. ACM (2001)
12. Landauer, T.K., Dumais, S.T.: A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological review **104**(2), 211 (1997)
13. Manning, C.D., Raghavan, P.: Introduction to Information Retrieval. Cambridge University Press. (2008)
14. McNee, S.M., Kapoor, N., Konstan, J.A.: Don't look stupid: avoiding pitfalls when recommending research papers. In: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work. pp. 171–180. ACM (2006)
15. Medrano, J.F., Figuerola, C.G., Berrocal, J.L.A.: Repositorios digitales en españa y calidad de metadatos. Scire: representación y organización del conocimiento **18**(2), 109–121 (2012)

16. Medrano, J.F.: Calidad en repositorios digitales en argentina, estudio comparativo y cualitativo. In: VII Conferencia Internacional BIREDIAL-ISTEC'17 y XII SIBD (La Plata, 2017) (2017)
17. MicrosoftResearch: Feature improvement: Related publications (2018), https://tinyurl.com/y9x8ptoc, consultado el: 2018-03-25
18. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. Neural Information Processing Systems pp. 3111–3119 (2013)
19. Nascimento, C., Laender, A.H., da Silva, A.S., Gonçalves, M.A.: A source independent framework for research paper recommendation. In: Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries. pp. 297–306. ACM (2011)
20. Orduña-Malea, E.: Impacto de los repositorios a través de técnicas cibermétricas: el caso general de latinoamérica y especial de costa rica. In: III Conferencia Bibliotecas y repositorios digitales de América Latina (BIREDIAL) (2013)
21. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford InfoLab (1999)
22. Paik, W., Yilmazel, S., Brown, E., Poulin, M., Dubon, S., Amice, C.: Applying natural language processing (nlp) based metadata extraction to automatically acquire user preferences. In: Proceedings of the 1st international conference on Knowledge capture. pp. 116–122. ACM (2001)
23. Palopoli, L., Rosaci, D., Sarné, G.M.: A multi-tiered recommender system architecture for supporting e-commerce. In: Intelligent Distributed Computing VI, pp. 71–81. Springer (2013)
24. Pinto, A.: Sedici en el ranking webometrics (2017), https://tinyurl.com/yb4uyqel, consultado el: 2018-03-30
25. Prasad, A., Patel, D.: Lucene search engine: An overview. DRTC-HP International **10** (2005)
26. Raamkumar, A.S., Foo, S., Pang, N.: Using author-specified keywords in building an initial reading list of research papers in scientific paper retrieval and recommender systems. Information Processing & Management **53**(3), 577–594 (2017)
27. Rashid, A.M., Albert, I., Cosley, D., Lam, S.K., McNee, S.M., Konstan, J.A., Riedl, J.: Getting to know you: learning new user preferences in recommender systems. In: Proceedings of the 7th international conference on Intelligent user interfaces. pp. 127–134. ACM (2002)
28. Řehůřek, R., Sojka, P.: Software Framework for Topic Modelling with Large Corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. pp. 45–50. ELRA, Valletta, Malta (May 2010)
29. Resnick, P., Varian, H.R.: Recommender systems. Communications of the ACM **40**(3), 56–58 (1997)
30. Ricci, F., Rokach, L., Shapira, B.: Introduction to recommender systems handbook. In: Recommender systems handbook, pp. 1–35. Springer (2011)
31. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: The adaptive web, pp. 291–324. Springer (2007)
32. Seroussi, Y.: Utilising user texts to improve recommendations. In: International Conference on User Modeling, Adaptation, and Personalization. pp. 403–406. Springer (2010)
33. Son, J., Kim, S.B.: Academic paper recommender system using multilevel simultaneous citation networks. Decision Support Systems **105**, 24 – 33 (2018)
34. Zhang, Q., Wu, D., Lu, J., Liu, F., Zhang, G.: A cross-domain recommender system with consistent information transfer. Decision Support Systems **104**, 49 – 63 (2017)

ToM-Dyna-Q: on the integration of reinforcement learning and machine Theory of Mind

Dan Kröhling, Ernesto Martínez

INGAR (CONICET/UTN)
Avellaneda 3657
Santa Fe, Argentina

Abstract. The capacity to understand others, or to reason about others' ways of reasoning about others (including us), is fundamental for an agent to survive in a multi-agent uncertain environment. This reasoning ability, commonly known as Theory of Mind, is instrumental for making effective predictions over others' future actions and learning from both real and simulated experience. In this work, a novel architecture for model-based reinforcement learning in a multi-agent setting is proposed. The proposed architecture, called ToM-Dyna-Q, integrates ToM simulation alongside with the well-known Dyna-Q architecture to account for artificial cognition in a shared environment inhabited by multiple agents interacting with each other. Results obtained for the two-player competitive game of Tic-Tac-Toe demonstrate the importance for a given agent of learning, reasoning and planning based on mental simulation modeling of other agents' goals, beliefs and intentions.

Keywords: Intelligent agents, prediction machines, reinforcement learning, Theory of Mind.

1 Introduction

From the beginning of mankind, we humans have experienced that it is practically impossible to survive without successfully competing and cooperating with others within a shared environment. The need to collaborate strategically with others has lead to the emergence of complex markets, where each individual must interact with his pairs in order to accomplish his own goals. Through time, this interaction contributed to the arise of increasingly sophisticated social skills that required humans to reason about the world they inhabit, what knowledge and understanding others have, what do they think we know they know and think, and so forth. This particular recursive type of reasoning about others has lately gained attention, and is usually referred to as Theory of Mind (ToM).

Theory of Mind denotes the ability to represent the mental states of others, including their beliefs, intentions and goals [2, 13], which can be used advantageously over shallower ways of reasoning [12]. Scientists have wondered if this

ability may be present not only in humans and other few species of animals [6], but also implemented by machines and rational artificial agents. Actually, it could [10, 11], and a number of recent research efforts have proven ToM effectiveness [3, 4, 14, 19].

As stated in [9], a great variety of methods try to address implementation details of ToM-based artificial agents [3, 8, 19] and that also happens with some learning agents in general [5, 7]. However, to the best of our knowledge, a conceptual architecture for tightly integrating reinforcement learning in a multi-agent setting with simulating ToM reasoning for planning and deliberation is lacking.

Bearing in mind model-based Q-learning in multi-agent systems, the novel ToM-Dyna-Q architecture is proposed. The Q-learning [18] algorithm is used for policy learning based on both actual and simulated experience. However, other reinforcement learning algorithms such as *Sarsa* or Bayesian Q-learning can be used. Q-learning will suffice to keep things simple enough and help us concentrate on answering the questions that drive our computational study: is ToM useful to gain a competitive advantage over other learning agents? Does ToM accelerate the learning curve of more sophisticated reasoning agents? Is an agent's best strategy to play always optimally?

To address the foregoing questions, the simple Tic-Tac-Toe game will be used. A group of agents was created (named Q, R, S, and T), some of them with the opportunity to learn about the game and their opponents, and some of them without that capability. We then provide certain agents the possibility to use different levels of ToM reasoning about the opponent, to test intuitions regarding the role to ToM-based learning in artificial agents. Finally, we set them against each other in tournaments of Tic-Tac-Toe games.

To begin with, a short introduction to previous works on Theory of Mind and model-based Q-learning is given in Section 2. Then, in Section 3, the proposed architecture for a ToM-Dyna-Q agent (Jack) is presented. In Section 4, a number of computational experiments made are discussed together with results obtained for different scenarios. Finally, some concluding remarks about our research efforts in implementing ToM for multi-agent learning are made in Section 5.

2 Previous work

2.1 Theory of Mind

Following the definition given by Singer et al. [13], it can be said that Theory of Mind, or the equivalent term "mentalizing", connotes an agent's ability to cognitively represent the mental states of others, including intentions, knowledge, goals, beliefs, and models. As mentioned earlier, this is not only a human capacity, but also is present in a few animal species and now intelligent artificial agents too [10, 11].

In previous works, it has been proved that deception and manipulation could provide agents a social competitive advantage [4]. Thus, we could state that

ToM, that is, the capacity to infer other agent's thoughts and states as well as effectively predicting their next actions [1], is a natural consequence of learning and deliberation in rational agents in order to effectively interact with others. In Section 4, this ability to understand others (or to reason about other agents' reasoning) in strategic interactions between two players is analyzed. A previous illuminating study could be found also in [17].

Fig. 1 will serve us to explain the concept of ToM. Suppose there are two agents, Jack (our focal agent) and Zack, interacting in a certain environment. As a ToM-0 agent (or T0), Jack makes a model out of the environment but does not take into account any other agent (Zack, for instance) in it. In other words, Jack only sees the effect of Zack's actions over the environment and himself, but could not identify Zack as an entity capable of responding to his actions and acting purposefully. As a T1 agent, Jack models not only the environment as a set of interacting passive objects, but also distinguishes the presence of others as T0 agents. That is, Jack interprets that a Zack thinks as he would do if the environment were agent-free. As could be imagined, a T2 Jack would model a Zack as if he were a T1 agent, and so it continues. In general, a Tm Jack would model a Zack as a T$m-1$ agent.

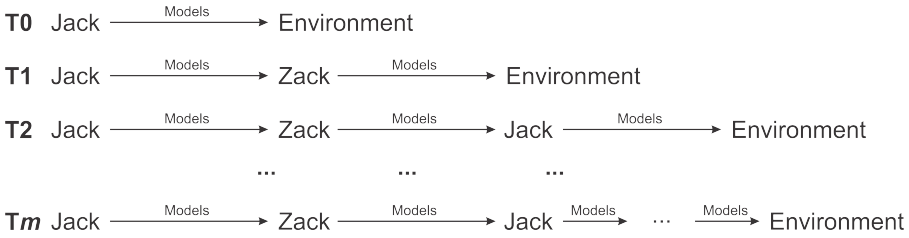


Fig. 1. Theory of Mind explained.

2.2 Dyna-Q (model-based Q-learning)

Q-learning [18] is a model-free algorithm created for learning from a sequence of reinforcements and state transitions, which consists of a function that iterates over the expected cumulative rewards for future time steps in episodic interactions between and agent and its environment. We will explain it succinctly: given a state s_t observed by a Q-learning agent in a certain environment, this algorithm determines the next action a_t from all the agent's possible actions A she can perform following a policy $Q(s, a)$. Then, the environment returns both a reward r, that will be used as a reinforcement signal or hint, and its next state s_{t+1} to the agent. The algorithm takes this information and actualizes the policy $Q(s, a)$ at the end of each episode according to the learning rule:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \left[r_t + \gamma \, max_a \, Q(s_{t+1}, a) - Q(s_t, a_t) \right] \qquad (1)$$

In equation 1 there are two hyper-parameters used to influence the learning curve and the long-term view of action effects on Q-values. Firstly, the learning rate α measures the importance given to prediction errors in Q-values updates. Secondly, the discount rate γ establishes how much our agent will look up into the future for planning and learning. As a common rule of thumb, these parameters are usually set to 0,1 and 0,9, respectively [15]. There is a third parameter, ϵ, that is used to determine the balance between exploration and exploitation of the learning agent based on what she knows and its uncertainty. When ϵ is set near to 0, the algorithm will exploit what it already knows for certain, not exploring for alternative courses of action, even if there may exist better ones. When ϵ is set near to 1, the algorithm will explore the state space as much as it can to reduce uncertainty, but would not consider those rewarding courses of action already found, despite how good they are.

Dyna-Q [15], or model-based Q-learning, is an extension to the mentioned Q-learning algorithm, that incorporates a model of the environment to combine real and simulated experience in Q-values updates. Dyna-Q uses a generative model of the environment and allows the Q-learning agent to plan in advance his actions based on simulating sequences of action-state-rewards. The main assumption behind the generative model is that it does not change in response to the agent's learning process. In Dyna-Q, the agent learns not only its policy but also improves the model of its environment. The Dyna-Q does not apply to a multi-agent setting because, as one agent learns, the generative model also changes due to the response of other agents in its environment. An extension to Dyna-Q is thus needed to account for the effect of one agent learning, reasoning and planning on the other agents in the environment.

3 ToM-Dyna-Q

In this section, we propose to overhaul completely the Dyna-Q architecture so that a learning agent such as Jack could not only model the perceivable changes (the state transitions) in the environment due to its actions, but also to identify and mentally represent other agents reasoning and learning processes. The resulting model-based learning architecture, aptly named ToM-Dyna-Q, makes room for a given agent Jack to elaborate and plan about the knowledge it has over other agents' knowledge, policies, models, goals, etc., including knowledge and models he believes the others think Jack has about them. This novel multi-agent learning and planning framework makes room for creating ToM learning agents with different levels of sophistication. In Fig. 2[1], the main building blocks of the ToM-Dyna-Q learning in a multi-agent environment are shown.

Reinforcement learning using ToM-Dyna-Q allows for a more comprehensive framework for learning a model of an agent's real world (or his reality) through reasoning about other agents and environmental objects separately. Real experience for the agent is now highly informative about other agents' models, beliefs

[1] The original figure was taken from [15] and conveniently adapted.

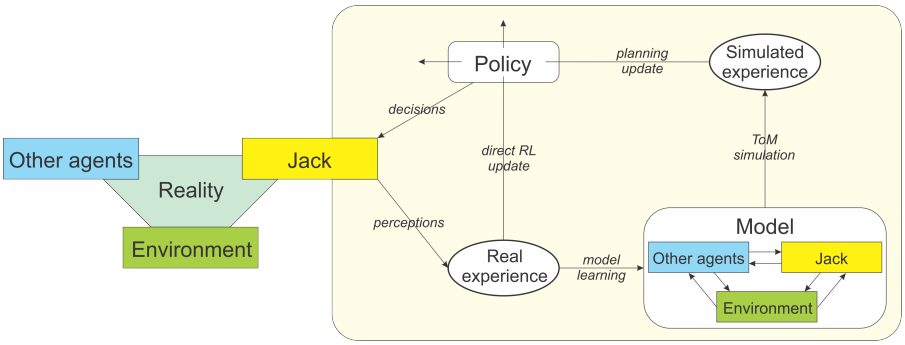


Fig. 2. Tom-Dyna-Q for model-based learning and reasoning in multi-agent systems.

and strategies. Based on ToM simulations, a given agent learns adaptations to his policy which, when implemented in the real world, gives rise to perceptions upon which direct reinforcement learning (RL) is practiced. Thus, by planning and reasoning using a gradually refined model of the multi-agent environment, the policy used while interacting with the actual environment probes other agents responses. When compared with predictions about expected actions, a signal error allows for further elaborations and refinements.

4 Computational experiments

4.1 Setting

In this section, the setting for the computational experiments is detailed. In order to show the advantages that could be obtained from ToM-Dyna-Q, the English version of the Tic-Tac-Toe is used as a representative example. Two agents, Jack and Zack (in the different configurations discussed below), will play against each other in sessions of 100 games to assess about the competitive edge of "mentalizing" the opponent.

The configurations Jack and Zack may adopt are:

- **Q-configuration.** In this configuration, an agent would have learned to play Tic-Tac-Toe using the well-known Q-learning algorithm playing against another instance of himself. Nineteen experience levels where created corresponding to the different number of games that the agent had played, with the 19th level corresponding to a total of 10 million episodes played. In general, we will name Qn an agent of an n level of experience.
- **R-configuration.** An R agent would be an agent that plays completely at random, regardless its opponent's mental state, expertise, or board state.
- **S-configuration.** Five heuristic strategies typically used by human players for Tic-Tac-Toe (e.g., as a start choice, to play always in the board center). From the simplest to the most complicated one, they will be called as Sl, l being their heuristic complexity level.

In Fig. 3, it is shown the different Qn levels obtained for an agent, playing against R and S agents, to highlight later to what extent ToM-based learning may improve the cumulative utilities the agents get.

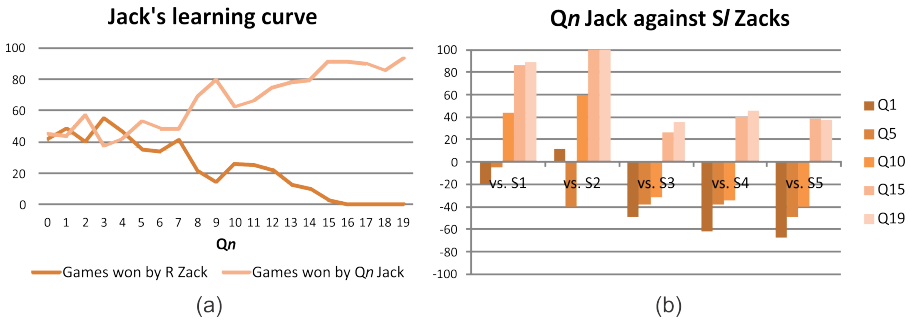


Fig. 3. (a) Amount of games won by a Qn Jack and a R Zack playing against each other in tournaments of 100 games, with Jack's Qn levels varying from 0 to 19. (b) Rewards obtained by five different Qn Jacks against five different Sl Zacks.

– **T-configuration.** ToM agents are identified by Tm, where m is the level of Theory of Mind our agent is going to use, as described earlier in Fig. 1. In this configuration, a Tm Qn agent is one such that it could play as a Qn agent, but may also behave as an agent with lower experience levels $[1; n-1]$ in order to mislead his opponents or resort to a higher ϵ in order to explore more and learn over his opponent knowledge and expertise. The ToM agent will model his opponent as explained later, and then will consider this information to be the state over which he will reason, plan and learn.

The rationale for ToM agents when modeling their opponents is based on the number of games won by a given opponent considering the last ten games. If an opponent has won ten out of the last ten, then he is certainly a tough opponent. Accordingly, the ToM agent will try to learn as much as it can from such an opponent. On the contrary, If the opponent has won zero, then it is easy to beat, and a ToM agent will try to mislead such an easy opponent to slow down its learning process. The demonstration of this seemingly intuitive reasoning is presented in Fig 4. The rewards gathered by any of our agents are based on the difference between the number of games won and the number of games lost, leaving aside the games that end up in a draw.

As can be seen, when the Qn level of Jack is lower than that of Zack, then the better choice for Jack is to explore more by increasing his ϵ. When the Qn level of Jack is almost equal than that of Zack, then for Jack is better just playing as he knows best, in order to win more games. Finally, when the Qn level of Jack is greater than that of Zack, then the better choice for him is to exploit his knowledge, but lowering the ToM level of his strategy, that is to say, picking

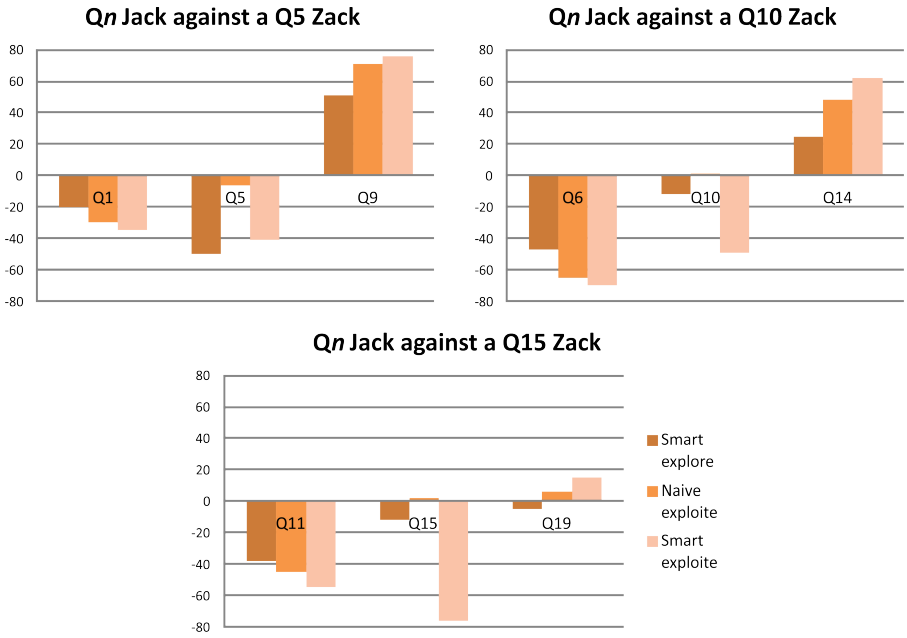


Fig. 4. The intuition that uses our agent to play given the experience he thinks his opponent has. In the vertical axis are presented Jack's rewards when playing against a Q5, a Q10, and a Q15 Zack in tournaments of 100 games.

consciously over his [1;n] available levels in order to prevent the opponent to learn from his own expertise.

4.2 Results and analysis

In this section, results obtained when a Tm Jack is set to play against a T0 Zack are presented. As was previously mentioned, these agents are engaged in playing tournaments of 100 games.

In Fig. 5, the rewards obtained by the agent Jack when playing with different levels of Theory of Mind against a T0 agent are shown. It is noteworthy that, whichever his Qn level of experience is, his best choice is always to play as a T1 agent. This is not a surprise, given that a T0 and a T2 agent construct models that do not correspond to the actual reality, and thus fail to predict Zack's playing strategy.

The same situation could be seen the next two graphics. However, the differences blur a little when the opponent's experience level is significantly high. Such a result may be attributed to the fact that, when Jack's opponent strategy is actually very hard to defeat, it cannot establish a policy sufficiently good to cope with such a rival. On the other hand, 100 games may be too small a sample of games, which does not allow Jack to gain sufficient knowledge to successfully be-

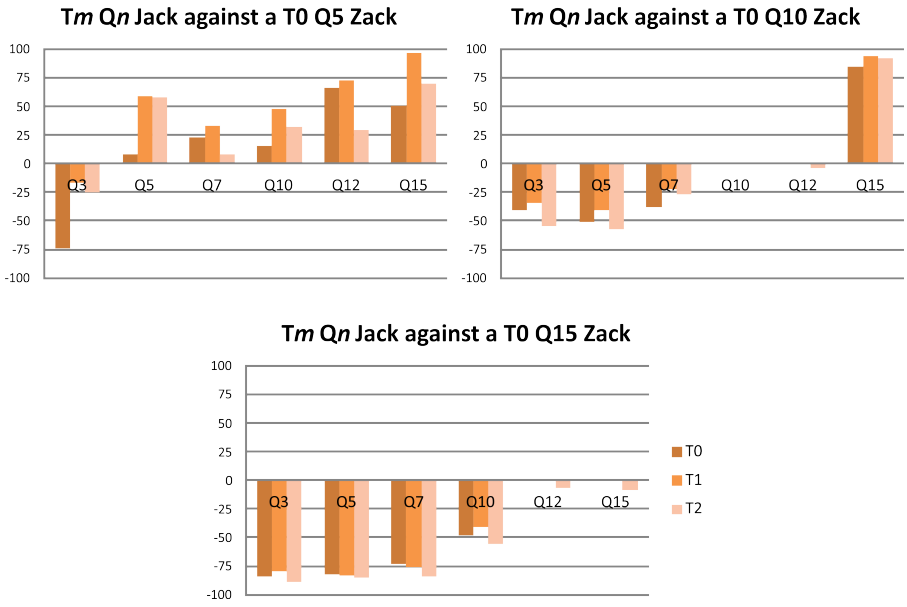


Fig. 5. In the vertical axis are presented Jack's rewards when playing with different levels of sophistication against a T0 Zack with Q5, Q10, and Q15 experience levels.

have in the game. For instance, a Q5 agent has only played two hundred learning episodes in Tic-Tac-Toe before the tournament, visiting only 156 possible states in the board, whereas a Q15 agent has experienced 500000 learning episodes, visiting as many as 3947 of a total of 8744 relevant states in the board, after the elimination of some symmetrical board states.

In Fig. 6 could be seen that a T2 Jack makes the best out of him when he is playing against a T1 Zack, in contrast to what happened in the T0 Zack scenarios. In this case, the model he makes of his opponent is accurate, and allows him to pick the right actions and get better results.

As a summary, it can be stated that, for a given agent, the use of ToM for planning and learning could give a competitive advantage over other learning agents, but should not be overestimated. If an agent fails to identify the depth of the reasoning his opponents resort to, then it may fool itself by taking suboptimal decisions based on too complex reasoning. Even worse, actions taken may be far away from those reasonable courses of action the agent may learn without modeling its opponent at all. The correct use of ToM does accelerate learning, while identifying the proper actions to take (to exploit or explore) based on the experience the opponent has in the environment. Finally, we could say that the best strategy of an agent is not always to play the one he knows is his best play, but to identify the goals, intentions and models other agents have about the environment and agents within it, as well as and the beliefs they think it has about the them, their intentions, goals and strategies.

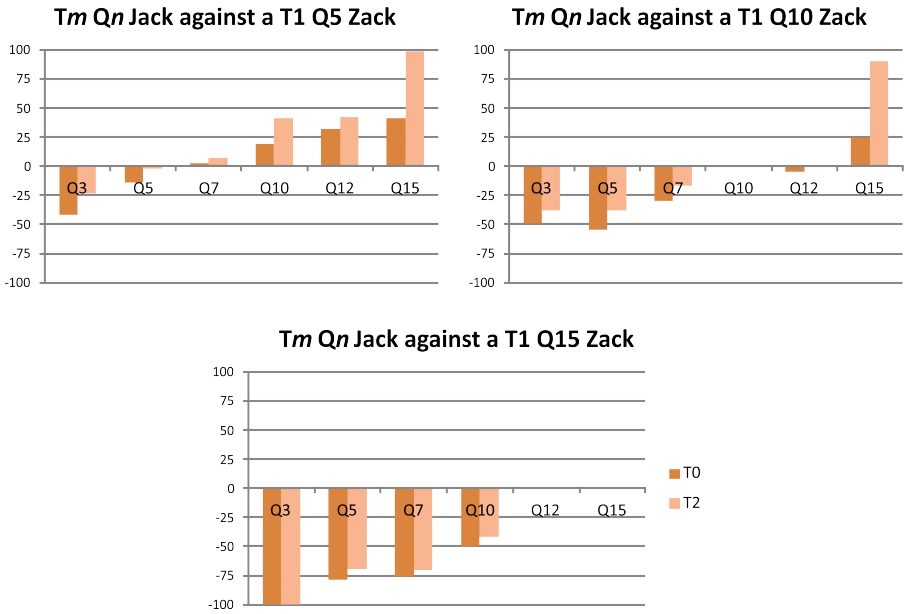


Fig. 6. In the vertical axis are presented Jack's rewards when playing with T0 and T2 against a T1 Zack.

5 Concluding remarks

In this work, a novel architecture for model-based reinforcement learning in a multi-agent environment is proposed. The novel architecture ToM-Dyna-Q may integrate different learning rules and models for accounting about other agents reasoning, planning and learning, including Bayesian Q-learning, actor-critic, eligibility traces [16], etc. Many applications for the Internet of Things could be approached by the proposed architecture. Automated negotiations, autonomous driving, and emergent scheduling, were the main drivers for our proposal.

The English version of the Tic-Tac-Toe game was presented as a toy example, where a proper integration of Theory of Mind with reinforcement learning show promising results. Through a number of experiments with different types of agents proposed, our research questions where addressed, demonstrating that the correct use of ToM by an agent may provide him a competitive advantage over other learning agents that do not reason and plan bearing in mind that other agents may also have a mind on their own.

References

1. A. Agrawal, J. Gans, and A. Goldfarb. *Prediction machines: The simple economics of artificial intelligence.* Harvard Business Review Press, Boston Massachusetts, 2018.

2. C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, 1(4):598, 2017.
3. S. de Jong, D. Hennes, K. Tuyls, and Y. Gal. Meta-strategies in the colored trails game. *Belgian/Netherlands Artificial Intelligence Conference*, (c):2–6, 2011.
4. H. de Weerd, R. Verbrugge, and B. Verheij. How much does it help to know what she knows you know? an agent-based simulation study. *Artificial Intelligence*, 199-200:67–92, 2013.
5. R. Dearden, N. Friedman, and S. Russell. Bayesian q-learning. In *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence*, AAAI '98/IAAI '98, pages 761–768, Menlo Park, CA, USA, 1998. American Association for Artificial Intelligence.
6. J. Joiner, M. Piva, C. Turrin, and S. W. C. Chang. Social learning through prediction error in the brain. *npj Science of Learning*, 2(1):8, 2017.
7. D. S. Leslie and E. J. Collins. Individual q-learning in normal form games. 44(2):495–514, 2005.
8. J. Pöppel and S. Kopp. Satisficing models of bayesian theory of mind for explaining behavior of differently uncertain agents. *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden.*, (July), 2018.
9. S. Qi and S.-C. Zhu. Intent-aware multi-agent reinforcement learning. *CoRR*, abs/1803.02018, 2018.
10. N. C. Rabinowitz, F. Perbet, H. F. Song, C. Zhang, S. M. A. Eslami, and M. Botvinick. Machine Theory of Mind. *ArXiv e-prints*, Feb. 2018.
11. B. T. Revell. Deepmind ai is learning to understand the 'thoughts' of others. *New Scientist*, (February):1–5, 2018.
12. I. Sher, M. Koenig, and A. Rustichini. Children's strategic theory of mind. *Proceedings of the National Academy of Sciences*, 111(37):13307–13312, 2014.
13. T. Singer and A. Tusche. Understanding others: Brain mechanisms of theory of mind and empathy. *Neuroeconomics: Decision Making and the Brain*, pages 513–534, 2014.
14. C. A. Stevens, N. A. Taatgen, and F. Cnossen. Instance-based models of metacognition in the prisoner's dilemma. *Topics in Cognitive Science*, 8(1):322–334, 2016.
15. R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. Adaptive computation and machine learning series. The MIT Press, Cambridge MA, second edition edition, 2018.
16. R. S. Sutton, D. Precup, and S. Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1-2):181–211, 1999.
17. F. B. von der Osten, M. Kirley, and T. Miller. The minds of many: Opponent modelling in a stochastic game. *IJCAI International Joint Conference on Artificial Intelligence*, pages 3845–3851, 2017.
18. C. J. C. H. Watkins. *Learning from Delayed Rewards*. PhD thesis, King's College, Cambridge, UK, 1989.
19. W. Yoshida, R. J. Dolan, and K. J. Friston. Game theory of mind. *PLoS Computational Biology*, 4(12), 2008.

Spectogram Prediction with Neural Networks

Mario Alejandro García ✉ and Eduardo Atilio Destéfanis

Universidad Tecnológica Nacional Facultad Regional Córdoba, Argentina
mgarcia@frc.utn.edu.ar

Abstract. A neural network model for spectrogram magnitude prediction is presented. It has one convolutional layer that computes the short-time Fourier transform. By choosing the magnitude of the spectrum as output and discarding the phase, it is possible to avoid complex number operations. The structure of the network and coefficients computation for this alternative are presented in detail. The model coefficients can be directly computed or trained with the gradient descent algorithm. In both cases, the results are satisfactory, but the obtained weights are different. An analysis of the differences is made. The main contribution of this article is to show that the proposed model is trainable. Consequently, the coefficients can be adapted to particular problems.

Keywords: discrete fourier transform, spectrogram, deep learning, convolutional neural network

1 Introduction

The spectrogram is a representation of the variation of the frequency spectrum of a signal. This variation can occur in time (audio, earthquake waves, etc.), space (images) and other domains.

In machine learning, it is common to use the spectral information data to find non-obvious features in the source domain. The frequency spectrum of a signal is obtained through the Fourier Transform (FT). For discrete data, the computational basis of spectral analysis is the Discrete Fourier Transform (DFT).

In the following paragraphs of this introduction, some key concepts will be stated.

Discrete Fourier Transform. The DFT converts a finite sequence of N complex numbers (samples) $\{x_n\} := x_0, x_1, ..., x_{N-1}$ into a sequence of $K = N$ complex numbers $\{X_k\} := X_0, X_1, ..., X_{N-1}$.

$$X_k = \sum_{n=0}^{N-1} x_n \; e^{-i2\pi kn/N} \tag{1}$$

According to the Euler's formula,

$$X_k = \sum_{n=0}^{N-1} x_n \left[\cos(-2\pi kn/N) + i\ \sin(-2\pi kn/N)\right] \tag{2}$$

It is important to note that the DFT is a linear operator [1]. The DFT can then be defined as the linear map $\mathcal{F} : \mathbb{C}^N \to \mathbb{C}^N$ such that $X = \mathcal{F}(x)$ with the following matrix representation.

$$\boldsymbol{X} = \boldsymbol{F}\boldsymbol{x}$$

where

$$\boldsymbol{x} = \begin{bmatrix} x_0 & x_1 & x_2 & \dots & x_{N-1} \end{bmatrix}^\intercal,$$
$$\boldsymbol{X} = \begin{bmatrix} X_0 & X_1 & X_2 & \dots & X_{N-1} \end{bmatrix}^\intercal$$

and according to equation 1,

$$\boldsymbol{F} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{-\frac{i2\pi}{N}} & e^{-2\frac{i2\pi}{N}} & \dots & e^{-(N-1)\frac{i2\pi}{N}} \\ 1 & e^{-2\frac{i2\pi}{N}} & e^{-4\frac{i2\pi}{N}} & \dots & e^{-2(N-1)\frac{i2\pi}{N}} \\ 1 & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-(N-1)\frac{i2\pi}{N}} & e^{-2(N-1)\frac{i2\pi}{N}} & \dots & e^{-(N-1)^2\frac{i2\pi}{N}} \end{bmatrix}$$

For the case of equation 2,

$$\boldsymbol{F} = \boldsymbol{F}_C + i\boldsymbol{F}_S$$

where

$$\boldsymbol{F}_C = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & cos(-1\frac{i2\pi}{N}) & cos(-2\frac{i2\pi}{N}) & \dots & cos(-(N-1)\frac{i2\pi}{N}) \\ 1 & cos(-2\frac{i2\pi}{N}) & cos(-4\frac{i2\pi}{N}) & \dots & cos(-2(N-1)\frac{i2\pi}{N)} \\ 1 & \vdots & \vdots & \ddots & \vdots \\ 1 & cos(-(N-1)\frac{i2\pi}{N}) & cos(-2(N-1)\frac{i2\pi}{N}) & \dots & cos(-(N-1)^2\frac{i2\pi}{N}) \end{bmatrix}$$

$$\boldsymbol{F}_S = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & sin(-1\frac{i2\pi}{N}) & sin(-2\frac{i2\pi}{N}) & \dots & sin(-(N-1)\frac{i2\pi}{N}) \\ 0 & sin(-2\frac{i2\pi}{N}) & sin(-4\frac{i2\pi}{N}) & \dots & sin(-2(N-1)\frac{i2\pi}{N)} \\ 0 & \vdots & \vdots & \ddots & \vdots \\ 0 & sin(-(N-1)\frac{i2\pi}{N}) & sin(-2(N-1)\frac{i2\pi}{N}) & \dots & sin(-(N-1)^2\frac{i2\pi}{N}) \end{bmatrix}$$

Spectrogram. The magnitude of the Short-Time Fourier Transform (STFT), yields the spectrogram. In the case of a discrete signal of length L, the STFT is simply the DFT of length N segments, where $N < L$, of the signal. The result is a complex matrix $\boldsymbol{S}$ with the magnitude and phase of the signal for each frequency in each segment (time). Generally, the columns in the matrix represent the time dimension and the rows represent the different frequencies. The choice of the N value depends on the goal of the spectral representation. For small N values, high definition is obtained in the time dimension and low definition is obtained in the frequency dimension, while for high N values, the effect is reversed. The segments can be overlapped on m samples, for m between 0 and $N - 1$.

Convolutional Neural Networks. Each neuron of a convolutional neural network (CNN) performs a linear transformation of the input vector before an activation function. The output of the neuron j is defined as:

$$y_j = g(\sum_{i=0}^{N} w_{ij}\ x_i)$$

where
$g()$ is the activation function
x_i is the input i
w_{ij} is the synaptic weight corresponding to the input i of the neuron j
w_{0j} is the bias. ($x_0 = 1$).

Each layer of a CNN is composed of kernels. All kernels have the same number of neurons. Within each kernel, the neurons share the synaptic weights, but each one has its own receptive field (neurons connect to a limited set of inputs). Neurons of the same location in each kernel share the receptive field, that is, they connect to the same inputs. In this way, the output of a CNN layer is the discrete convolution of the inputs and the weights of each kernel. Once the network is trained, each kernel specializes in recognizing or transforming a certain pattern of the input.

1.1 Objectives

The objective of this work is to design an artificial neural network that predicts the spectrogram and to train the network with audio data. For this, the network must calculate the DFT.

The weights of the network can be set with the elements of the matrix $\boldsymbol{F}$ or be trained.

The main contribution of this work is the calculation of the synaptic weights by training. Demonstrating that the network can learn the coefficients of the DFT, suggests that other representations of the signal, better suited to the classification needs, could also be learned.

2 Motivation and related work

There are many deep learning works that use the magnitude of the frequency spectrum as input [2–4]. This article shows that it is possible to compute the same information by adding layers at the beginning of the deep network. The advantage is that the calculation of the spectrum can be adapted to a particular case.

Moreira et al. proposed the calculation of the DFT with cellular neural networks in [5]. They divide the weights into two groups that represent the real and imaginary parts. The training is not carried out, the weights are directly assigned.

Velik, in [6], predicts the DFT with weights calculated from complex exponential functions, also directly assigned, and reports that the neural network cannot be trained.

On the other hand, Anderson and Mallat [7] propose the replacement of the DFT by the Deep Scattering Spectrum (DSS) technique, based on wavelet transforms, because DSS is able to represent invariant characteristics over time (or space in the case of the images). Sometimes these deviations in time are significant for recognition. Our work is carried out in the field of vocal quality classification, where a deviation of the vibration frequency of the vocal cords is important [8].

3 Methods and Materials

3.1 Data

The neural network was designed to predict the spectrogram magnitudes of two seconds audio signals.

The audios are part of the Voice Disorders Database (VDD) [9], recorded by the *Universidad Politécnica de Madrid* in collaboration with the *Hospital Universitario Príncipe de Asturias*. The audio files contain approximately 2 seconds of a sustained vowel /a/. These were recorded from people with vocal pathologies and healthy people.

Inputs. The audios are in WAV format with a sample rate of 25000 samples per second. All files where duration ≥ 2 seconds were selected. Then the 50000 central samples were taken.

The input data is defined as 430 vectors of length $L = 50000$. 300 vectors were randomly chosen for training and 130 for validation.

Outputs. The outputs are calculated as the absolute value of the STFT of the inputs, with segments of size $N = 1760$ and overlap $m = 1540$ elements. It implies a 220 elements displacement by transformation. The module is calculated in order to obtain the spectrogram magnitude. The spectrogram calculated in this way has 220 columns (time) and 881 rows (frequency). Due to the nature

of the audios (voices in the usual pitch), the energy is concentrated almost completely in the first 200 rows (0-2826 Hz). For this reason, it was decided to train the network to predict only the magnitude of rows 1 to 200 of the spectrogram. For this case $K = 200$.

Then, the output is a vector of size 200×220. Note that the output values are in $\mathbb{R}$ because the module of the spectrum is taken.

3.2 Neural Network

To obtain the output defined in the previous section, the neural network must consist of two parts, one that calculates the STFT and another that obtains the absolute value. If the operation of calculating the absolute value is not included in the network, it is not possible to perform the training with the data as defined.

The STFT computation is done with a convolution layer, where the synaptic weights are the elements of the matrix $\boldsymbol{F}$ and the activation function is linear. This is possible because both, the DFT and the operation performed by each neuron, are linear transformations. It is important to note that the values of the matrix $\boldsymbol{F}$ are constant.

There are two ways to implement the calculation, depending on whether the DFT equation 1 or 2 is chosen. If the matrix $\boldsymbol{F}$, corresponding to the equation 1 is used, the weight matrix $\boldsymbol{W}$ consisting in the complex coefficients w_{ij}, will have size $N \times K$. In the case of the equation 2, $\boldsymbol{W}$, of size $N \times 2K$ will consist in the (real) values of the matrices $\boldsymbol{F}_C$ and $\boldsymbol{F}_S$.

In terms of efficiency, there is no difference between the two alternatives. For this work the second one is chosen because, since the output only conserves information of the spectrum magnitude, it is possible to avoid complex numbers operations. This could be a practical advantage because, among the neural networks software libraries, there is still no (in general) support for the complex numbers [10]. For the rest, the two approaches are equivalent.

Convolution layer. In order to calculate the STFT with a convolution layer, it is convenient to write the equation 2 in the following way.

$$X_k = \sum_{n=0}^{N-1} x_n \ \cos(-2\pi kn/N) + \sum_{n=0}^{N-1} i \ x_n \ \sin(-2\pi kn/N) \tag{3}$$

In matrix form,

$$\boldsymbol{X} = \boldsymbol{F}_C\boldsymbol{x} + i\boldsymbol{F}_S\boldsymbol{x}$$

Fig. 1 shows the proposed neuronal model. It can be seen that the output of the convolutional layer contains $\boldsymbol{F}_C\boldsymbol{x}$ and $\boldsymbol{F}_S\boldsymbol{x}$ values. These are the convolution of the input and each one of the $2K(400)$ kernels.

In the case of the direct assignment of the synaptic weights (without training), the first K kernels take the values of the K columns of matrix $\boldsymbol{F}_C$, and

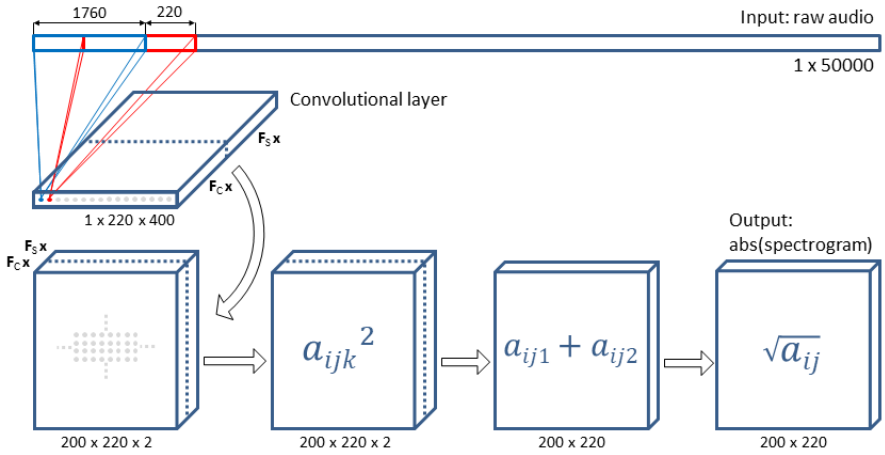


Fig. 1. Artificial neural network model that receives audio as input and predicts the magnitude of the spectrogram.

remaining kernels take the values of $\boldsymbol{F}_S$ columns. In this model, the neurons do not have the w_{0j} weight because the DFT has no term independent. Formally, the assignment of the weights is done as follows:

$$w_{ijk} = f_{Cik} \tag{4}$$

$$w_{ij(k+K)} = f_{Sik} \tag{5}$$

where

w_{ijk} is the synaptic weight of the input i of the neuron j of the kernel k.

f_{Cik} is the element in the row i and column k of $\boldsymbol{F}_C$.

f_{Sik} is the element in the row i and column k of $\boldsymbol{F}_S$.

Since the weights of neurons at the same kernel are shared, only one set of weights per kernel is stored. Note that there is not subindex j in the second terms of the equations 4 and 5. Then, the weights matrix W has size $N \times 2K$ (1760×400).

Magnitude of the spectrum. From equation 3, the magnitude of the frequency spectrum is as follows.

$$|X_k| = \left| \sum_{n=0}^{N-1} x_n \ \cos(-2\pi kn/N) + \sum_{n=0}^{N-1} i \ x_n \ \sin(-2\pi kn/N) \right|$$
$$= \sqrt{\left(\sum_{n=0}^{N-1} x_n \ \cos(-2\pi kn/N) \right)^2 + \left(\sum_{n=0}^{N-1} x_n \ \sin(-2\pi kn/N) \right)^2} \quad (6)$$

The results of the summations in the equation 6 are the scalars in the location k of the $\boldsymbol{F}_C\boldsymbol{x}$ and $\boldsymbol{F}_S\boldsymbol{x}$ vectors. Thus,

$$|X_k| = \sqrt{(\boldsymbol{F}_C\boldsymbol{x})_k^2 + (\boldsymbol{F}_S\boldsymbol{x})_k^2}$$

where $(\boldsymbol{F}_C\boldsymbol{x})_k$ and $(\boldsymbol{F}_S\boldsymbol{x})_k$ are the kth elements in $\boldsymbol{F}_C\boldsymbol{x}$ y $\boldsymbol{F}_S\boldsymbol{x}$ vectors respectively.

The output of the convolution layer of the model in Fig. 1 is an array containing the vectors $\boldsymbol{F}_C\boldsymbol{x}$ and $\boldsymbol{F}_S\boldsymbol{x}$ corresponding to all segments in the input signal. After the convolution, four operations arranged in layers are carried out. The first operation is a change in the form of the array in order to simplify the third operation, the second calculates the square of each element, the third adds pairs of values corresponding to the same frequency (k) and time segment, and finally, the calculation of the square root of each element is performed. The result is a matrix of size 200×220 with the magnitudes of the spectrum elements.

Training. Instead of assigning the weights directly, these can be trained through the gradient descent method on the mean squared error (MSE) function. The calculation of the gradient includes the derivatives of the three last layers (operations) of the model. In the next section, the results of this process are presented and compared with those obtained through direct assignment.

4 Results

Below, the result of 30,000 training cycles of the proposed model is shown. The weights were initialized with random values between -10^6 and 10^6 uniformly distributed. The optimization method Adam [11] was used with the parameters provided by the authors. The weights were updated in batches of size 300 (the whole of training data). The calculations were performed on an NVIDIA Titan Xp GPU donated through NVIDIA's GPU Grant Program.

The MSE reached on validation data was 1.41×10^{-6} (9.79×10^{-6}% of the mean expected output), while for the same model with directly assigned weights an MSE $< 10^{-9}$ was achieved. Fig. 2 shows the expected output for one vector in validation dataset and the output obtained by the network after the training. At first glance there are no differences.

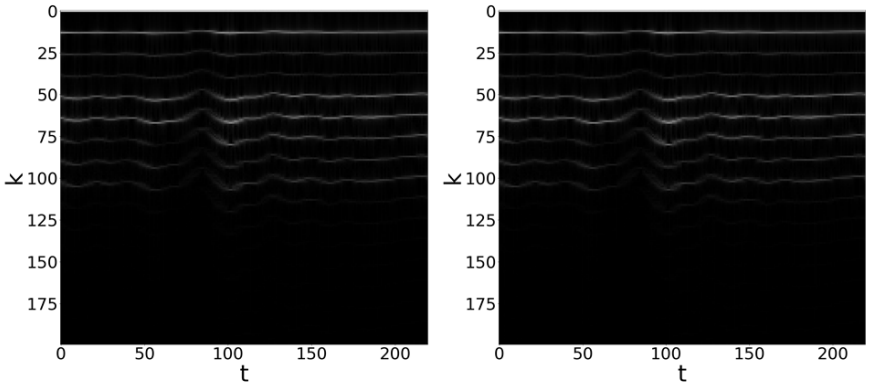


Fig. 2. Output (spectrogram) of the Neural network. Expected output (left) and output obtained with trained weights (right). At first glance there are no differences.

The result obtained by training is widely satisfactory. A comparison between the theoretical weights, calculated from the equations of the DFT, and the trained ones is presented below.

In Fig. 3 It can be seen the values of $\boldsymbol{F}_C$ and $\boldsymbol{F}_S$ for weights assigned directly (theoretical) and for trained weights. In all cases it is clearly observed that the upper rows of the transposed matrices represent the low frequencies and lower rows represent high frequencies.

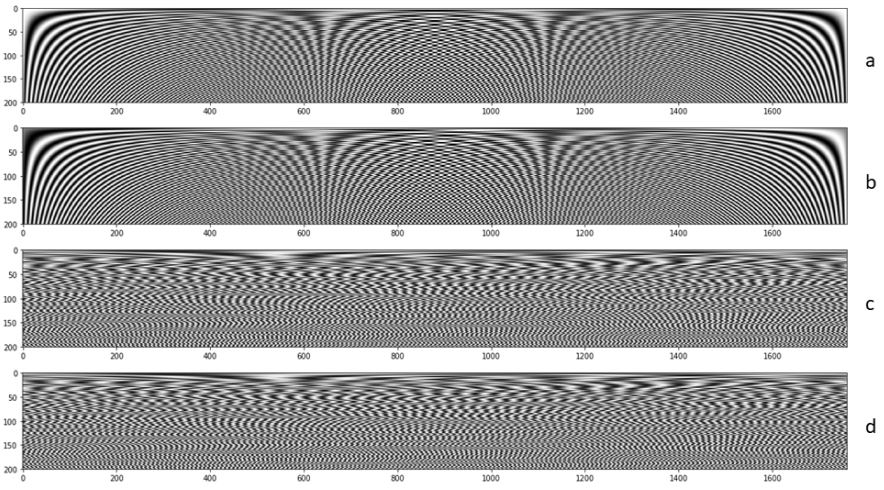


Fig. 3. Transposed matrices of coefficients. $\boldsymbol{F}_C$ with theoretical weights (a), $\boldsymbol{F}_S$ with theoretical weights (b), $\boldsymbol{F}_C$ with trained weights (c) and $\boldsymbol{F}_S$ with trained weights (d).

It is also evident that the two groups, theoretical and trained weights, present different image patterns. The trained weights have a "messy" appearance. The

origin of this phenomenon is that TDF performs a decomposition of the input into a weighted sum of the sinusoidal signals found in the matrices $\boldsymbol{F}_C$ and $\boldsymbol{F}_S$. The training method, for each value of k, finds a pair $\boldsymbol{F}_C$ and $\boldsymbol{F}_S$ formed by sine waves with a 90° phase offset allowing the desired decomposition, but this solution is not necessarily the same as the equation 3.

Fig. 4 shows three examples of trained vs. theoretical weights for particular values of k. Note that, for both theoretical and trained weights there is always a 90° phase shift between $\boldsymbol{F}_C$ and $\boldsymbol{F}_S$. This can be checked by calculating $mod = \sqrt{\boldsymbol{F}_C^2 + \boldsymbol{F}_S^2}$ for any value of k. For theoretical weights, obviously $mod = 1$, while for trained weights a value very close to 1 is always obtained. In this way, an orthogonal base is found to perform the decomposition.

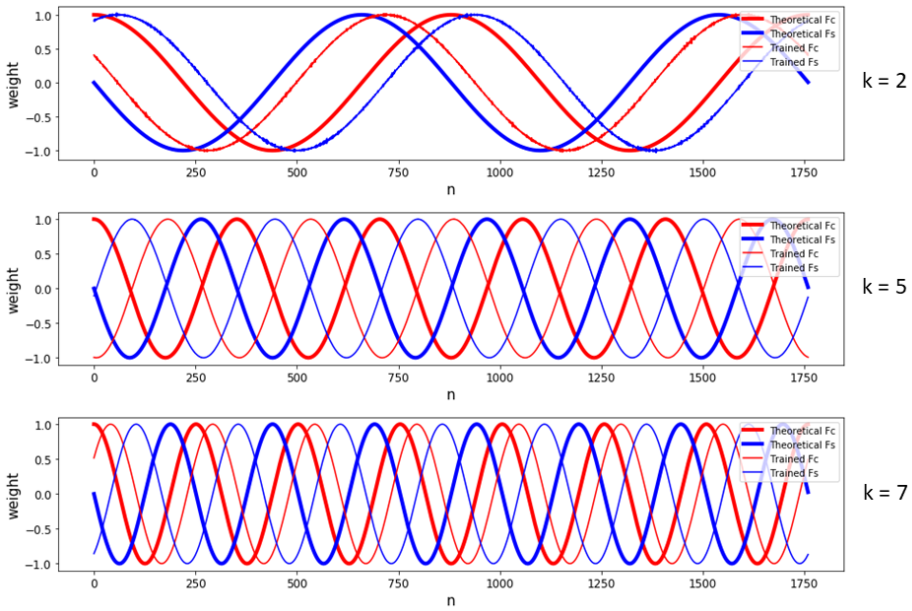


Fig. 4. Trained synaptic weights (thin lines) and theoretical weights (thick lines) for $k = 2$, 5 and 7. The values of $\boldsymbol{F}_C$ are shown in red and the values of $\boldsymbol{F}_S$ in blue.

5 Conclusion

It is concluded that a neuronal model is able to calculate the DFT, both for theoretical and trained weights, and that the trained weights do not necessarily tend towards the theoretical ones, although they share frequency and orthogonality condition. In addition, a convolutional network with the presented characteristics can be trained to calculate the spectrogram of the input signal. If the expected output is the magnitude of the spectrum, it is possible to avoid the operations

of complex numbers by adding operations in layers that calculate the Euclidean distance between the components of the same frequency.

The advantage of training the model is that it can be adapted to particular problems. For example, in the STFT computation it is common to use a window that softens the signal ends. These functions can be achieved with the proposed network by attenuating the ends of the weights. There are several window functions, but for a particular problem a different one may be better.

Then, to solve a particular problem, the weights can be initialized with random values, theoretical weights of the DFT or theoretical weights modified by some function (window function for example) and then, trained to find the optimal combination for the problem.

References

1. McClellan, J., Parks, T.: Eigenvalue and eigenvector decomposition of the discrete fourier transform. IEEE Transactions on Audio and Electroacoustics **20**(1) (1972) 66–74
2. Collobert, R., Puhrsch, C., Synnaeve, G.: Wav2letter: an end-to-end convnet-based speech recognition system. arXiv preprint arXiv:1609.03193 (2016)
3. Xu, Y., Du, J., Dai, L.R., Lee, C.H.: An experimental study on speech enhancement based on deep neural networks. IEEE Signal processing letters **21**(1) (2014) 65–68
4. Putten, M.J., Olbrich, S., Arns, M.: Predicting sex from brain rhythms with deep learning. Scientific reports **8**(1) (2018) 3069
5. Moreira-Tamayo, O., De Gyvez, J.P.: Filtering and spectral processing of 1-d signals using cellular neural networks. In: Circuits and Systems, 1996. ISCAS'96., Connecting the World., 1996 IEEE International Symposium on. Volume 3., IEEE (1996) 76–79
6. Velik, R.: Discrete fourier transform computation using neural networks. In: 2008 International Conference on Computational Intelligence and Security, IEEE (2008) 120–123
7. Andén, J., Mallat, S.: Deep scattering spectrum. IEEE Transactions on Signal Processing **62**(16) (2014) 4114–4128
8. García, M.A., Destéfanis, E.A.: Deep neural networks for shimmer approximation in synthesized audio signal. In: Argentine Congress of Computer Science, Springer (2017) 3–12
9. Arias-Londoño, J.D., Godino-Llorente, J.I., Markaki, M., Stylianou, Y.: On combining information from modulation spectra and mel-frequency cepstral coefficients for automatic detection of pathological voices. Logopedics Phoniatrics Vocology **36**(2) (2011) 60–69
10. Trabelsi, C., Bilaniuk, O., Zhang, Y., Serdyuk, D., Subramanian, S., Santos, J.F., Mehri, S., Rostamzadeh, N., Bengio, Y., Pal, C.J.: Deep complex networks. arXiv preprint arXiv:1705.09792 (2017)
11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

Transferencia de Aprendizaje para Clasificación de Peatones

Genaro Camele[1], Facundo Quiroga[1], Franco Ronchetti[1], Waldo Hasperué[1,2], and Laura Lanzarini[1]

[1] Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata,
[2] Investigador Asociado - Comisión de Investigaciones Científicas (CIC)
{gcamele,fquiroga,fronchetti,whasperue,laural}@lidi.info.unlp.edu.ar

Resumen Desde la aparición, en el año 2005, de los Histogramas de Gradientes Orientados (*HOG*) como descriptor para detección de peatones, han aparecido numerosas publicaciones que permitieron mejorar la clasificación agregando mejoras o comparando con nuevos descriptores. Del mismo modo, año a año aparecen nuevas bases de datos con imágenes de entornos reales, que permiten la evaluación de los modelos desarrollados. No obstante, la utilización de un modelo entrenado en un entorno real nuevo no siempre resulta trivial y es un tema pendiente de estudio para este dominio.
En este artículo, presentamos un protocolo para evaluar la transferencia de aprendizaje entre tres de las bases de datos más utilizadas en la literatura: *INRIA, Daimler* y *TUD-Brussels*. Comparamos los descriptores HOG y Patrones Binarios Locales (LBP) en conjunto con Máquinas de Vectores de Soporte (SVM) como clasificador de base. Los resultados obtenidos muestran que si bien cada conjunto de datos presenta escenas del mundo real, existen diferencias significativas que hacen que un modelo entrenado con un conjunto de imágenes no funcione apropiadamente con otro. Por otro lado, encontramos que al entrenar un modelo con la mezcla de diferentes bases de datos permite una mayor transferencia de aprendizaje, si bien no siempre ayuda al entrenamiento de un conjunto de datos particular.

Keywords: detección de peatones, transferencia de aprendizaje, SVM, HOG, LBP, Daimler, Inria, TUD-Brussels

1. Introducción y estado del arte

La detección de peatones puede verse como uno de los temas más importantes de la detección de objetos dentro del área de visión por computador. En los últimos años ha atraído la atención de muchos investigadores, lo que resultó en grandes avances debido a su impacto social y aplicaciones en seguridad vial [12][1]. En los últimos años, ya superada la barrera de la detección con una tasa de acierto relativamente alta, los investigadores comenzaron a enfocarse en

problemas puntuales como la detección de personas parcialmente ocultas, o la detección de imágenes borrosas o con poca resolución[8].

La detección de peatones, como otras tareas de detección de objetos, consiste principalmente de tres partes: el cálculo de un descriptor de imágenes adecuado para representar a un peatón, un modelo de clasificación que distinga imágenes de peatones (positivas) de las que no son (negativas) en base al descriptor, y un sistema de ventana deslizante que encuentra los peatones en una escena realizando varias ejecuciones del clasificador. En este artículo nos enfocamos en las primeras dos etapas ya que son las más importantes para evaluar la transferencia de aprendizaje.

Existen numerosas aproximaciones para llevar a cabo esta tarea en cuanto al cómputo de descriptores y métodos de clasificación. Si bien existen diversos artículos que investigan descriptores como Haar, detección de bordes, o histogramas de color, el enfoque más ampliamente estudiado, y que mejores resultados ha dado, es la combinación de descriptores HOGs (*Histogram of Oriented Gradients* con SVM [3][4]. En el mismo sentido, podemos encontrar en segundo lugar a los descriptores LBP (*Local binary pattern*) generalmente utilizados en combinación con los descriptores HOGs [10][9][6].

Una suposición importante en muchos enfoques tradicionales de detección de peatones es que el conjunto de datos de entrenamiento y el conjunto de datos de prueba deben ser similares. Esto se deduce del hecho de que las imágenes de entrenamiento y las de prueba son tomadas de la misma base de datos. Esta premisa dificulta transferir un modelo entrenado a un entorno real, donde las condiciones no son iguales a las que había en la base de datos original. Es decir, las características de las imágenes de las distintas bases de datos siguen una distribución diferente [1].

Hay varias razones por las cuales las características de distintas bases de datos varian. Por un lado, puede diferir el proceso de captura, por ejemplo debido a variaciones climatológicas que inciden en la luz o la temperatura de color. En el mismo sentido, las especificaciones técnicas de la cámara de captura podrían no ser similares. Por otro lado, el contexto en que se captura la base de datos nunca es el mismo, presentando problemas aún mayores como las variaciones en la vestimenta de los peatones y el fondo de la escena. Por ejemplo, algunas bases de datos presentan escenas con naturaleza de fondo y otras se capturaron en contextos de tráfico en ciudades o autopistas [3][5][11]. Si bien la utilización de un descriptor apropiado ayuda a que exista una relación más cercana entre el espacio de las imágenes de entrenamiento con respecto a las de prueba, no son lo suficientemente invariantes a estas diferencias en las características de las bases de datos.

En este contexto, existen muy pocos trabajos que utilicen diversas bases de datos existentes para evaluar la posibilidad de utilizar un modelo entrenado en escenas no vistas. En [2] realizan un proceso de transferencia con el objetivo de generar un modelo más robusto a variaciones en las imágenes, pero se enfocan en aumentar el desempeño en una base de datos particular y no en evaluar la capacidad de transferir el aprendizaje entre bases de datos. En [1] los autores

analizan la transferencia de aprendizaje para algunas bases de datos. No obstante, al utilizar la tarea de detección como medida de desempeño en lugar de la de clasificación introducen factores de variabilidad mayores que dificultan la interpretación de la transferencia de aprendizaje.

En este trabajo nos proponemos evaluar la transferencia de aprendizaje de modelos de *clasificación* de peatones en tres bases de datos ampliamente utilizadas: INRIA [3], Daimler [5] y TUD-Brussels [11]. Para ello establecimos un protocolo de experimentación para generar lo que definimos como *matrices de transferencia*. Las mismas reflejan, aproximadamente, la similitud entre las distintas bases de datos, así como la complementariedad de las mismas para generar un modelo de clasificación robusto y capaz de generalizar a nuevas escenas. De esta manera, podemos evaluar posibilidad de transferir un modelo entrenado con una base de datos a nuevas escenas con características de imagen diferentes.

2. Bases de datos y descriptores

A continuación presentamos las tres bases de datos que comparamos en los experimentos, Daimler Mono Pedestrian Detection, INRIA Person y TUD-Brussels Motionpairs. Además, se introducen brevemente los dos descriptores analizados, elegidos en base a los más utilizados en el estado del arte.

2.1. Bases de datos

La base de datos **INRIA Person** [3] (Figura 1) contiene imágenes de personas en varias situaciones. Fue generada en base a varios corpus e imágenes recolectadas de forma independiente. Las mismas son a color y de tamaño 64×128 píxeles. El conjunto de entrenamiento contiene 3.030 y 34.892 ejemplos positivos y negativos respectivamente, y el de prueba 1.028 y 453.



Figura 1: Ejemplos positivos (izquierda) y negativos (derecha) de la base de datos INRIA Person[3].

La base de datos **Daimler Mono Pedestrian Detection** [5] (Figura 2) se enfoca en la detección de peatones en cámaras monoculares. Contiene imágenes en escala de grises tomadas desde un auto en movimiento. El conjunto de

entrenamiento contiene 15.560 imágenes de peatones de 48×96 píxeles. Además, se incluyen 6.744 imágenes originales (tamaño completo) en donde no hay peatones con el objetivo de ser utilizadas para extraer patrones negativas. Con estas imágenes se generaron 32.000 patrones negativos del mismo tamaño que las imágenes positivas tomando muestras de forma aleatoria. Luego, dividimos los conjuntos de imágenes utilizando el 90 % para entrenamiento y el 10 % para evaluación.



Figura 2: Ejemplos positivos (arriba) y negativos (abajo) de la base de datos Daimler Mono Pedestrian Detection[5].

La base de datos **TUD-Brussels Motionpairs** [11] (Figura 3) también se enfoca en la detección de peatones desde un auto en movimiento, con imágenes a color de 640×480 píxeles. Contiene 551 peatones etiquetados para el conjunto de entrenamiento y 311 para el de prueba. El cuadro 1 resume la información básica de las tres bases de datos utilizadas.



Figura 3: Ejemplos de la base de datos TUD-Brussels Motionpairs [11].

2.2. Descriptores

El **Histograma de Gradientes Orientados (HOGs)** [3] se calcula en base al gradiente de la imagen. Dado que dicho gradiente tiende a ser más grande en los bordes de los objetos, es apto para representar la forma de los mismos. En

Cuadro 1: Detalle de los conjuntos de datos utilizados.

Base	Entrenamiento		Prueba		Color	Contexto
de datos	Positivas	Negativas	Positivas	Negativas		
Daimler	14.000	30.000	1.560	2.000	No	Peatones
INRIA	3.030	34.879	1.028	453	Sí	Personas
TUD-Brussels	982	9.064	110	1.010	Sí	Peatones

base al gradiente, representado como ángulo-magnitud, se divide la imagen en celdas de igual tamaño. Para cada celda se calcula un histograma con n cubetas. Cada cubeta corresponde a un rango de ángulos, y cada punto del gradiente contribuye a la cubeta en forma proporcional a su magnitud. Luego los histogramas de las celdas contiguas se agrupan en bloques, típicamente superpuestos, que se normalizan. Los bloques normalizados luego pierden su estructura 2D y se concatenan en un vector unidimensional de modo de formar el descriptor final. La última fila de la Figura 4 presenta una visualización de las celdas del HOG de imágenes positivas y negativas de cada dataset.

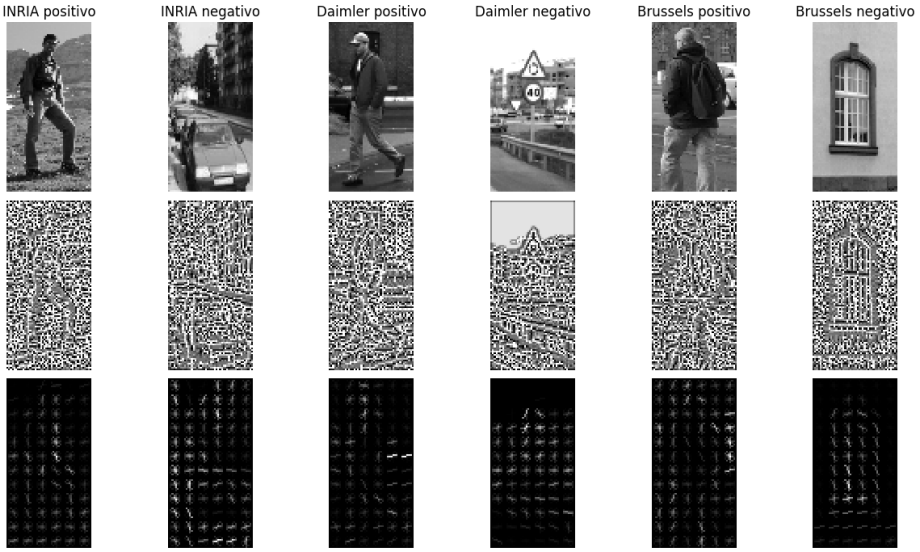


Figura 4: Imágenes positivas y negativas originales de cada una de las bases de datos (fila 1). LBPs y HOGs calculados para cada imagen (filas 2 y 3).

Patrones Binarios Locales (LBP). Los **Patrones Binarios Locales (LBP)** [7] son descriptores de textura. Se calculan dividiendo la imagen en celdas, y por

cada pixel de la celda se calcula un número binario de n dígitos que representa el gradiente en un camino de la vecindad del pixel, donde n es el tamaño del camino. Luego, se genera un histograma con 2^n cubetas para toda la celda, donde cada pixel incrementa en 1 la cubeta correspondiente a ese número binario. Al igual que con los HOGs los histogramas se concatenan para formar el descriptor final. La segunda fila de la Figura 4 también presenta una visualización del descriptor.

3. Experimentos

3.1. Protocolo del experimento

A continuación presentamos el protocolo de experimentación para generar las matrices de transferencia de aprendizaje. En primer lugar, entrenamos un modelo distinto para cada una de las tres bases de datos, utilizando su conjunto de entrenamiento respectivo. Luego, por cada modelo, se realiza la evaluación del mismo con los tres conjuntos de evaluación de las bases de datos. Como métrica de desempeño se utilizó la medida-F (*F-Score*), de modo de combinar la precisión (*precision*) y la exhaustividad (*recall*) en una sola métrica.

De este modo, obtuvimos una matriz de transferencia D, de tamaño 3×3, entre las tres bases de datos, donde $D_{i,j}$ nos indica la medida-F del modelo al ser entrenado con la base de datos i y evaluado con la base de datos j. De este modo, podemos observar, la similitud que hay entre las características de las imágenes de las diferentes bases de datos, para cada descriptor.

3.2. Preprocesamiento de los datos y entrenamiento del modelo

Dado que no todas las bases de datos tienen el mismo tamaño de imagen, unificamos el mismo a 48×96 píxeles como medida estándar para llevar a cabo los experimentos. Esto no supone ninguna deformación de relación de aspecto debido a que en todos los casos es de $1:2$.

Convertimos las imágenes a color de INRIA y TUD-Brussels a escala de grises de modo que el cálculo del HOG y los LBPs sea el mismo para los distintos conjuntos de datos. Normalizamos el rango de las imágenes al intervalo $[0 \dots 1]$. No realizamos una ecualización del histograma ni ningún otro tipo de preprocesamiento adicional a las imágenes para preservar las distribuciones originales de los datos.

Calculamos los HOGs sobre la imagen resultante con un tamaño de celda de 8×8 píxeles, con bloques de 2×2 celdas. Aplicamos normalización L2 a los descriptores obtenidos, siguiendo la metodología de [3]. Los LBPs se calculan con 8 puntos de muestreo en un radio de 1 pixel, siguiendo también la metodología de [10].

Como modelo de clasificación utilizamos Máquinas de Vector de Soporte ya que es el modelo más utilizado en la literatura al mismo tiempo de ser simple y muy eficiente computacionalmente para luego incorporarlo a un esquema de detección. En este caso utilizamos un SVM con una configuración similar a [3]

donde $C = 0{,}1$, el núcleo es lineal, igual peso para las dos clases y regularización L2. Entrenamos el modelo mediante el método SMO de la librería *liblinear* con la formulación dual, verificando que el modelo converge y no sobreajusta.

3.3. Resultados y discusión

El cuadro 2 muestra las matrices de transferencia entre los tres conjuntos de datos, para cada uno de los descriptores utilizados. Claramente, la transferencia de aprendizaje no es trivial. Esto queda reflejado al ver que la diagonal principal de cada matriz, que contiene los resultados de entrenar y evaluar un modelocon la misma base de datos, tiene una medida-F muy superior al resto de las entradas. Si bien Daimler parece tener mejor capacidad de transferencia, es posible que se deba a la mayor cantidad de ejemplos de esa base de datos.

Por otro lado, podemos observar que el mejor resultado obtenido de manera general es utilizando el descriptor HOG. El agregado del descriptor LBP como parte representativa de una imagen no parece generar gran impacto sobre la precisión del modelo, considerando el aumento significativo en la dimensión del descriptor final. No obstante, este agregado demuestra mejorar significativamente la transferencia de aprendizaje para algunos casos como por ejemplo al entrenar con la base de datos INRIA y evaluar con las otras dos. Esto indica que la transferencia es muy dependiente del descriptor utilizado. Es posible que esto se deba también al hecho de que INRIA es una base de datos de personas y no de peatones exclusivamente.

Cuadro 2: Matrices de transferencia entre las tres bases de datos utilizadas. Las entradas muestran los *F-Score*. Filas: Entrenamiento. Columnas: Evaluación. Abreviaciones: I = INRIA, D = Daimler, B = TUD-Brussels.

(a) HOG

	I	D	B
I	0.619	0.297	0.064
D	0.841	0.984	0.627
B	0.335	0.396	0.803

(b) LBP

	I	D	B
I	0.176	0.078	0.097
D	0.693	0.887	0.278
B	0.167	0.151	0.218

(c) HOG+LBP

	I	D	B
I	0.611	0.351	0.210
D	0.837	0.979	0.556
B	0.376	0.369	0.748

3.4. Transferencia al entrenar con múltiples bases de datos

Una extensión natural del experimento anterior consiste en unir los conjuntos de entrenamiento de dos bases de datos para generar el modelo, y realizar la evaluación de igual forma que en el experimento anterior. De esta forma podemos evaluar si el agregado de datos diversos contribuye a la mejora de la transferencia. Es decir, analiza la complementariedad de las bases de datos.

El Cuadro 3 muestra las matrices de transferencia resultantes. Nuevamente, la métrica utilizada es la medida-F.

Cuadro 3: Matrices de transferencia entre las tres bases de datos utilizadas al entrenar con múltiples bases de datos. Las entradas muestran los *F-Score*. Filas: Entrenamiento. Columnas: Evaluación. Abreviaciones: I = INRIA, D = Daimler, B = TUD-Brussels.

(a) HOG

	I	D	B
I+D	0.837	0.960	0.578
I+B	0.747	0.645	0.658
D+B	0.816	0.976	0.850

(b) LBP

	I	D	B
I+D	0.444	0.709	0.215
I+B	0.273	0.149	0.176
D+B	0.581	0.829	0.303

(c) HOG+LBP

	I	D	B
I+D	0.844	0.946	0.472
I+B	0.697	0.562	0.497
D+B	0.789	0.965	0.766

Podemos observar que al entrenar con varias bases de datos aumenta la medida-F en casi todos los casos con respecto a usar una sola base de datos. No obstante, debemos considerar un efecto de tamaño por la unión de los conjuntos de entrenamiento. Por ejemplo, TUD-Brussels no aumenta de forma considerable los ejemplos al sumarlos a los de INRIA y sin embargo aumenta significativamente la medida-F de su conjunto de evaluación. La figura 5 resume los resultados obtenidos en todos los experimentos utilizando el descriptor HOG, que fue el que mejor desempeño logró. Las diferentes series muestran las distintas configuraciones de entrenamiento, mientras que las 3 columnas principales representan las tres bases de datos de evaluación.

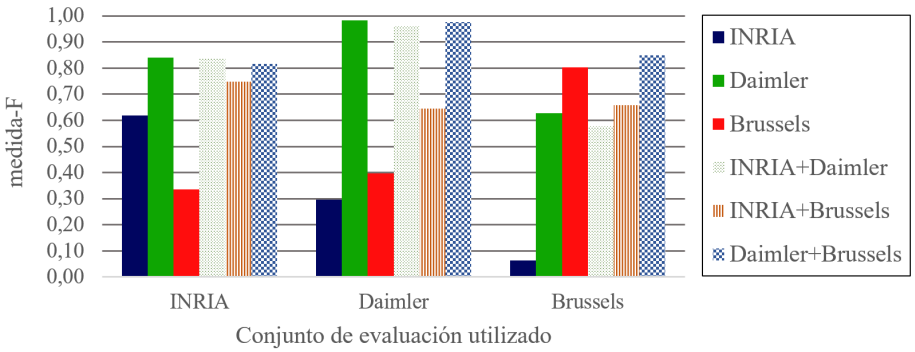


Figura 5: Gráfico comparativo de la medida-F para la evaluación de cada base de datos al entrenar con todas las combinaciones de bases de datos analizadas, utilizando el descriptor HOG.

4. Conclusión

En este artículo presentamos un protocolo para evaluar la transferencia de aprendizaje entre bases de datos de peatones. Los resultados indican que si bien los conjuntos de datos más ampliamente utilizados para la detección de peatones tienen la misma finalidad, presentan diferencias singificativas que hacen no trivial la utilización de un modelo entrenado en un entorno real, con cambios en la luminosidad o condiciones de escena. Ninguna de las tres bases de datos resulta ideal como único modelo de entrenamiento para llevar a cabo un proceso de transferencia de aprendizaje, aunque, en base a los resultados Daimler, parece la más apta de las evaluadas.

En trabajos futuros, buscaremos ampliar el rango de bases de datos para aumentar la riqueza de la comparación, así como cuantificar el efecto del tamaño de las mismas. Además, esperamos evaluar modelos más recientes del estado del arte como aquellos basados en Redes Convolucionales. Una vez terminada la evaluación de la transferencia para clasificación, proponemos ampliar los experimentos para incluir tareas de detección con el fin de evaluar su indicidencia.

Referencias

1. Benenson, R., Omran, M., Hosang, J., Schiele, B.: Ten years of pedestrian detection, what have we learned? In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) Computer Vision - ECCV 2014 Workshops. pp. 613–627. Springer International Publishing, Cham (2015)
2. Cao, X., Wang, Z., Yan, P., Li, X.: Transfer learning for pedestrian detection. Neurocomputing 100, 51–57 (2013), http://www.sciencedirect.com/science/article/pii/S0925231212003256, special issue: Behaviours in video
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 1, pp. 886–893. IEEE (2005)
4. Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: An evaluation of the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. 34(4), 743–761 (Apr 2012), http://dx.doi.org/10.1109/TPAMI.2011.155
5. Enzweiler, M., Gavrila, D.M.: Monocular pedestrian detection: Survey and experiments. IEEE Transactions on Pattern Analysis & Machine Intelligence (12), 2179–2195 (2008)
6. Gan, G., Cheng, J.: Pedestrian detection based on hog-lbp feature. 2011 Seventh International Conference on Computational Intelligence and Security pp. 1184–1187 (2011)
7. Mu, Y., Yan, S., Liu, Y., Huang, T., Zhou, B.: Discriminative local binary patterns for human detection in personal album. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. pp. 1–8. IEEE (2008)
8. Ouyang, W., Wang, X.: Single-pedestrian detection aided by multi-pedestrian detection. 2013 IEEE Conference on Computer Vision and Pattern Recognition pp. 3198–3205 (2013)
9. Pei, W.J., Zhang, Y.L., Zhang, Y., Zheng, C.H.: Pedestrian detection based on hog and lbp. In: Huang, D.S., Bevilacqua, V., Premaratne, P. (eds.) Intelligent Computing Theory. pp. 715–720. Springer International Publishing, Cham (2014)

10. Wang, X., Han, T.X., Yan, S.: An hog-lbp human detector with partial occlusion handling. In: Computer Vision, 2009 IEEE 12th International Conference on. pp. 32–39. IEEE (2009)
11. Wojek, C., Walk, S., Schiele, B.: Multi-cue onboard pedestrian detection. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). vol. 00, pp. 794–801 (06 2009), doi.ieeecomputersociety.org/10.1109/CVPRW.2009.5206638
12. Yan, J., Zhang, X., Lei, Z., Liao, S., Li, S.Z.: Robust multi-resolution pedestrian detection in traffic scenes. 2013 IEEE Conference on Computer Vision and Pattern Recognition pp. 3033–3040 (2013)

XIX Workshop Procesamiento Distribuido y Paralelo (WPDP)

Coordinadores

Marcela Printista (UNSL) - Laura De Giusti (UNLP) - Carlos García Garino (UNCuyo)

Temas

Algoritmos Paralelos
Métricas de Performance y Eficiencia en Paralelismo
Escalabilidad
Optimización de algoritmos paralelos
Modelos de Arquitecturas paralelas
Computación de Altas Prestaciones
Predicción de Performance en Algoritmos Paralelos
Balance de Carga en Sistemas Distribuidos/Paralelos
Mapping, scheduling y virtualización en arquitecturas paralelas
Arquitecturas multiprocesador orientadas a Paralelismo
Software de base para multicores, simétricos y asimétricos
Arquitecturas distribuidas como Procesadores Paralelos
Evaluación de performance en sistemas paralelos homogéneos y heterogéneos
Cluster, Grid y Cloud Computing
Middleware para arquitecturas de procesamiento paralelo
Fallas de hardware y software en Arquitecturas Paralelas. Tolerancia a Fallas
Aplicaciones de Sistemas Distribuidos/Paralelos
Simulación de Alta Performance

Factores que afectan el consumo energético de operaciones de checkpoint y restart en clusters

Marina Morán[1], Javier Balladini[1], Dolores Rexachs[2], Emilio Luque[2]

[1]Facultad de Informática, Universidad Nacional del Comahue
Buenos Aires 1400, 8300 Neuquén, Argentina
{javier.balladini, marina}@fi.uncoma.edu.ar
[2]Departamento de Arquitectura de Computadores y Sistemas Operativos, Universitat Autònoma de Barcelona
Campus UAB, Edifici Q, 08193 Bellaterra (Barcelona), España
{dolores.rexachs, emilio.luque}@uab.es

Resumen El método de tolerancia a fallos mas usado actualmente en Cómputo de Altas Prestaciones es el de rollback-recovery mediante el uso de checkpoints. Éste, como cualquier otro método de tolerancia a fallos, agrega un consumo energético adicional al propio de la ejecución de la aplicación. El objetivo de este trabajo es determinar los factores que afectan el consumo energético de los nodos de cómputo de un cluster homogéneo, al ejecutar operaciones de checkpoint y restart, sobre aplicaciones SPMD (Single Program Multiple Data). Nos hemos enfocado en el estudio energético de nodos de cómputo, contemplando diferentes configuraciones de parámetros de hardware y software. Se estudió el efecto de los estados de rendimiento (estados P) y potencia (estados C) de los procesadores, el tamaño del problema de la aplicación, la configuración del software de checkpoint utilizado (DMTCP), y del sistema de archivos distribuído (NFS). El análisis de los resultados permitió identificar oportunidades que permiten disminuir el consumo energético de las operaciones de checkpoint y restart.

1. Introducción

La computación de altas prestaciones (HPC, High Performance Computing) continúa aumentando su poder de cómputo al mismo tiempo que aumenta su consumo energético. Dadas las limitaciones que existen para abastecer de energía a este tipo de computadoras, se hace necesario conocer el comportamiento del consumo energético en estos sistemas, y encontrar formas de disminuirlo o acotarlo. En especial, para la era exaescala, se estima un límite máximo de 20 MW [4].

El método de tolerancia a fallos mas usado actualmente en CAP es el de rollback-recovery mediante el uso de checkpoints. Éste, como cualquier otro método de tolerancia a fallos, agrega un consumo energético adicional al propio de la ejecución de la aplicación [16]. Debido a ello, es importante conocer y predecir el comportamiento energético de métodos de tolerancia a fallos, de manera de

poder gestionar su impacto en el consumo total de energía durante la ejecución de una aplicación.

Un sistema de cómputo de tipo cluster tiene nodos de cómputo, nodos de almacenamiento, y al menos una red de interconexión. El objetivo de este trabajo es determinar los factores que afectan el consumo energético de los nodos de cómputo de un cluster homogéneo, al ejecutar operaciones de checkpoint y restart (C/R), sobre aplicaciones SPMD (Single Program Multiple Data). Nos hemos enfocado en el estudio energético de nodos de cómputo, contemplando diferentes configuraciones de parámetros de hardware y software. El consumo energético del nodo de almacenamiento y la red de almacenamiento no se consideran en este artículo.

Las contribuciones de este artículo son:

- Un estudio de factores propios del sistema (hardware y software) y de las aplicaciones que impactan en el consumo energético producido por las operaciones de checkpoint y restart.
- La identificación de oportunidades que permitan disminuir el consumo energético de las operaciones de checkpoint y restart.

La sección 2 presenta algunos trabajos relacionados, mientras que en la sección 3 se identifican los factores que inciden en el consumo energético. En la sección 4 se detalla la plataforma experimental y el diseño de experimentos, cuyos resultados y su análisis son presentados en la sección 5. Finalmente, se presentan las conclusiones y trabajos futuros en la sección 6.

2. Trabajos Relacionados

En [8] y [11] evalúan el comportamiento energético del C/R coordinado y no coordinado con logeo de mensajes. En [11] además evalúan la recuperación paralela y proponen un modelo analítico para predecir el comportamiento de dichos protocolos a escala exa. En [12] utilizan un modelo analítico para comparar el tiempo de ejecución y la energía consumida de la replicación y del C/R coordinado. En [2] y [6] presentan un modelo analítico para estimar el intervalo óptimo de un checkpoint multinivel en términos de consumo energético. No miden potencia disipada sino que utilizan valores de otras publicaciones. En [7] miden la potencia disipada y el tiempo de ejecución de las operaciones de alto nivel involucradas en el checkpoint (coordinados, no coordinados y jerárquicos) variando el número de núcleos involucrados. No utilizan diferentes frecuencias de procesador, ni indican si el checkpoint es comprimido o no. En [14], [5] y [15] presentan un marco para el ahorro energético del C/R. En [14] reemplazan muchas operaciones de E/S pequeñas por pocas grandes a cargo de un solo núcleo, para hacer más eficiente energéticamente el checkpoint y el restart. Utilizan RAPL para medir y limitar el consumo energético. En [15] diseñan un *runtime* que permite modificar la frecuencia de reloj y el número de procesos que realizan las operaciones de E/S del C/R de manera de optimizar el consumo energético. Otro trabajo donde se analiza el impacto de del escalado dinámico de frecuencia

y tensión en el consumo energético de las operaciones de checkpoint es en [13]. En [9] evalúan el consumo energético de una aplicación que utiliza checkpoints comprimidos. Muestran que al utilizar compresión se gasta más energía pero se ahorra tiempo, por lo que la ejecución completa de la aplicación con todos sus checkpoints puede verse beneficiada desde el punto de vista energético.

Nuestro trabajo se centra en C/R coordinado a nivel de sistema. Los valores de potencia disipada de las operaciones de checkpoint y restart son mediciones obtenidas con un medidor físico externo. No hemos encontrado trabajos que evalúen el impacto de los estados C y de las configuraciones del NFS en el consumo energético de las operaciones de C/R.

3. Identificación de factores que inciden en el consumo energético

La energía se puede calcular como el producto entre la potencia y el tiempo. Cualquier factor que pueda incidir sobre alguno de estos dos parámetros debe ser contemplado para luego analizar de qué manera afecta al consumo energético. Enfocándonos en el estudio energético de nodos de cómputo, los factores están relacionados a diferentes niveles: Hardware, Software de Aplicación y Software de Sistema.

Hardware: La especificación de Interfaz de Potencia y Configuración Avanzada (ACPI, acrónimo de Advanced Configuration and Power Interface) provee un estandar abierto que permite al sistema operativo gestionar la energía de los dispositivos y del sistema de cómputo entero [1]. Permite gestionar el comportamiento energético del procesador, el componente que más energía consume en un sistema de cómputo. ACPI define Estados de Potencia del Procesador (estados Cx), donde C0 es el estado de ejecución, y C1... Cx son estados de inactividad.

Un procesador que está en estado C0, también estará en un Estado de Rendimiento (estados Px). El estado P0 significa una ejecución a la máxima capacidad de rendimiento y demanda de potencia. A medida que se aumenta el número del estado P, se reduce su rendimiento y potencia demandada. Los procesadores implementan los estados P utilizando la técnica de Escalado Dinámico de Frecuencia y Tensión (DVFS, acrónimo de Dynamic Voltage and Frequency Scaling) [10]. Reduciendo la tensión suministrada se reduce el consumo energético. Sin embargo, se incrementa el retardo de las compuertas lógicas, entonces es necesario reducir la frecuencia de reloj de la CPU para que el circuito funcione correctamente. En ciertos procesadores multicore, se permite que cada núcleo esté en un estado P diferente. Normalmente, la técnica DVFS logra mejoras cuando una carga de trabajo está limitada por memoria, es decir, el procesador malgasta muchos ciclos de CPU esperando obtener datos desde la memoria.

Cuando no hay instrucciones para ejecutar, el procesador puede ser puesto en un estado C mayor a 0, para ahorrar energía. Existen diferentes niveles de estado C, donde cada uno de los niveles podría apagar determinados relojes, reducir ciertas tensiones suministradas a componentes ociosos, apagar la memoria cache, etc. Mientras más elevado sea el número del estado C, menor será la potencia

demandada pero mayor será la latencia requerida para regresar al estado C0 (estado de ejecución). Algunos procesadores permiten la elección de un estado C por núcleo.

Como ambos estados, C y P, tienen incidencia en la potencia y el tiempo, es necesario evaluar el impacto de ellos en el consumo energético durante operaciones de C/R

Software de Aplicación: Básicamente, un checkpoint consiste en guardar el estado de una aplicación, tal que, en caso de fallo, se pueda reiniciar la ejecución desde ese punto guardado. Cuanto mayor sea el tamaño del problema de la aplicación, mayor será el tiempo requerido para guardar su estado. Su incidencia, al menos en el tiempo, convierte al tamaño de problema en un factor que afecta el consumo energético de operaciones de C/R.

Software de Sistema: Hay dos tipos de software de sistema altamente involucrados en las operaciones de C/R. Por un lado, el sistema que se ocupa de realizar esas operaciones. Por otro lado, como el archivo de checkpoint requiere ser resguardado en almacenamiento estable y remoto, es necesario utilizar un sistema de archivos distribuído. En nuestro caso, el software de sistema que utilizamos es Distributed MultiThreaded CheckPointing (DMTCP)[3] y Network File System (NFS). Ambos softwares tienen opciones de configuración que inciden sobre el tiempo de ejecución y/o la potencia, y por lo tanto son factores que afectan el consumo energético de las operaciones de C/R.

En el caso de NFS, se permite montar carpetas de forma síncrona (opción *sync*) o asíncrona (opción *async*). Si una carpeta de NFS se monta con la opción *sync*, las escrituras en dicho punto de montaje causarán que los datos sean descargados completamente en el servidor NFS, y escritos en almacenamiento persistente, antes de retornar el control al cliente[1]. Así, el tiempo de ejecución de una operación de escritura se ve afectado al variar esta configuración.

En el caso de DMTCP, es una herramienta que realiza checkpoints de forma transparente sobre un grupo de procesos diseminados entre muchos nodos y conectados por sockets, tal como ocurre con aplicaciones MPI (*Message Passing Interface*). DMTCP es capaz de comprimir (utilizando el programa gzip) el estado de un proceso para requerir un menor espacio de almacenamiento en disco y reducir la cantidad de datos transmitidos por la red (entre el nodo de cómputo y el nodo de almacenamiento). El uso o no de la compresión impacta en el tiempo de ejecución y en la potencia requerida (al realizar un trabajo diferente), por lo tanto es otro factor que incide en el consumo energético.

4. Plataforma y diseño experimental

Plataforma experimental

Los experimentos se realizaron sobre un cluster de computadoras con una red Ethernet de 1 Gbps. Cada nodo, tanto de cómputo como de almacenamiento, posee 4 GB de memoria principal, un disco rígido SATA de 500 GB y 7200 rpm,

[1] https://linux.die.net/man/5/nfs

y un procesador Intel Core i5-750, con un rango de frecuencia de 1,2 GHz a 2,66 GHz (sin utilizar el mecanismo Intel Turbo Boost[2]), cuatro núcleos (sin multithreading), y 8 MiB de memoria caché. Los nodos utilizan el sistema operativo GNU/Linux Debian 8.2 Jessie (kernel version 3.16 de 64 bits), OpenMPI version 1.10.1 (implementación de MPI), y la herramienta de checkpoint DMTCP [3] versión 2.4.2. El sistema de archivos de red utilizado para hacer la escritura remota de los archivos de checkpoint es NFS v4.

Las mediciones de potencia se realizaron con el osciloscopio PicoScope 2203 (cuya presición es del 3 %), la sonda diferencial activa TA041, y la pinza de corriente PP264 60 A AC/DC, todos productos de Pico Technology. Las mediciones de tiempo de ejecución del checkpoint y del restart se realizaron utilizando la opción provista por DMTCP.

La aplicación seleccionada para la caracterización del sistema es una aplicación de transferencia de calor del tipo SPMD escrita en MPI que utiliza el tipo de dato float. Esta aplicación describe, mediante una ecuación, el cambio de temperatura en el tiempo, sobre un plano, dada una distribución de temperaturas inicial y ciertas condiciones de borde.

Diseño experimental

Se utilizan dos nodos de cómputo (salvo especificado lo contrario), y cada nodo de cómputo escribe en un nodo de almacenamiento diferente a través de un NFS configurado en modo asíncrono (salvo especificado lo contrario). La tasa de muestreo utilizada para ambos canales del osciloscopio se estableció en 1000 Hz. Las mediciones de potencia corresponden a la potencia disipada por el nodo completo incluyendo la fuente. El DMTCP está configurado para comprimir los archivos de checkpoint (salvo especificado lo contrario). Las pruebas se realizaron con la opción *c-states* del procesador activa (salvo especificado lo contrario). Para cambiar la frecuencia del procesador se utiliza el *governor userspace* de GNU/Linux, el cual permite modificar el archivo sysfs *scaling_setspeed* disponible para cada núcleo. Al modificar la frecuencia siempre se indica la misma para todos los núcleos.

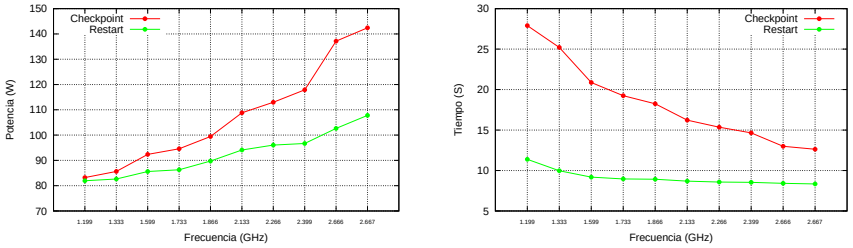
Cada experimento consiste en lanzar la aplicación con un proceso por núcleo, dejarla ejecutar durante un período de precalentamiento de 20 segundos, realizar un checkpoint manualmente, abortar la aplicación (desde el coordinador del DMTCP), y luego volver a reinicarla desde la línea de comandos con el script de restart generado por DMTCP. El tiempo precalentamiento se definió empíricamente observando cuándo la curva de potencia disipada llega a los valores máximos para esa aplicación. Cada experimento se repite tres veces (suficiente para la poca variabilidad de los instrumentos de medición utilizados) y se obtiene la media.

[2] https://www.intel.com/content/www/us/en/architecture-and-technology/turbo-boost/turbo-boost-technology.html

5. Experimentos y Análisis de Resultados

5.1. Estados P del procesador

Se evaluó el impacto de los estados P del procesador en el consumo energético de las operaciones de C/R. La figura 1 muestra la potencia disipada y el tiempo de ejecución para diferentes frecuencias del procesador. Se observa que a medida que aumenta la frecuencia de reloj, la potencia disipada aumenta y el tiempo de ejecución disminuye. Las funciones obtenidas son estrictamente crecientes para el caso de la potencia disipada y decrecientes para el caso del tiempo de ejecución. La diferencia de la potencia disipada y el tiempo de ejecución a la máxima y a la mínima frecuencia llega a ser de casi 60 W y 15 segundos para el checkpoint, y de casi 26 W y 3 segundos para el restart. También se puede observar cómo los cambios en la frecuencia de reloj afectan más a la potencia disipada y al tiempo de ejecución del checkpoint que del restart.



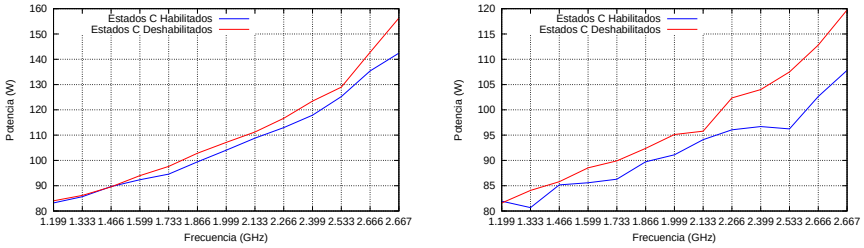
(a) Potencia disipada (b) Tiempo de ejecución

Figura 1: Incidencia de los estados P

5.2. Estados C del procesador

Durante la escritura o lectura de un archivo de checkpoint es posible que el procesador tenga momentos ociosos y por lo tanto sucedan transiciones entre los estados C que afecten la disipación de potencia del procesador. Para estudiar su comportamiento, se realizaron operaciones de C/R con los estados C habilitados y deshabilitados, para todas las frecuencias del procesador.

En la figura 2 se observa que las mediciones de potencia con los estados C habilitados muestran mayor variabilidad, especialmente en el restart. En este gráfico también se observa cómo la diferencia entre la potencia disipada con estados C habilitados y deshabilitados aumenta al aumentar la frecuencia del procesador. Esta diferencia llega a ser aproximadamente del 9 % para el checkpoint (en la frecuencia máxima), y del 10 % para el restart (en la frecuencia 2,533 GHz). Además, impacta en un mayor consumo de energía cuando los estados C están deshabilitados. Para el caso del checkpoint, el consumo llega a ser un 13 % mayor, y para el caso del restart, un 20 % mayor. Los tiempos de ejecución no mostraron variación al habilitar o deshabilitar los estados C.



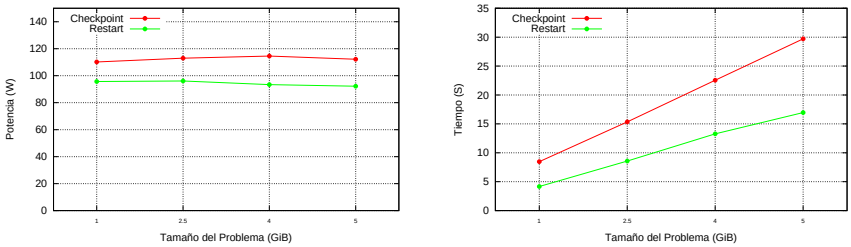
(a) Checkpoint (b) Restart

Figura 2: Incidencia de los estados C en la potencia

5.3. Tamaño del Problema

Se evaluó el impacto del tamaño del problema en el consumo energético de las operaciones de C/R. Se realizaron mediciones de la potencia disipada y el tiempo de ejecución, para diferentes tamaños de problema. No se usaron tamaños que superen la memoria principal disponible en el nodo de cómputo.

En el gráfico 3 se observa que la potencia disipada casi no varía al variar el tamaño del problema (no supera el 4 %), mientras que el tiempo de ejecución, como era de esperar, aumenta a medida que aumenta el tamaño del problema.



(a) Potencia disipada (b) Tiempo de ejecución

Figura 3: Incidencia del tamaño del problema

5.4. Configuración del NFS

Se evaluó el impacto en el consumo energético del uso de la opción *sync* o *async* en el montado del NFS. La figura 4 muestra gráficos que comparan la potencia disipada, el tiempo de ejecución y la energía consumida por un checkpoint almacenado sobre un sistema de archivos de red montado con la opción *sync* y *async*, para tres frecuencias de reloj diferentes (mínima, media y máxima del procesador). Para las tres frecuencias, la potencia disipada es mayor y el tiempo de ejecución es menor cuando se usa la configuración asincrónica. El menor tiempo se explica por el modo de funcionamiento del modo asincrónico, que no necesita esperar que los datos sean descargados en el servidor para

avanzar. La mayor potencia disipada puede deberse a que el modo asincrónico disminuye los tiempos ociosos del procesador. Para la frecuencia mínima y media, las diferencias son pequeñas, dando como resultado un consumo energético similar (un 1.5 % mayor para el caso asincrónico en la mínima frecuencia, y un 7 % menor para el caso asincrónico para la frecuencia media). En ambos casos, al ir mas lento el procesador del nodo de cómputo, no necesita detenerse tanto a esperar la respuesta del servidor NFS. En cambio para la máxima frecuencia esta diferencia se amplía, llegando a un ahorro energético de hasta un 25 % al usar el modo asincrónico.

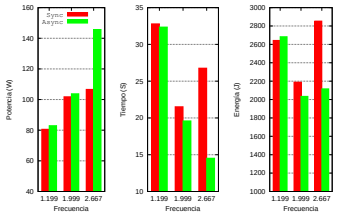
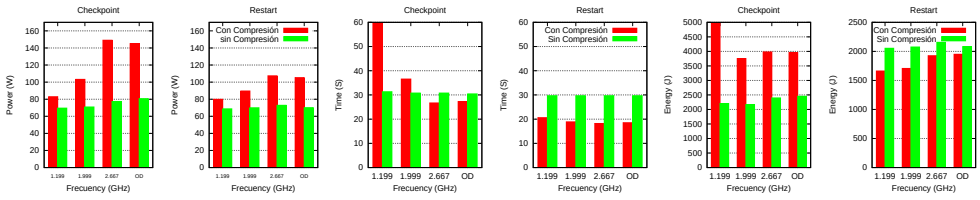


Figura 4: Potencia disipada, tiempo de ejecución y energía consumida por el checkpoint para tres frecuencias.

Debido a su menor consumo energético, y a la menor variabilidad en el tiempo de ejecución, es preferible almacenar los checkpoints en un NFS montado con la opción *async*.

5.5. Compresión de los archivos de checkpoint

Se analizó el impacto de la compresión de los archivos de checkpoint en el consumo energético del checkpoint y del restart. Los experimentos se realizaron sobre un único nodo de cómputo escribiendo en un nodo de almacenamiento.



(a) Potencia Disipada (b) Tiempo de Ejecución (c) Energía

Figura 5: Análisis del impacto de la compresión en el consumo energético.

La figura 5 muestra los gráficos que comparan la potencia disipada, el tiempo de ejecución y la energía consumida por el checkpoint y el restart, con y sin compresión de los archivos de checkpoint, para tres frecuencias de reloj diferentes (mínima, media y máxima disponibles en el procesador) y para el *governor ondemand*.

Los resultados obtenidos muestran lo siguiente:

- La potencia disipada, tanto en el checkpoint como en el restart, es mayor al usar la compresión. Esto se debe al mayor uso de la CPU que se requiere para ejecutar el programa de compresión (gzip) que utiliza DMTCP.
- Sin compresión, el tiempo de ejecución casi no varía para diferentes frecuencia de reloj, tanto para el checkpoit como para el restart.
- El consumo de energía al utilizar el *governor ondemand* es similar al obtenido al utilizar la máxima frecuencia.
- En el checkpoint, el consumo energético siempre es mayor al usar compresión (hasta un 55 % en la frecuencia mínima). Ocurre lo contrario para el restart, donde el consumo energético siempre es menor al usar compresión (hasta un 20 % para el tamaño analizado).
- El menor consumo energético del checkpoint se da para la frecuencia media (1,999 GHz) y sin usar compresión. El menor consumo energético del restartse da para la frecuencia baja (1,199 GHz) y usando compresión.

Teniendo en cuenta que en general se realizan más operaciones de checkpoint que de restart, es conveniente no utilizar compresión para reducir el consumo energético.

6. Conclusiones y trabajos futuros

Este trabajo muestra, a partir de una serie de experimentos sobre un cluster homogéneo ejecutando una aplicación SPMD, cómo impactan diferentes factores del sistema y de las aplicaciones, en el consumo energético del C/R coordinado utilizando DMTCP. Se estudió el impacto de los estados P y C del procesador. Se encontró que las operaciones de checkpoint son más sensibles a cambios en los estados P que las operaciones de restart. Por el contrario, los cambios de los estados C del procesador afectan más al restart. En la plataforma evaluada, el uso de los estados C permite ahorrar energía (hasta 13 % en el checkpoint y 20 % en el restart). El aumento en el tamaño del problema de la aplicación en estudio siempre resulta en un aumento en el consumo energético, debido a su alta incidencia en el tiempo de ejecución de las operaciones. Se encontró que al utilizar la máxima frecuencia de reloj se llega a obtener hasta un 25 % de ahorro energético al utilizar la configuración asíncrona de montaje del NFS. La compresión de los archivos de checkpoint es beneficiosa para el restart, registrando hasta un 20 % de ahorro energético al usar archivos comprimidos, mientras que impacta negativamente en la operación de checkpoint, llegando a tener hasta un 55 % más de gasto de energía al usar la mínima frecuencia de reloj. Entre los trabajos futuros se espera evaluar el impacto en el consumo energético de las operaciones de C/R, del modelo de programación de la aplicación paralela, y de otras herramientas de tolerancia a fallos, así como evaluar la energía consumida por el nodo de almacenamiento.

Referencias

1. Acpi - advanced configuration and power interface. Available at: http://www.acpi.info, accessed on 2018-07-09
2. Amrizal, M.A., Takizawa, H.: Optimizing Energy Consumption on HPC Systems with a Multi-Level Checkpointing Mechanism. Networking, Architecture, and Storage (NAS), 2017 International Conference on. IEEE, 2017 (2017)
3. Ansel, J., Arya, K., Cooperman, G.: DMTCP: Transparent Checkpointing for Cluster Computations and the Desktop pp. 1–12 (2009)
4. Bergman, K., Borkar, et al.: Exascale computing study: Technology challenges in achieving exascale systems. Defense Advanced Research Projects Agency Information Processing Techniques Office (DARPA IPTO), Tech. Rep 15 (2008)
5. Cui, X., Znati, T., Melhem, R.: Adaptive and power-aware resilience for extreme-scale computing. In: 2016 Intl IEEE Conferences on Ubiquitous Intelligence. vol. 00, pp. 671–679 (July 2016), doi.ieeecomputersociety.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0111
6. Dauwe, Daniel, e.a.: Optimizing checkpoint intervals for reduced energy use in exascale systems. Eighth International Green and Sustainable Computing Conference (IGSC). IEEE, 2017. (2017)
7. Diouri, Gluck, O., Lefevre, L., Cappello, F.: Ecofit: A framework to estimate energy consumption of fault tolerance protocols for hpc applications. In: 2013 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)(CCGRID)
8. Diouri, M., Gluck, O., Lefevre, L., Cappello, F.: Energy considerations in checkpointing and fault tolerance protocols. In: Dependable Systems and Networks Workshops (DSN-W), 2012 IEEE/IFIP 42nd International Conference on. pp. 1–6
9. Ibtesham, Dewan, e.a.: Coarse-grained energy modeling of rollback/recovery mechanisms. ependable Systems and Networks (DSN), 2014 44th Annual IEEE/IFIP International Conference on. IEEE, 2014. (2014)
10. Le Sueur, E., Heiser, G.: Dynamic voltage and frequency scaling: The laws of diminishing returns. In: Proceedings of the 2010 international conference on Power aware computing and systems. pp. 1–8 (2010)
11. Meneses, E., Sarood, O., Kale, L.V.: Assessing energy efficiency of fault tolerance protocols for hpc systems. In: Proceedings of the 2012 IEEE 24th International Symposium on Computer Architecture and High Performance Computing. pp. 35–42. SBAC-PAD '12, IEEE Computer Society, Washington, DC, USA (2012)
12. Mills, et al.: Energy consumption of resilience mechanisms in large scale systems. In: Proceedings of the 2014 22Nd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing
13. Mills, B., Grant, R.E., Ferreira, K.B., Riesen, R.: Evaluating energy savings for checkpoint/restart. In: Proceedings of the 1st International Workshop on Energy Efficient Supercomputing. pp. 6:1–6:8. E2SC '13, ACM, New York, NY, USA
14. Rajachandrasekar, R., et al.: Power-check: An energy-efficient checkpointing framework for hpc clusters. 2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing pp. 261–270 (2015)
15. Saito, Takafumi, e.a.: Energy-aware I/O optimization for checkpoint and restart on a NAND flash memory system. Proceedings of the 3rd Workshop on Fault-tolerance for HPC at extreme scale. ACM, 2013. (2013)
16. Shalf, J., Dosanjh, S., Morrison, J.: Exascale computing technology challenges. In: International Conference on High Performance Computing for Computational Science. pp. 1–25. Springer (2010)

Rendimiento del algoritmo AES sobre arquitecturas de memoria compartida

Adrián Pousa[1], Victoria Sanz[12], Marcelo Naiouf[1], and Armando De Giusti[13]

[1] III-LIDI, Facultad de Informática, Universidad Nacional de La Plata, Argentina
[2] CIC, Buenos Aires, Argentina
[3] CONICET, Argentina
{apousa,vsanz,mnaiouf,degiusti}@lidi.info.unlp.edu.ar

Resumen Actualmente AES (Advanced Encryption Standard) es uno de los algoritmos de cifrado simétrico más utilizados para encriptar información. El volumen de datos sensibles que se trasmiten en las redes se incrementa constantemente y cifrarlos puede requerir un tiempo significativo. Por lo anterior, es importante adaptar este algoritmo para aprovechar la potencia de cómputo de las arquitecturas paralelas emergentes. En este trabajo presentamos un análisis del rendimiento de AES sobre diversas arquitecturas de memoria compartida (multicore Intel E5-2695v4, Xeon Phi 7230 y GPU Nvidia GTX 960), para datos de entrada de distinto tamaño. Los resultados revelan que la GPU es la mejor alternativa para cifrar datos de entrada que no superan los 32MB. Sin embargo, para un volumen mayor de datos, el multicore alcanza el mejor rendimiento, seguido por el Xeon Phi.

Keywords: AES, cifrado de información, multicore, GPU, Xeon PHI

1. Introducción

En la actualidad, el volumen de datos que se trasmite por redes públicas se incrementa constantemente. En ocasiones esta información suele ser sensible y por esto es importante cifrarla. AES (Advanced Encryption Standard) [1] es uno de los algoritmos de cifrado simétrico por bloques más utilizados para encriptar información. AES es considerado como un algoritmo seguro por el gobierno de los Estados Unidos para protección nacional de información [2].

El proceso de cifrado y descifrado requiere de un tiempo de cómputo que dependiendo del volumen de datos puede ser significativo. Diversas investigaciones se han enfocado en reducir este tiempo, adaptando distintos algoritmos de cifrado para aprovechar la potencia de cómputo de arquitecturas paralelas emergentes.

Por un lado, varios autores [3,4,5] presentaron distintas estrategias para paralelizar AES sobre GPUs Nvidia, obteniendo buen rendimiento respecto a la versión secuencial provista por OpenSSL [6].

Por otro lado, otros autores paralelizaron AES sobre GPUs Nvidia y sobre CPUs multicore y compararon el rendimiento de ambas versiones. A continuación resumimos los resultados de algunos de estos trabajos.

Ortega y otros [7] paralelizan AES sobre multicore usando OpenMP y sobre GPU usando CUDA. Ejecutan ambos algoritmos sobre dos escenarios hardware: 1) multicore Intel i5 M450, GPU Nvidia Geforce GT 330M, 2) multicore Intel i7 960, GPU Nvidia Quadro FX 1800. Muestran que la GPU alcanza el mejor rendimiento en ambos escenarios para datos de entrada que no superan los 256MB.

Duta y otros [8] paralelizan AES (versión OpenSSL) con CUDA y OpenMP. Reportan que obtienen mejor rendimiento con la GPU para datos de entrada de hasta 1.4GB. Sin embargo, no describen las arquitecturas utilizadas.

Nishikawa y otros [9] paralelizan AES (OpenSSL) sobre un multicore (Intel Core i7-920) y dos GPUs Nvidia (Tesla C2050 y GTX 285) y muestran que las GPUs alcanzan el mejor rendimiento para datos menores a 256MB.

En [10] comparamos el rendimiento de dos versiones de AES: la primera realiza cómputo intensivo [11] y la segunda, basada en OpenSSL, reduce el volumen de cómputo reemplazando operaciones por accesos a datos pre-calculados almacenados en memoria (es decir, realiza acceso intensivo a memoria). Los resultados revelan que la segunda versión alcanza mejor rendimiento. Paralelizamos esta versión y la ejecutamos sobre una GPU Nvidia GTX 560TI y una máquina multicore compuesta por dos Intel Xeon E5405 (total 8 cores). Mostramos que el mejor rendimiento es alcanzado sobre la GPU para datos de entrada menores a 256MB.

Desde nuestro conocimiento, al momento no existen trabajos que realicen una comparación de rendimiento de AES sobre GPU, multicore y Xeon Phi. Asimismo, los trabajos de otros autores citados con anterioridad utilizan multicores de escritorio que no superan los 4 u 8 cores.

En este trabajo presentamos un análisis del rendimiento de AES sobre diversas arquitecturas de memoria compartida (multicore Intel E5-2695v4, Xeon Phi 7230 y GPU Nvidia GTX 960), para datos de entrada de distinto tamaño (hasta 2GB). Para esto, desarrollamos dos implementaciones paralelas de AES: AES-CUDA desarrollada con CUDA para ser ejecutada sobre GPUs Nvidia, y AES-OMP desarrollada con OpenMP para ser ejecutada sobre multicores y Xeon Phi. Los resultados revelan que la GPU es la mejor alternativa para cifrar datos de entrada que no superan los 32MB. Sin embargo, para un volumen mayor de datos, el multicore alcanza el mejor rendimiento, seguido por el Xeon Phi.

El resto del paper está organizado de la siguiente manera. La Sección 2 repasa las arquitecturas multicore y manycore actuales y sus características. La Sección 3 describe el algoritmo AES. La Sección 4 introduce nuestras implementaciones paralelas de AES para ser ejecutadas sobre multicore, Xeon Phi y GPUs Nvidia. La Sección 5 muestra los resultados experimentales. Por último, la Sección 6 presenta las principales conclusiones y algunas ideas para investigación futura.

2. Arquitecturas multicore y manycore emergentes

En la actualidad existen arquitecturas paralelas de memoria compartida muy diversas. Por un lado, las CPUs multicore incorporan decenas de núcleos com-

plejos que operan a alta frecuencia. Por ejemplo: las microarquitecturas Intel Broadwell e Intel Skylake (línea de escritorio i9) poseen 22 y 18 núcleos, respectivamente.

Por otro lado, las arquitecturas many-core tales como las GPUs y el Intel Xeon Phi integran gran cantidad de cores simples que operan a menor frecuencia. Ambas arquitecturas utilizan multithreading por hardware para acelerar los cambios de contexto [12,13]. En una CPU convencional el cambio de contexto es costoso ya que el SO debe almacenar en memoria el estado del programa (el contador de programa, el valor de los registros y la información del stack). Sin embargo, en las arquitecturas many-core el costo del cambio de contexto es menor debido al soporte hardware nativo: la arquitectura dispone de un banco de registros más amplio. Una aplicación hará uso efectivo de esta tecnología si soporta alto grado de paralelismo a nivel de hilo. Un gran número de hilos permite ocultar latencias de acceso a memoria y de operación.

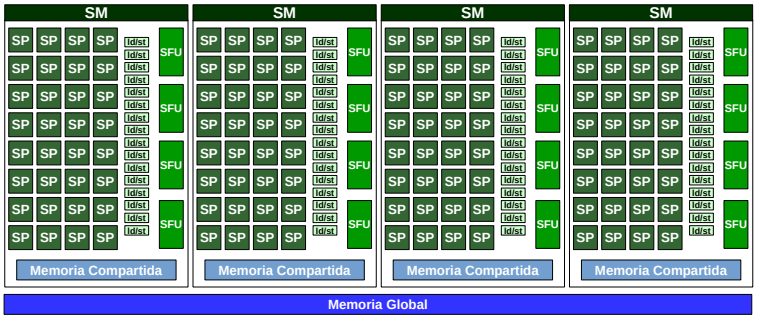


Figura 1. Estructura de una GPU Nvidia.

Las GPUs (Graphics Processing Units) han ganado importancia debido a que alcanzan alto rendimiento para aplicaciones de propósito general. Una GPU actúa como un coprocesador (*device*) conectado a una CPU convencional (*host*) vía PCIe. El cómputo sobre una GPU requiere realizar transferencias desde la memoria del host a la memoria del device (host-to-device o H2D) y viceversa (device-to-host o D2H). La Figura 1 muestra los componentes principales de una GPU Nvidia. En general, una GPU Nvidia está compuesta por un conjunto de multiprocesadores llamados Streaming Multiprocessors (SMs). Cada SM posee varios cores simples denominados Streaming Processors (SPs), un conjunto de unidades para resolver accesos a memoria (unidades load/store) y unidades para funciones especiales (SFUs). Asimismo, las GPUs poseen una jerarquía de memoria de varios niveles, que se diferencian en tamaño y latencia de acceso: la memoria global (mayor latencia), la memoria de constantes (latencia intermedia) y una memoria compartida ubicada en cada SM (menor latencia). Para

programar una GPU Nvidia es necesario utilizar extensiones al lenguaje C, como OpenCL o CUDA, y compilar los programas con un compilador específico.

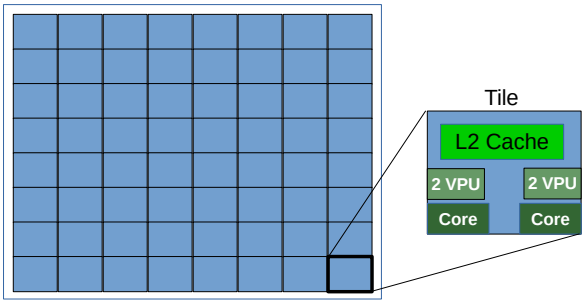


Figura 2. Estructura de un Xeon Phi de segunda generación.

El primer modelo de Xeon Phi conocido como Knights Corner se comercializó como coprocesador, por ello estaba sujeto a limitaciones tales como un tamaño de memoria limitado y podía utilizarse sólo como *device* conectado a un procesador *host* vía PCIe, requiriendo transferencias H2D y D2H. La segunda generación de Xeon Phi conocida como Knights Landing se comercializa además como procesador: puede utilizarse como una CPU independiente en lugar de ser sólo una placa complementaria, por esto no posee las limitaciones del modelo anterior. [14]

El Xeon Phi Knights Corner está compuesto por unos pocos procesadores basados en una modificación de la tecnología Intel Pentium Core (P54C) conectados a través de un bus en anillo. El Xeon Phi Knights Landing está compuesto por procesadores basados en una modificación de la tecnología Silvermont Atom en 14nm. La Figura 2 ilustra su estructura. Específicamente, este modelo está compuesto por varios *tiles* conectados a través de una interconexión en malla 2D. Cada tile posee dos cores, cada core está conectado a dos unidades de procesamiento vectorial (VPUs) AVX-512 y a una caché L2 compartida de 1MB. Cada core soporta cuatro hilos hardware (multithreading por hardware).

Los Xeon Phi son compatibles con procesadores Intel convencionales, es decir pueden ejecutar binarios creados para arquitecturas x86 o x86-64, sin requerir modificaciones en el código fuente. Asimismo, a diferencia de las GPUs Nvidia, no requieren utilizar extensiones a lenguajes ni compiladores específicos. Sin embargo, hacer uso del compilador propietario de Intel puede llevar a obtener un mejor rendimiento.

3. Algoritmo AES

AES (Advanced Encryption Standard) [1] es un algoritmo de cifrado por bloques, es decir los datos a cifrar se dividen en bloques de tamaño fijo (128 bits) y cada bloque se representa como una matriz de 4x4 bytes llamada *estado*, como se muestra en la Figura 3.

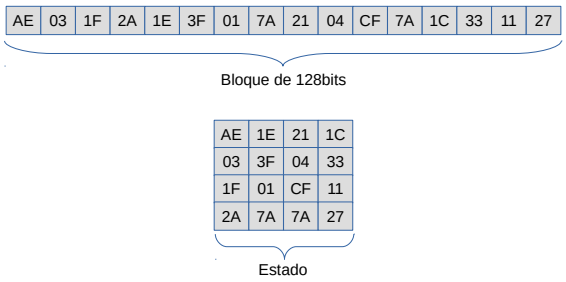


Figura 3. Estado AES

El algoritmo aplica a cada estado once *rondas*, cada una está compuesta por un conjunto de operaciones. Las once rondas se pueden clasificar en tres tipos: una ronda *inicial*, nueve rondas *estándar* y una ronda *final*. AES es un algoritmo simétrico, por esto utiliza la misma clave para cifrar y descifrar los datos, cuyo tamaño es de 128 bits, según lo indica el estándar. Esta clave es conocida como *clave inicial*, y a partir de ella se generan diez claves más mediante un procedimiento matemático. Las diez claves resultantes junto con la clave inicial son denominadas *subclaves* y cada una es utilizada en una de las rondas.

La ronda inicial realiza una única operación:

- **AddRoundKey:** realiza un XOR byte a byte entre el estado y la clave inicial.

Cada una de las siguientes nueve rondas (rondas estándar) aplican 4 operaciones en este orden:

- **SubBytes:** reemplaza cada byte del estado por otro de acuerdo a una tabla de sustitución de bytes con valores predeterminados. Este valor resultante es obtenido accediendo a la tabla tomando como índice de fila los primeros 4 bits del byte a reemplazar y como índice de columna los últimos 4 bits. El tamaño de la tabla es de 16x16 bytes.
- **ShiftRows:** rota cíclicamente a izquierda los bytes de las filas del estado de la siguiente manera: una vez en la segunda fila, dos veces en la tercera y tres veces en la cuarta. Notar que la primera fila del estado no se modifica.

- **MixColumns:** aplica a cada columna del estado una transformación lineal y la reemplaza por el resultado de esta operación.
- **AddRoundKey:** opera de la misma manera que en la ronda inicial pero utilizando la siguiente subclave.

La ronda final consiste de 3 operaciones:

- **SubBytes** y **ShiftRows:** operan de la misma manera que en las rondas estándar.
- **AddRoundKey:** opera de la misma manera que en las rondas anteriores pero utilizando la última subclave.

La Figura 4 muestra un esquema del algoritmo.

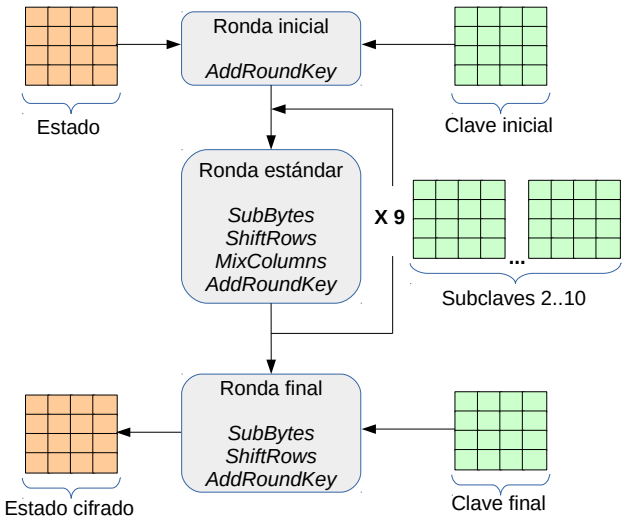


Figura 4. Rondas del algoritmo AES sobre un estado

4. Implementaciones AES

Todas nuestras implementaciones están basadas en el algoritmo AES secuencial provisto por la biblioteca de código abierto OpenSSL [6]. OpenSSL reúne un conjunto de herramientas de administración y bibliotecas relacionadas con criptografía, que proporcionan funciones criptográficas a distintos servicios (OpenSSH, HTTPS, etc).

Cabe destacar que no paralelizamos el proceso de generación de subclaves debido a que su tiempo de ejecución es despreciable.

4.1. AES-OMP

OpenMP es una API para los lenguajes C, C++ y Fortran que permite escribir y ejecutar programas paralelos utilizando memoria compartida. Esta API permite crear un conjunto de hilos que trabajan concurrentemente para aprovechar la potencia de cómputo de arquitecturas multicore y manycore Xeon Phi.

La implementación paralela de AES con OpenMP (AES-OMP) considera a los datos de entrada como bloques consecutivos de 16 bytes. Cada uno de los hilos toma proporcionalmente un conjunto de bloques consecutivos (estados) y aplica el proceso de cifrado.

4.2. AES-CUDA

El algoritmo AES sobre GPU (AES-CUDA) realiza los siguientes pasos. El host copia en la memoria de constantes del device las subclaves y la tabla de sustitución de bytes, ya que ambas son de sólo lectura; luego copia los datos a cifrar en la memoria global del device (transferencia host-to-device o H2D) e invoca al kernel.

Al igual que AES-OMP, AES-CUDA considera los datos de entrada como bloques consecutivos de 16 bytes (estados). Los hilos pertenecientes a un mismo bloque CUDA trabajan sobre estados consecutivos. Para esto, cooperan para cargar de manera eficiente la información a cifrar sobre la memoria compartida de alta velocidad. Luego cada hilo se encarga de cifrar un estado. Por último, los hilos cooperan para trasladar los datos desde la memoria compartida a la memoria global.

Por último, el host recupera los datos cifrados desde la memoria global del device (transferencia device-to-host o D2H).

5. Análisis experimental

Ejecutamos AES-OMP sobre una máquina con dos procesadores Intel Xeon E5-2695v4 y 128GB de RAM. Cada procesador cuenta con 18 cores que operan a 2.10GHz, por lo tanto la máquina posee 36 cores en total. Asimismo, ejecutamos AES-OMP sobre un Intel Xeon Phi 7230 (Knights Landing) y 128GB de RAM. El Xeon Phi posee 64 cores que operan a 1.30GHz y permite ejecutar 4 hilos hardware por core (256 hilos en total). Utilizamos gcc para compilar el código sobre estas plataformas.

Ejecutamos AES-CUDA sobre una GPU Nvidia Geforce GTX 960 compuesta por 1024 cores y 2GB GDDR5 de memoria. Los cores operan a una frecuencia normal de 1127Mhz. Generalmente, los algoritmos CUDA pueden alcanzar distinto rendimiento según la cantidad de hilos por bloque CUDA utilizada. Nuestra implementación obtiene el mejor rendimiento con 256 hilos por bloque CUDA.

Para realizar el análisis de rendimiento ciframos datos de distinto tamaño 16MB, 32MB, 64MB, 128MB, 256MB, 512MB, 1GB y 2GB. En este trabajo mostramos sólo los resultados del proceso de cifrado, ya que el proceso de descifrado presenta un rendimiento similar. Para cada escenario tomamos el tiempo de ejecución promedio de 100 ejecuciones.

Cabe destacar que nuestro análisis considera el tiempo de transferencia de datos (H2D y D2H) en AES-CUDA ya que representa una porción significativa del tiempo total de ejecución. Para demostrar esto, comparamos AES-CUDA considerando las transferencias H2D y D2H con AES-CUDA sin considerar dichas transferencias. La Figura 5 muestra el tiempo normalizado de ejecución de ambas versiones.

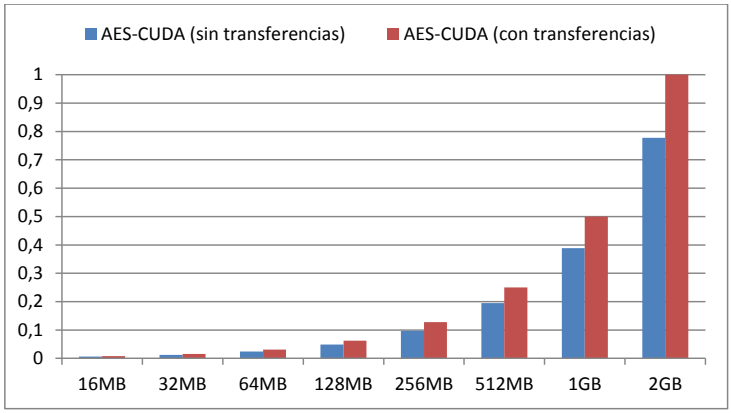


Figura 5. Tiempo de ejecución normalizado de AES-CUDA.

La Figura 6 muestra el Speedup[4] de AES-CUDA sobre la GPU con 1024 cores, AES-OMP sobre el Xeon Phi con 256 hilos y AES-OMP sobre el multicore Intel E5 con 36 hilos/cores. Calculamos el Speedup con respecto al algoritmo secuencial ejecutado sobre 1 core del Intel E5.

Los resultados revelan que el speedup alcanzado por AES-OMP sobre Intel E5 aumenta a medida que incrementa el tamaño de los datos de entrada. Lo mismo ocurre con el speedup alcanzado por AES-OMP sobre el Xeon Phi, aunque el factor de crecimiento es menor.

[4] El Speedup es definido como $\frac{T_s}{T_p}$, donde T_s es el tiempo de ejecución del algoritmo secuencial y T_p es el tiempo de ejecución del algoritmo paralelo.

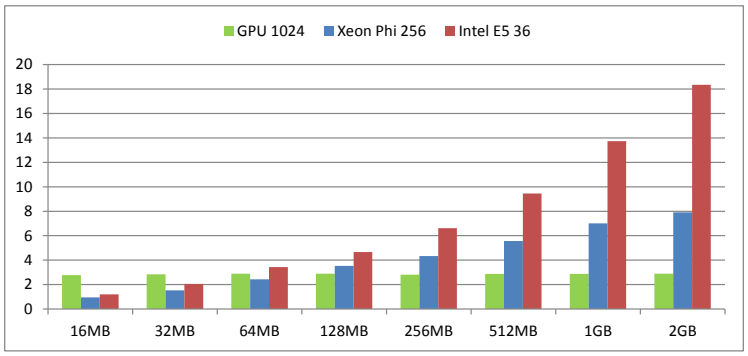


Figura 6. Speedup alcanzado por AES-OMP y AES-CUDA.

Por otro lado, los resultados muestran que el speedup de AES-CUDA no aumenta a medida que incrementa el tamaño de los datos de entrada. Esto se debe a que el tiempo de transferencia de datos entre CPU y GPU crece linealmente con el volumen de datos a transferir.

Como puede observarse, la GPU alcanza el mejor rendimiento cuando el tamaño de los datos a cifrar no supera los 32MB, seguido por el Intel E5 y el Xeon Phi. Para tamaños mayores, el mejor rendimiento lo obtiene el Intel E5, seguido por el Xeon Phi y la GPU.

6. Conclusiones y Trabajo Futuro

En este trabajo presentamos un análisis del rendimiento del algoritmo de cifrado simétrico AES sobre diversas arquitecturas de memoria compartida (multicore Intel E5-2695v4, Xeon Phi 7230 y GPU Nvidia GTX 960), para datos de entrada de distinto tamaño.

Para esto, desarrollamos dos implementaciones paralelas de AES basadas en el algoritmo secuencial provisto por la biblioteca OpenSSL: AES-CUDA desarrollada con CUDA para ser ejecutada sobre GPUs Nvidia, y AES-OMP desarrollada con OpenMP para ser ejecutada sobre multicores y Xeon Phi.

El análisis experimental revela que la GPU es la mejor alternativa para cifrar datos de entrada que no superan los 32MB. Para un volumen de datos mayor, el multicore alcanza el mejor rendimiento, seguido por el Xeon Phi.

Como trabajo futuro, planeamos comparar el rendimiento que proveen estas arquitecturas para aplicaciones con distintos perfiles de cómputo. Asimismo, planeamos utilizar estas arquitecturas de forma conjunta, como un sistema híbrido, para acelerar distintos tipos de aplicaciones.

Referencias

1. Federal Information Processing Standard Publication 197 (FIPS PUB 197): The Official AES Standard, http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf
2. Lynn Hathaway (June 2003): National Policy on the Use of the Advanced Encryption Standard (AES) to Protect National Security Systems and National Security Information, http://csrc.nist.gov/groups/ST/toolkit/documents/aes/CNSS15FS.pdf
3. Manavski, S. A. et al: CUDA compatible GPU as an efficient hardware accelerator for AES cryptography. In: Proc. of the 2007 IEEE International Conference on Signal Processing and Communications (ICSPC'07), pp. 65-–68 (2007).
4. Di Biagio, A. et al: Design of a parallel AES for graphics hardware using the CUDA framework. In: Proc. of the 2009 IEEE International Symposium on Parallel & Distributed Processing, pp. 1–8 (2009).
5. Iwai, K. et al: Acceleration of AES encryption on CUDA GPU. International Journal of Networking and Computing. 2 (1), pp. 131-–145 (2012).
6. The OpenSSL Project: Cryptography and SSL/TLS Toolkit, https://www.openssl.org
7. Ortega, J. et al: Parallelizing AES on multicores and GPUs. In: Proc. of the 2011 IEEE INTERNATIONAL CONFERENCE ON ELECTRO/INFORMATION TECHNOLOGY, pp. 1–5 (2011)
8. Duta, C.L. et al: Accelerating Encryption Algorithms Using Parallelism. In: Proc. of the 19th International Conference on Control Systems and Computer Science, pp. 549–554 (2013).
9. Nishikawa, N. et al: High-Performance Symmetric Block Ciphers on Multicore CPU and GPUs. International Journal of Networking and Computing. 2(2), pp. 251–268 (2012)
10. Pousa, A. et al: Comparación de rendimiento de algoritmos de cómputo intensivo y de acceso intensivo a memoria sobre arquitecturas multicore. Aplicación al algoritmo de criptografía AES. In: Actas del XXI Congreso Argentino de Ciencias de la Computación (CACIC 2015), pp. 163–172 (2015).
11. Pousa, A. et al: Performance Analysis of a Symmetric Cryptographic Algorithm on Multicore Architectures. In: De Giusti A, Diaz J (eds.) Computer Science & Technology Series. XVII Argentine Congress of Computer Science Selected Papers. pp. 57–66. EDULP, Argentina (2012)
12. Hyesoon, K. et al: Performance Analysis and Tuning for General Purpose Graphics Processing Units (GPGPU). Morgan & Claypool, (2012)
13. Rahman, R.: Intel Xeon Phi Coprocessor Architecture and Tools: The Guide for Application Developers. Apress, Berkeley, CA (2013)
14. Jeffers, J. et al: Intel Xeon Phi Processor High Performance Programming Knights Landing Edition. Morgan Kaufmann (2016)

CLOUD ROBOTICs: Navegación de un vehículo autónomo en un entorno con obstáculos

Manuel Costanzo[1], Marcos Boggia[1], Ismael Rodriguez[1], Armando De Giusti[1,2]

[1] Instituto de Investigación en Informática LIDI, Facultad de Informática, UNLP-CIC
[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
{mcostanzo, mboggia, ismael, degiusti}@lidi.info.unlp.edu.ar

Resumen. Este trabajo presenta el desarrollo de una plataforma WEB en el Cloud y el despliegue de un sistema Multi-Robot, compuesto por un chasis de auto Robot Rover 2WD, con sensores integrados, como así también, un Robot cámara aérea, que emula la captura de imágenes desde un cuadricóptero/drone, conectados al Cloud público de Amazon Web Services (AWS). Se detallan los prototipos desarrollados y el protocolo de comunicación utilizado; se mencionan los algoritmos implementados para el procesamiento y reconocimiento de imágenes, la simulación y determinación del camino óptimo, y la detección de obstáculos y espacios transitables, con el objetivo de que el vehículo no tripulado pueda llegar a un punto de destino de forma autónoma.

Palabras Clave: Auto Autónomo, Cloud Robotics, IoT, MQTT, Node-RED.

1 Introducción

Los avances en los paradigmas de Cloud Computing y Cloud Robotics han provocado un factor disruptivo de las TI en la industria tecnológica de sistemas Multi-Robot [1,2,3]. Estas tecnologías permiten tratar los datos de los componentes de hardware de un robot (sensores, actuadores/motores, cámaras, memoria, etc.), sin importar las limitaciones de cómputo y almacenamiento de las placas de desarrollo. En otras palabras, permite a los robots obtener resultados de tareas de cómputo intensivo, tales como: procesamiento de imágenes, determinación de rutas, confección de mapas, acciones cognitivas, etc., sin tratamiento local, sino en el Cloud [4].

Estos paradigmas brindan la capacidad de establecer escenarios para sistemas de Multi-Robot, donde cada robot se integra de un hardware mínimo con conectividad WiFi, donde los datos de los sensores y la adquisición de imágenes se procesan en el Cloud, y los actuadores de cada robot realizan las operaciones necesarias [5].

El propósito de investigación del presente trabajo es, el despliegue de un sistema de Multi-Robot, conectados al Cloud Público de AWS [6] con el fin de simular y ejecutar la navegación autónoma de un vehículo de 4 ruedas, como así también, la gestión de una cámara aérea de video, ambos conectados al Cloud, donde se realiza el procesamiento de imágenes y la planificación de caminos autónomos para que el vehículo efectúe la navegación a un punto de destino establecido.

Para ello, se determinó el método de comunicación con el Cloud; se registraron los robots en el servicio de AWS IoT [7]; se desarrollaron los algoritmos para

procesamiento y reconocimiento de imágenes, para simulación y determinación del camino óptimo al punto de destino establecido, y para detección de obstáculos y espacios transitables, con un bajo costo computacional.

A continuación, el presente trabajo se divide de la siguiente manera: en la sección 2, se introducen algunos conceptos elementales; en la sección 3, se presentan los prototipos confeccionados y sus componentes; en la sección 4, se describe el trabajo experimental realizado. Por último, la Sección 5, expone los resultados obtenidos y las líneas de trabajo futuras en relación a esta investigación.

2 Conceptos elementales

En esta Sección, se mencionan algunos conceptos empleados en la investigación realizada.

- **Amazon Web Services**

Es una plataforma de Cloud Público, que provee "Infraestructura como servicio"; a través de la tecnología de virtualización. De sus servicios más destacados, utilizamos "Elastic Compute Cloud" (EC2) [8] y "AWS Internet of Things" (AWS IoT).

- **Internet of Things (IoT)**

Es un nuevo paradigma que considera a los dispositivos como "objetos" conectados a Internet, utilizando protocolos de comunicación estándar [9].

- **Protocolo Message Queuing Telemetry Transport (MQTT)**

Es un protocolo de comunicación ligero [10,11], especialmente diseñado para tolerar conexiones intermitentes y reducir los requisitos de ancho de banda de red; soporta comunicación segura con TLS y calidad de servicio (QoS).

Implementa la comunicación de mensajes por medio de la publicación/suscripción sobre un canal de comunicación (tópico), como se puede observar en la figura siguiente:



Figura 1. MQTT: publicación/suscripción.

- **AWS IoT**

Es un servicio que proporciona AWS con el fin de conectar, administrar y operar grandes conjuntos de dispositivos [12]. Para cada dispositivo, se emite un certificado y un par de llaves privada-pública, que junto al certificado de la Entidad Certificante (CA), permite conectar el dispositivo al servicio de AWS IoT, en forma segura.

- **Node-RED**

Es una herramienta de programación visual, que permite programar algoritmos basados en flujos para IoT, sin la necesidad de escribir código. Provee de un editor de flujo basado en navegador WEB, que brinda una amplia paleta de nodos con diversa funcionalidad. El flujo se confecciona conectando los nodos entre sí. [13].

3 Prototipos y componentes

Para el presente trabajo se ha ensamblado un chasis de robot Rover de cuatro ruedas, emulando un vehículo tradicional de propulsión trasera, que cuenta con la capacidad de orientar las ruedas delanteras para la dirección del mismo. Se utiliza una placa micro-controladora de bajo costo y un módulo de expansión compatible, para controlar el motor, un sensor de ultrasonido y un módulo giroscopio para orientar al vehículo. Por otro lado, se realizó el despliegue de una cámara aérea que cuenta con una placa de desarrollo Raspberry Pi 3 y una cámara de video. A continuación se detallan los componentes:

Componentes	Descripción
Chasis Rover 2WD con servo S3003	Chasis compuesto de dos capas de acrílico, con un motor de 3 a 24v DC y velocidades con carga de 258 a 347 rpm/min, cuatro ruedas y un servo S3003 que permite direccionar el chasis con giros de hasta 45 grados. Dimensiones: 248mm (L) x 146mm (W) x 70mm (H).
Microcontroladora Raspberry Pi 3	Placa de bajo costo de desarrollo que brinda conexión WiFi, almacenamiento en memoria MicroSD de 32Gb y soporta alimentación de energía de 5v DC [14].
RaspiRobot Board V3	Módulo de expansión [15] compatible con Raspberry Pi 3, que brinda la interfaz de conexión entre la placa de desarrollo y los motores; incluye una fuente de alimentación conmutada para el suministro de energía, tanto de la placa Raspberry Pi como de los actuadores.
Sensor Ultrasónico HC-SR04	Sensor de distancias por ultrasonidos, capaz de detectar objetos y calcular la distancia a la que se encuentra, en un rango de 2 a 450 cm; cuantifica el delay de la señal, es decir, el tiempo que demora en regresar su eco.
Fuente conmutada Step-Down	Fuente de alimentación con regulador Step-Down DC-DC XL4015, que brinda una salida regulada de 1.25 a 36V.
Acelerómetro y Giroscopio de 3 ejes	Módulo Acelerómetro MPU, que posee un giroscopio de tres ejes con el que podemos medir velocidad angular, y un acelerómetro también de 3 ejes, con el que medimos los componentes X, Y y Z de la aceleración. Utilizamos el sensor de giroscopio para que el robot pueda posicionarse en la dirección correcta.
Raspberry Pi Camera Board 2	Pi Camera V2.1, es un módulo de video alta calidad de 8MP con sensor de imagen Sony IMX219 y un lente de foco fijo. Es capaz de tomar imágenes estáticas de 3280 x 2464 pixeles como así también, capturar vídeo de 1080p30, 720p60 y 640x480p90.

4 Trabajo experimental

El trabajo de investigación consistió en desarrollar una plataforma WEB que permite procesar, en el Cloud, los datos generados en un sistema de Multi-Robot que emula un vehículo de 4 ruedas y una cámara aérea de video, y planificar un camino autónomo para que el vehículo realice la navegación a un punto de destino.

El escenario está compuesto por el Robot Rover 2WD (RR2WD), obstáculos y un punto de destino; cada elemento es identificado por medio de una figura geométrica

(cuadrado, triángulo y circulo respectivamente). Quien se encarga de identificar cada componente es el Cloud, por medio de las imágenes provenientes del Robot cámara aérea. Una vez identificadas las figuras geométricas, se planifica el camino óptimo para que el vehículo se desplace hacia el punto destino sorteando los obstáculos.

A continuación, se describen los procedimientos para el procesamiento de imágenes y la determinación del camino óptimo.

4.1 **Procesamiento de imágenes**

El procesamiento de imágenes se realiza en una instancia de EC2 en el Cloud, utilizando la librería OpenCV [16] y lenguaje de programación Python [17]. Se implementó un algoritmo que permite detectar e identificar sobre una imagen de captura aérea, figuras geométricas. Luego de la detección, se extraen las coordenadas de ubicación del RR2WD (cuadrado), el destino (círculo) y los obstáculos (triángulo) con el propósito de utilizar estos datos en una posterior planificación del camino óptimo.

4.2 **Cámara aérea**

La cámara aérea emula la captura de video que puede realizar un drone. Dentro del escenario experimental se ubicó a una altura de tres metros sobre el nivel de la superficie del RR2WD, enfocando el escenario completo. Dicha cámara está integrada a una placa de desarrollo Raspberry Pi 3.

Cuando el Robot aéreo recibe una señal vía el protocolo MQTT, desde el Cloud, solicitando una captura de imagen, la misma se realiza en una resolución de 640x480 pixeles, se comprime y se codifica en base64, para por último enviarla vía MQTT a la instancia EC2, con el fin que el Cloud pueda realizar el reconocimiento y ubicación del RR2WD, los obstáculos y el punto de destino.

Todo el procesamiento mencionado se implementó con el lenguaje de programación visual Node-RED, como se muestra en la figura 2.

4.3 **Instancia EC2**

Cuando la instancia EC2, recibe una imagen, vía MQTT, se procede a identificar las figuras geométricas en la misma. Como ya se ha mencionado, se debe detectar un “cuadrado blanco con bordes negros” para identificar al RR2WD, la figura geométrica de un “círculo blanco con bordes negros” para identificar al punto de destino; y un “triángulo blanco con bordes negros” que identifica los obstáculos.

Una vez detectadas las figuras, se procede a calcular el camino óptimo entre el vehículo y el destino. Previo a este procesamiento, es necesario haber calibrado la cámara.



Figura 2. Flujo de envío de imagen.

4.4 **Detección de figuras geométricas**

Se implementó un algoritmo que detecta e identifica figuras geométricas dentro de una imagen. El proceso de detección consiste en transformar una captura habitual en una imagen en donde lo único que se pueda visualizar sean los vértices de las figuras en cuestión. Una vez transformada la captura, se identifican las figuras en función de la cantidad de bordes que tengan los objetos que sobresalen en la foto.

Cuando se recibe una captura aérea del escenario, se procede a transformar la imagen a escala de grises, como se muestra en la figura 3 y 4.



Figura 3. Imagen original.



Figura 4. Imagen en escala de grises.

Posteriormente, se suaviza la imagen como muestra la figura 5 y se aplica un umbral a la misma haciendo uso de la función de OpenCV "adaptiveThreshold" de la siguiente manera: "cv2.adaptiveThreshold (blurred, 255, 1, 1, 9, 5)", para tener como resultado una captura como se visualiza en la figura 6.



Figura 5. Imagen suavizada.

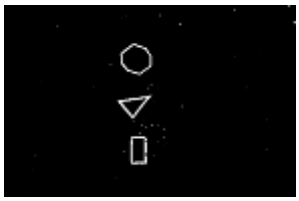


Figura 6. Imagen post umbral.

Por último, en función de los contornos detectados, se aproximan los mismos y se detecta la figura geométrica con respecto a la cantidad de vértices que posea el contorno. Si la cantidad de bordes de un contorno es igual a 3, se identifica al mismo como un triángulo, si la cantidad de vértices es 4, se trata de un cuadrado/rectángulo, y si es mayor a 4, se identifica como un círculo.

4.4.1 **Calibración cámara aérea**

Para lograr obtener una correspondencia entre el escenario real y el que muestra la imagen, se debe efectuar previamente una calibración de la cámara. Antes de la calibración se necesita conocer el ancho (en centímetros) de la figura geométrica del "cuadrado blanco con bordes negros" que representa al RR2WD.

El objetivo de la calibración es determinar la cantidad de centímetros contenidos en un pixel. Con esta referencia, se podrá conocer la distancia en centímetros entre cualquier par de píxeles, logrando así conocer la dimensión del escenario real.

La calibración se alcanza, calculando la distancia Euclidiana entre dos extremos de la figura geométrica del cuadrado, y se divide por el ancho de referencia de la figura anteriormente mencionada. Este procedimiento permite conocer cuántos pixeles forman un centímetro. Con esta información, se podrá calcular la distancia entre cualquier par de coordenadas (x,y) sobre la imagen.

4.5 **Planificación de camino óptimo**

Una vez encontradas e identificadas las figuras geométricas en la imagen, se procede a planificar el camino óptimo. Un camino está compuesto por un conjunto de movimientos que guiarán al vehículo autónomo desde la posición donde se encuentra, hacia el punto de destino, evitando que colisione con los obstáculos que se presenten. El RR2WD puede realizar movimientos de giro en grados a derecha e izquierda (de 0 a 180 grados y de 0 a -180, respectivamente), como así también, movimientos para avanzar una determinada distancia en centímetros.

La planificación del camino consta de 3 fases: Búsqueda del camino, optimización del camino y transformación del camino.

4.5.1 **Búsqueda del camino**

La técnica de planificación elegida es la denominada Informed RRT* (Informed Rapidly Exploring Random Trees Star) (IRRT*) [18]. Se implementó un algoritmo que se basa en la construcción de un árbol de configuraciones, que crece en forma aleatoria, con la búsqueda, a partir de un punto origen.

En la figura 7, se puede observar el resultado de simular la ejecución del algoritmo IRRT* tomando como punto de inicio la ubicación del RR2WD (punto verde), el punto de destino (punto azul) y los obstáculos de por medio (puntos rojos). Como resultado de la planificación se obtiene un vector, donde cada posición contiene un nodo que representa una coordenada (x,y).

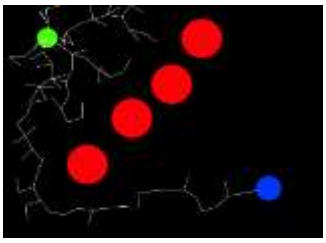


Figura 7. Simulación de la ejecución de la búsqueda del camino.

4.5.2 **Optimización del camino**

Luego de encontrar las coordenadas del camino desde el inicio al destino, se procede a optimizar el mismo. Este algoritmo itera por los nodos resultantes del procesamiento de búsqueda IRRT*, descartando aquellos puntos intermedios que no intersectan con un obstáculo, obteniendo como resultado un vector de coordenadas acotado.

En la figura 8 se puede observar el resultado luego de la optimización de la figura 7.

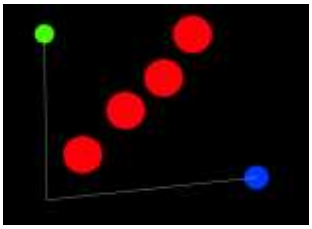


Figura 8. Resultado de la optimización del camino.

4.5.3 **Transformación del camino**

Para que el RR2WD pueda interpretar las acciones a realizar, se debe transformar el vector de coordenadas en movimientos. Cada movimiento consta de un grado de giro y una distancia en centímetros. Esta información se enviará al RR2WD para que realice la navegación autónoma hasta el punto de destino.

4.6 **Plataforma WEB**

Se desarrolló una plataforma WEB basada en “Node-Red-Dashboard” [19] que brinda una interfaz gráfica, la cual permite realizar la calibración de la cámara, como así también, visualizar el resultado de la ejecución del algoritmo de planificación de caminos simulado. Por otro lado, permite dar inicio al sistema de Multi-Robot, mostrando en tiempo real todas las acciones que se estén llevando a cabo con el RR2WD, los resultados de la planificación de la trayectoria, y la imagen capturada de la cámara aérea.

4.6.1 **Calibración**

Esta sección de la plataforma WEB, permite al usuario el ingreso de la información necesaria para calibrar las cámaras, por ejemplo: tipo de figura geométrica (“cuadrado” o “círculo”) y medida de ancho de la misma. Además, permite efectuar la calibración notificando con un mensaje de éxito o de error, según corresponda.

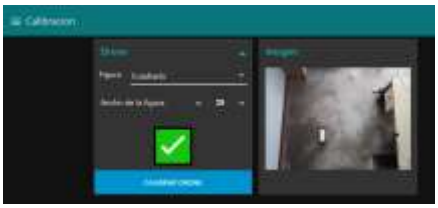


Figura 9. Interfaz de Calibración.

4.6.2 **Simulación**

Esta sección brinda la posibilidad al usuario de simular la planificación de un camino, permitiendo ingresar un punto de inicio, un punto de destino y diversos puntos de obstáculos, dando a elegir el radio y color de los mismos, junto con opciones para eliminar los obstáculos y para resetear la simulación. Además, permite la posibilidad de visualizar el resultado final luego de la optimización o la animación de cómo el algoritmo de planificación RRT fue creando el árbol de nodos.



Figura 10. Simulación con animación.



Figura 11. Simulación sin animación.

4.6.3 **Interfaz en Tiempo Real (Live)**

En esta última pantalla, se permite dar inicio y fin al sistema Multi-Robot, y visualizar en tiempo real la información que se está intercambiando y la situación en la que se encuentra cada actuador. Se puede apreciar gráficamente el resultado de la planificación del camino, y ver las imágenes que las cámaras están enviando a la instancia EC2 en el Cloud, como muestra en la siguiente figura:



Figura 12. Interfaz de gestión del sistema Multi-Robot.

5 Resultados obtenidos y líneas de trabajo a futuro

En esta investigación se ha implementado un sistema Multi-Robot conectado al Cloud público de AWS. Se han confeccionado los prototipos para emular un vehículo autónomo de 4 ruedas y una cámara aérea. Se han desarrollado los algoritmos que permiten realizar procesamiento de imágenes, reconocimiento de figuras geométricas y determinación del camino óptimo.

Además, se implementó una plataforma WEB para gestionar el sistema Multi-Robot, como así también, visualizar el resultado de simular la ejecución del algoritmo de planificación de camino óptimo IRRT*.

Uno de los avances más significativos de esta investigación consiste en que, actualmente, el vehículo de cuatro ruedas conectado a AWS, se puede orientar, desplazar y arribar al destino establecido, de forma autónoma evitando los obstáculos del escenario.

Concluimos y reafirmamos que Cloud Robotics es una tecnología que favorece la evolución de los sistemas Multi-Robot, tendiente a reducir muchas de las limitaciones que se presentan en los sistemas de robots tradicionales.

En cuanto a las líneas de trabajo a futuro, se pretende integrar un drone en reemplazo de la cámara aérea fija, cuya navegación se ejecute en forma autónoma. Por otro lado, se espera adicionar uno o más robots Rover 2WD y 4WD, adaptando la solución para emular una ciudad en la cual los sistemas Multi-Robot puedan trabajar colaborativamente.

Referencias

1. A. De Giusti, I. Rodríguez, M. Costanzo and M. Boggia, “Cloud Robotics: Auto Rover 4WD y Cuadricóptero controlados remotamente desde AWS” in CACIC ’17. Proceedings of XVIII Workshop de Procesamiento Distribuido y Paralelo (WPDP) – XXIII Congreso Argentino de Ciencias de la Computación. La Plata, Argentina. 2017.
2. Costanzo M., Boggia M., Rodriguez I., De Giusti A.: “CLOUD ROBOTICS: Vehículo autónomo conectado a AWS”. VI Jornadas de Cloud Computing & Big Data (JCC&BD 2018). La Plata, Argentina 2018.
3. L. Wang, M. Liu, M. Meng, R. Siegwart. “Towards Real-Time Multi-Sensor Information Retrieval in Cloud Robotic System”. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI). Hamburgo, Alemania, 2012.
4. J. Kuffner. “Cloud-enabled robots”. IEEE-RAS International Conference on Humanoid Robot. Nashville, USA, 2010.
5. L. Turnbull: “Cloud Robotics: Formation Control of a Multi Robot System Utilizing Cloud Infrastructure”. Proceedings of IEEE – Southeastcon. Jacksonville, USA, 2013.
6. I. Bermudez, S. Traverso, M. Mellia and M. Munafò. “Exploring the cloud from passive measurements: The Amazon AWS case”. Proceedings of IEEE – INFOCOM, 2013.
7. “AWS IoT” Available at: https://aws.amazon.com/es/iot. Accessed on 2018-06-06. [10] I. Bermudez, S. Traverso, M. Mellia and M. Munafò. “Exploring the cloud from passive measurements: The Amazon AWS case”. Proceedings of IEEE – INFOCOM, 2013.
8. “Tipos de instancias de Amazon EC2” Available at: https://aws.amazon.com/es/ec2/instance-types/. Accessed on 2018-06-06.
9. L. Atzori, A. Iera and G. Morabito. “The Internet of Things: A survey”. Journal Computer Networks ELSEVIER. Volume 54, Issue 15. 2010.
10. “Protocolo MQTT” Available at: http://mqtt.org. Accessed on 2018-06- 06.
11. “OASIS MQTT”, Available at: https://www.oasis-open.org. Accessed on 2018-06-06.
12. “AWS IoT” Available at: https://aws.amazon.com/es/iot/. Accessed on 2018-06-06.
13. “Node-RED” Available at: https://nodered.org. Accessed on 2018-06-06.
14. “The official site RapsberryPi” Available at: https://www.raspberrypi.org/products/raspberry-pi-3-model-b- plus/. Accessed on 2018-06-06.
15. “Raspirobot Board V3” Available at: https://www.monkmakes.com/rrb3/. Accessed on 2018-06-06.
16. “Open Source Computer Vision Library” Available at: https://opencv.org/. Accessed 2018-06-06
17. “Python Software Fundation” Available at: https://www.python.org/. Accessed on 2018-06-06.
18. Jonathan D. Gammell, Siddhartha S. Srinivasa and Timothy D. Barfoot. “Informed RRT*: Optimal Sampling-based Path Planning Focused via Direct Sampling of an Admissible Ellipsoidal Heuristic” arXiv:1404.2334v3 [cs.RO] 28 Nov 2014.
19. “Node-Red-Dashboard” Available at: https://github.com/node-red/node-red-dashboard. Accessed on 2018-06-06.

A Cloud Powered Relaxed Heterogeneous Distributed Shared Memory System

Matías Teragni[1] , Gonzalo Zabala[1], Sebastian Blanco[1]

[1] CAETI - Facultad de Tecnología Informática, Universidad Abierta Interamericana, Buenos Aires, Argentina
{Matias.Teragni, Gonzalo.Zabala, Sebastian.Blanco}@uai.edu.ar

Abstract. Distributed systems allow the existence of impressive pieces of software, but usually impose strict restrictions on the implementation language and model. We propose a distribution system model that enables the incorporation of any hardware device connected to the internet as its nodes, and places no restriction on the execution engine, allowing the transparent incorporation of any existing codebase into a Distributed Shared Memory.

Keywords: Cloud Computing, Distributed Programming, Distributed Shared Memory, Multi-Platform.

1 Introduction

Modern applications have unprecedented requirements both in the available resources to them and the capacities associated with the platform that host them, including scalability and availability. The way we have transcended the limits of the current hardware is by building distributed systems.

1.1 Distributed System

A distributed system is a collection of independent computers that appears to its users as a single coherent system [1]. We understand a computer as either a memory module, a processor module or a combination of both. To build an application that makes use of all the available hardware, we can consider two main approaches:

Operative Level support [2], which today can be found as a cloud farm, where the distribution is hidden from the programmer, restricting the control over it

Application Level support, where the distribution is provided by a framework or an execution program, like a cluster built using Hadoop[1], forcing the developer to use specific languages and tools.

A Distributed Shared Memory [3] is a way of building a distributed system, by creating a shared memory system between the nodes and including an abstraction layer (either by software or hardware) that makes it work as a huge shared memory system, bringing the advantages in simplicity [4] and development cost of a single computer execution, and the scalability and performance gains of a distributed system [5].

[1] http://hadoop.apache.org/

To achieve this goal a Distributed Shared Memory includes a set of algorithms, protocols, and guidelines that allows two or more computers to work as if their combined memory cells were available for both to use. This enables computer programs to exist across more than one computer, and if built correctly, take advantage of having more memory and available processors.

1.2 Memory Model

A memory model is the axiomatic formalization of the legal behaviors [6] regarding memory access provided by an execution engine (either a physical computer or a virtual machine) and can be used to detect anomalies [7] and ensure the correct execution of a program according to the specified semantics of the language used to define it. Studying these issues becomes fundamental when the desired code can be executed in parallel [8], because most anomalies and errors that can be produced by a faulty model will not happen under the strict serialization imposed by having a single processing unit, but even in the case of a single processing unit, if more than a single process is executing concurrently, then some anomalies can arise as well. In practice, once an anomaly is detected and analyzed, the correct serialization techniques can be applied to the code, for example by ensuring that the compiler inserts fences or locks when they are required to guarantee the semantics of the program.

The cost of synchronization.

In the specific case of a multicore system, the serialization is an expensive operation. If one must stop every processor while the state of a register is replicated to each node, then every advantage that comes with having the multicore in the first place is lost [9]. In the case of a Distributed system, like the ones behind Cloud Computing, the cost must include the network latencies, making it even worse, and other failure condition that may arise, such as network partitioning, where a subset of the nodes get disconnected to the others, and can cause a divergence in the distributed system. These extra circumstances are the root of the CAP theorem, which establishes a limit in guaranteeing consistency, availability and partition-tolerance of a distributed storage system [10].

Relaxed memory models.

To address these issues and take advantage of the current available hardware, relaxed memory models are being used. A relaxed memory model [11] is one that provides certain useful guarantees of the semantics but is not as strict as a complete serialization of the operations, allowing different execution paths to be correct, and diminishing or eliminating completely the synchronization cost.

Eventual Consistency.

One of the most popular relaxed models is Eventual Consistency [12]. This model guarantees that every node of the system will, eventually, have the same state, even in the presence of network partitions or node disconnections.

This is incredibly useful for distributed databases (specially NoSQL ones) and content distribution networks, where the mutation of the data is limited, and the

robustness of the program is not compromised by discarded intermediate values. The only thing that matters in those applications is the last state.

Causal Consistency.

Causal consistency [13] comes into play as stronger restrictions are applied to the order of the operations, which must guarantee that some subset of them cannot be executed out of order. This can ensure the semantic of a given operation or algorithm but requires synchronization between the nodes. It is still weaker than a complete sequential order, because all the operations that are not related by a direct cause can be executed in any order.

To provide causal consistency the program code must be analyzed to correctly determine the happens-before relation that restricts the execution re-ordering. This can be achieved in a statically manner, analyzing the source code, or in a more dynamical one by having the code be interpreted, and that relationship determined at runtime.

Both methods can be built on top of the eventual consistency [14], but require the use of a special compiler, or a specifically designed execution engine.

Relaxed Distributed Memory.

Building a distributed system often requires standardized and controlled hardware and network arrangements but using a relaxed model could allow the creation of distributed systems over a wide area network, and heterogeneous devices, like the internet. The issue preventing the adoption of weak models for general purpose applications is that they may force the developer to be aware and explicitly manage the distribution and synchronization forbidding the use of several programming languages not equipped with the synchronization tools required to guarantee the correct behavior of the program. It leaks the abstraction provided by the distributed system.

In the following section we propose a distribution model to support this, including the specific requirements leading to it. In section 3 we discuss an implementation prototype, including the design choices made during development, after which we expose future work and conclusions.

2 Proposed Distribution Model

To address these issues we propose building, on top of the guarantees provided by causal consistency, a distributed system platform composed by a series of libraries or modules that can be imported into an existing codebase, enabling the transparent incorporation of heterogeneous nodes, over an internet connection, simplifying the development process by allowing the use of a general-purpose programming language. Allowing in the process the inclusion of hardware such as mobile phones, or IoT devices, and the use of different programming languages to define the behavior of the different components of the system.

2.1 Requirements of The Distribution Model

Transparent.

Transparency in the context of a distributed system is defined as hiding the fact that its processes and resources are physically distributed [1].When a distributed system is built on top of an unmanaged network, nodes can connect and disconnect at any time. Handling these contingencies adds complexity to the code. To avoid it, it is desirable that the nodes can connect and disconnect transparently, and the information exchange between them to be performed almost anonymously. This allows any part of the system to request actions without worrying about exactly which node will perform them and provides fault tolerance to the system.

Local Access Time.

The principal performance benefit to use Message Passing instead of a Shared memory resides on data locality [4] [15]. To take advantage of data locality and cope with the unknown network characteristics without giving an unpredictably slow access time, we must sacrifice memory on the devices by building a cache of the shared data, guaranteeing a low access time by ensuring that the data is already present on every node of the network, and making any access to the shared data work as a local access to the device memory.

Heterogeneous.

Considering that the network which connects the nodes of the proposed distributed system is the internet, the nodes themselves can be hardware of any kind. Today we have a huge number of smartphones and IoT devices connected that could be potentially part of a distributed system and including that hardware could provide a huge opportunity to capitalize existing processing power.

Partitionable.

If the system should be able to incorporate different kinds of hardware, each with its own resource limitations, then the distributed shared memory cannot be a complete snapshot of the program, causing some of the incorporated hardware, like a smartphone, to crash due to insufficient memory. Therefore the synchronized state should be a part of the whole memory, and each device should be able to copy just the parts that it needs.

2.2 Replication Scheme

Given the requirement that the nodes are connected over a wide area network, considering that the nodes of our distributed system can be the huge number of IoT devices and smartphones out there, building a peer to peer replication scheme over the internet would not be practical. First, the network devices that separate the local area networks from the wide area networks filter broadcast messages, making it difficult to build a real peer to peer communication system over a wide area network. In addition to that, if every phone would have to process incoming messages sent by every other phone out there, the exponentially big number of messages would saturate the network infrastructure and would enqueue faster than the devices could process them, creating a new source of latency.

2.3 Data Hub replication

To control the replication, a few nodes on the proposed network have the specific job of replicating the state of the application to every other node. When any node needs to modify the shared memory space it informs the corresponding data node, and the data node ensures the correct replication to every other node in the network.

The data node can serialize modification requests if they collide and can ensure the consistency of the data stored in the shared memory. To achieve consensus a data node will always have the last word over the specific data it handles.

These specific nodes could be deployed on Cloud Servers, making them responsible for receiving the modification requests and broadcasting those changes to the connected devices, building a distributed shared memory between the several hundreds of heterogeneous devices connected to them, and taking advantage of the power and availability of the current cloud infrastructure.

Requirements of the cloud server.

Current cloud infrastructure allows us to handle several hundreds of incoming connections, and provide the basis for a successful replication scheme, but to get the proper functionality out of a distributed system built this way, there are two guarantees that need to be provided on top of the basic cloud functionality.

Low Replication Time.

One of the most difficult problems to solve correctly is when a collision and anomalies occur, mainly because exactly how the system should resolve it depends heavily on the semantics of the programming language and specific situation.

A collision in this case might be defined as two nodes trying to operate on the same piece of information at the same time. An anomaly is a behavior that is not consistent with the sequential execution of the program. Since the information takes time to propagate over the network, then these problems are exacerbated because they cannot arise only when two nodes act at the same time, but when two nodes act inside a window of time smaller than the time required to propagate the information between them.

If the propagation of information is slow, more collisions will take place, and the system will lose coherency, breaking the abstraction of the shared memory. The system, then, must reduce the amount and weight of the messages that circulate over the network, for example using a technology like web sockets instead of HTTP request-response cuts the overhead of the messages sent by removing the unnecessary http headers on each piece of data that is moved around.

Optimized Replication Scheme.

The replication protocol can introduce latency by itself, every message sent by the nodes must be processed by the server, and every message sent by the server must be processed by the clients. The main risk in this case is that if any node on the network cannot process the incoming messages in time, creating an ever-increasing queue of messages to process.

This can be tackled using two approaches: we can reduce the number of messages that are sent, and we can reduce the processing cost of each individual message. If the data hub takes the responsibility of broadcasting the changes, then the clients do not need to incorporate a polling system to obtain the latest modifications, reducing greatly the number of messages that the server must process.

If the server consolidates the changes instead of sending each change as an individual message, then the number of messages the client must process is greatly reduced.

3 Implementation prototype

3.1 Datahub implementation

A prototype of a Distributed Shared Memory between heterogeneous devices was built on top of a Firebase Database (functioning as a Data Hub using our replication scheme). The client side was built on JavaScript because it can be executed on any device that can run a web browser, serving as a fair test ground of a distributed system on heterogeneous devices over a wide area network.

Cloud server options.

Most of the cloud engines available today (including those provided by Amazon[2], Microsoft[3] and Google[4]) provide different services, from hosting complete virtual machines to websites, services, or functional programs. They usually include also some sort of storage, either a relational SQL database, a NoSql document-oriented database, or a file system to store the information generated by our cloud application.

If we wanted to use one of these traditional providers as our data hubs we would have to implement a program that take advantage of the available storage to persist the information and we would need to implement the handling of the modifications, collisions, and replications. This means that we would have to implement the whole functionality required of the Data Hubs in the language and format supported by that specific host.

On the other hand, real-time cloud databases like Firebase[5], or Cloudant[6] provide an interesting service in which you have a NoSql persistence scheme that can be accessed simultaneously by several hundreds of devices, with the objective of building a server-less application, where your clients know how to handle those changes. This service is intended, initially, for mobile and web applications, but can in theory be used from any platform, assuming the implementation of the correct binding.

The NoSql data that is synchronized takes the form of a JSON defined tree, where the specific semantics of the nodes in that tree are defined by the user of the service,

[2] https://aws.amazon.com/
[3] https://azure.microsoft.com
[4] https://cloud.google.com/
[5] https://firebase.google.com/
[6] https://www.ibm.com/cloud/cloudant

this means that if we want to interact with this database we must define the structure of the data that is contained by it, to allow our clients to be aware of the changes received from this service. Additionally, it usually has an integration with the other Cloud platform services, allowing us to implement any functionality that we required on top of the data synchronization on the server side, like conflict resolution, or garbage collection.

Using this service, we can avoid implementing most of the requirements of our Data Hubs, since the persistence and synchronization of the data is solved out of the box, we only need clients that handle those changes correctly.

The prototype was built using Google Firebase because, unlike its competition, it guarantees that any modification performed by any of the devices is synchronized to every other device connected in a matter of milliseconds thanks to the connection between the clients and the database uses streaming http and web sockets to avoid unnecessary overhead.

Distributed Shared Memory using Firebase.

Although the service provided by Google handles the replication of the data among the devices, it does not constitute a distributed shared memory, to achieve this functionality an abstraction layer must be built and deployed on each client. This layer must process every incoming message with modifications to the synchronized data model and apply the equivalent changes to the local memory converting every change received from the Cloud Database into the corresponding mutation of the memory of the device. This conversion is closely related to the semantics of the client programming language, and the nature of the mutation to the data model.

The other responsibility of the abstraction layer is to detect every change of the local memory and impact an equivalent modification into the synchronized data model. Detecting the changes is not necessarily an easy task, since polling every memory section to be shared is inviable, the only way left to gain this functionality is to intercept the changes of the client programming language by using the tools available in that context. This issue has been addressed by creating proxies that allowed us to detect the modifications without a major performance penalty.

3.2 Heterogeneous Node Implementation

Defined Data structure.

An example was provided where each device handles the movement of a rectangle on the screen, but the complete state is shown to all the connected nodes[7]. The data structure on this prototype is composed by two sets JSON elements, an Array containing all the rectangles, and an Object for each rectangle, containing floats for the X and Y positions, and a string for the hexadecimal color value. Once connected a computer will create a new rectangle of a random color and add it to the rectangle set, after which a node can only modify its own rectangle but has access to the whole collection.

[7] http://hiveproject.github.io/Firebase/Demo/Square/

Prototype Functionality.

The implementation assumes that all the memory structures are local and have no explicit reference to the distribution, as seen in the code examples below. The only exception is the first line, that obtains a reference to an array present in the shared memory.

The first function inserts a newly created rectangle object (composed by x, y and color) into an array, that happens to be shared, and subscribes to the event related to mouse movement a function responsible of updating the position of that specific rectangle to the coordinates of the mouse in that device.

```
squares=hive.get("SquareDemoPosition");
…
let pos = {x:50, y:50, color:"#"+myColor};
pos= squares[squares.push(pos)-1];
canvas.onmousemove = function(e){
  var mouseX, mouseY;
  if(e.offsetX) {
    mouseX = e.offsetX;
    mouseY = e.offsetY;
  } else if(e.layerX) {
    mouseX = e.layerX;
    mouseY = e.layerY;
  }
  pos.x=mouseX; pos.y=mouseY;
}
```

Once a local object is referenced by a shared object, then the local object becomes a shared one, enabling synchronization of its data to every other device connected.

The second function handles Drawing, by clearing the screen, and going through the array of rectangles, painting them on an HTML5 canvas. This function executes periodically with a low interval to refresh the display.

```
var ctx=canvas.getContext("2d");
function draw(){
  ctx.fillStyle = "black";
  ctx.fillRect(0,0,500,500);
  for (let k in squares)
  {
    if(squares[k]) {
      let current = squares[k];
      ctx.fillStyle=current.color;
      ctx.fillRect(current.x-5,current.y-5,10,10);

    }
  }
}
setInterval(draw,10);
```

Guaranteeing the Program Semantics

Since all the shared memory can be accessed by any node in the distributed system, there is a gain to be made with the available parallelism, but as a drawback the risk of data races is high, and unpredictable or undefined behavior may occur. Usually to solve this a compiler or an interpreter adds fences or locks to guarantee the correct synchronization of the shared memory, but since this prototype was intended as a library, it cannot modify the execution engine.

Instead this implementation provides the tools required to do a mutual exclusion to the developer, and with them a compare and set and other synchronization techniques can be built. According to the demonstration in [16], it is not possible to have stronger consistency than causal consistency having a partitioned system without sacrificing availability or endangering divergence. Therefore, availability will be sacrificed to provide the desired semantics, and a critical section will never trigger unless the requesting node is currently connected to the corresponding Data Hub.

The lock mechanism, like the one provided by Java and other general-purpose languages, can be accessed as a function in the synchronization library, and receives two arguments.

```
hive.lock(object,callback);
```

- A shared object over which the lock will be held
- A callback function that will be executed once the lock is acquired

This function is asynchronous, because a blocking function would not work on the single-threaded JavaScript execution and guarantees that the callback will be executed only if the calling node is the owner of the lock. Once the execution of the callback is completed, all the modifications to the shared state performed will be pushed, and the lock is released automatically. If a node is disconnected, the callback will never be executed.

4 Future Work

The developed model and prototype serve as a proof of concept, and several issues still need to be addressed. For now, at least, the prototype can only synchronize the shared state, but it may be worthwhile to explore the synchronization of behavior, being able to send complete working objects from one node to the others. An optimization is required in the synchronization engine to avoid unnecessarily replication of the parts of the shared memory that are not accessed by the local node.

The locking mechanism still has issues regarding deadlocks, and a disconnection of a node that owns a lock does not release it, making it unavailable to everyone.

5 Conclusions

This work has proposed a model that enables building a distributed system platform that integrates heterogeneous nodes over an unmanaged network, permitting the incorporation of an existing codebase because the model works as a library, and without imposing a programing language, paradigm, or platform. The viability of this model was tested by the construction of a prototype that exhibits the desired features. To guarantee the correct execution, the correct implementation is left to the programmer,

and tools were provided to address the need of critical sections in a code. Looking forward, more experiences need to take place to correctly determine the application area where these kinds of systems can be more beneficial.

References

1. Tanenbaum, A., van Steen, M.: Distributed Systems: Principles and Paradigms. Prentice-Hall, Upper Saddle River, NJ, USA (2006)
2. Tanenbaum, A., Van Renesse, R.: Distributed operating systems. ACM Computing Surveys (CSUR) 17(4), 419-470 (1985)
3. Protic, J., Tomasevic, M., Milutinovic, V.: Distributed shared memory: concepts and systems. In : IEEE Parallel & Distributed Technology: Systems & Applications (1996)
4. Ngo, T., Snyder, L.: On the influence of programming models on shared memory computer performance. In : Proceedings Scalable High Performance Computing Conference SHPCC-92, Williamsburg, VA, USA, USA (1992)
5. Nitzberg, B., Lo, V.: Distributed shared memory: a survey of issues and algorithms. In : Computer - IEEE Computer Society (1991)
6. Manson, J., Pugh, W., Adve, S.: The Java memory model. In : roceedings of the 32nd ACM SIGPLAN-SIGACT symposium on Principles of programming languages, Long Beach, California, USA (2005)
7. Kwiatkowski, J., Pawlik, M., Konieczny, D.: Parallel Program Execution Anomalies. In : Proceedings of Large Scale Computations on Grids, Wisla, Poland (2006)
8. Boehm, H.-J., Adve, S.: Foundations of the C++ concurrency memory model. In : Proceedings of the 29th ACM SIGPLAN Conference on Programming Language Design and Implementation, Tucson, AZ, USA (2008)
9. Kuznetsov, P., Ravi, S.: On the Cost of Concurrency in Transactional Memory. In : Principles of Distributed Systems. OPODIS 2011, vol. 7109 (2011)
10. Gilbert, S., Lynch, N.: Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. ACM SIGACT News 33(2), 51-59 (2002)
11. Boudol, G., Petri, G.: Relaxed memory models: an operational approach. In : Proceedings of the 36th annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages, Savannah, GA, USA (2009)
12. Burckhardt, S.: Principles of Eventual Consistency. Foundations and Trends in Programming Languages 1(1-2), 1-150 (2014)
13. Mosberger, D.: Memory consistency models. ACM SIGOPS Operating Systems Review 27(1), 18-26 (1993)
14. Lloyd, W., Freedman, M., Kaminsky, M., Andersen, D.: Don't settle for eventual: scalable causal consistency for wide-area storage with COPS. In : SOSP '11 Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles, Cascais, Portugal (2011)
15. LeBlanc, T., Markatos, E.: Shared Memory vs. Message Passing in Shared-Memory Multiprocessors. In : Proceedings of the Fourth IEEE Symposium on Parallel and Distributed Processing, Arlington, TX, USA, USA (1992)
16. Prince, M., Alvisi, L., Dahalin, M.: Consistency, Availability and Convergence. Technical Report, University of Texas, Austin (2011)

Análisis de consumo energético en Cluster de GPU y MultiGPU en un problema de Alta Demanda Computacional

Erica Montes de Oca[1], Laura De Giusti[1,3], Armando De Giusti[1,2] ,Marcelo Naiouf[1]

[1] Instituto de Investigación en Informática LIDI (III LIDI)
Facultad de Informática, Universidad Nacional de La Plata
Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires (CIC
La Plata, Buenos Aires, Argentina
[2] CONICET
[1] Investigador Asociado CIC
{emontesdeoca,ldgiusti,degiusti,mnaiouf}@lidi.info.unlp.edu.ar

Resumen. En este trabajo se realiza un análisis de consumo energético en dos Cluster de GPU y una MultiGPU utilizando como caso de estudio el problema de los N Cuerpos. Se describen las soluciones implementadas con MPI+CUDA para las arquitecturas usadas. Se muestran los resultados y un análisis de performance y consumo energético.

Palabras claves: GPU, Cluster de GPU, MultiGPU, N Cuerpos, Green Computing, Consumo Energético.

1 Introducción

Los profundos cambios generados por el avance tecnológico han modificado las formas de comunicación. En muchos casos, el resultado de este desarrollo ha traído aparejado un impacto ambiental. En las últimas décadas, Green Computing, ha emergido como un nuevo concepto para hacer frente a las preocupaciones sobre el cuidado del medio ambiente cuando las Tecnologías de la Información y la Comunicación están presentes [1].
En la actualidad, la generación de grandes cantidades de datos requiere reducir los tiempos de ejecución de las aplicaciones que los procesan. Sin embargo, acelerar el procesamiento ocasiona un mayor consumo de energía consumida, por lo que es necesario también minimizar esta variable [2]. Una opción de aceleramiento de cómputo son las GPU (Unidad de Procesamiento Gráfico) [3]. En los últimos años, se ha comenzado a tomar consciencia en el ámbito del software en desarrollar algoritmos más eficientes, no sólo desde el punto de vista de la performance, sino también del consumo energético [4][5]. Green Computing ya es un concepto instalado y adoptado. y las Ciencias de la Computación son las responsables de la investigación y promoción de sus prácticas [6][7].
En este trabajo, la Sección 2 introduce los conceptos básicos de Green Computing; la Sección 3, describe la simulación del problema de Atracción Gravitacional de los N

Cuerpos. La Sección 4 muestra los resultados obtenidos en el trabajo experimental. Por último, la Sección 5, presenta las conclusiones y trabajos futuros.

2 Green Computing

Desde la Revolución Industrial, el avance tecnológico alcanzado por el hombre inició la degradación del medio ambiente. El desequilibrio en los Gases de Efecto Invernadero (GEI) ocasiona los cambios climáticos actuales [8]. Los sectores generadores de las mayores emisiones de Dióxido de Carbono (el más conocido de los GEI) son: la generación de energía eléctrica, la industria, el transporte, el comercio y las residencias [9].
Green Computing comprende el uso eficiente de los recursos, la reducción del uso de materiales peligrosos, maximizar la eficiencia en la manufactura de productos, promover el reciclaje o biodegrabilidad de los productos y desechos fabriles [10].
El precio de la energía ha ido en aumento, por lo que la reducción del consumo energético es un punto clave. En este sentido, un sistema de cómputo no sólo debe ser escalable y reducir los tiempos de procesamiento, sino que debe ser eficiente en cuanto a consumo de energía, reduciendo los costos del gasto de consumo eléctrico y la disminución del impacto ambiental.

3 Arquitectura GPU

Las Unidades de Procesamiento Gráfico (GPU), comenzaron a ser utilizadas en las últimas décadas para cómputo de propósito general (GPGPU). Es una arquitectura paralela desde su nacimiento, y cada nueva generación doblan el número de cores. e incrementan la cantidad de operaciones de punto flotante por segundo. [11] [12].
La programación en GPU se facilitó a través del surgimiento de CUDA (Compute Unified Device Architecture) en 2007, por parte NVIDIA. CUDA es una plataforma hardware-software que le permite al desarrollador crear código en lenguaje C.
La arquitectura de una GPU-Nvidia está organizada en Streams Multiprocessors (SMs), los cuales tienen un determinado número de Streams Processors (SPs). Los SPs comparten la lógica de control y la caché de instrucciones. Nvidia [12] varía la cantidad de SMs en cada nueva generación logrando que las mismas sean escalables en performance, además de variar el número de DRAM para lograr escalar el ancho de banda de memoria y la capacidad. Cada SM provee los suficientes threads, cores y memoria shared para ejecutar uno o más bloques de threads CUDA. Los que realizan verdaderamente la ejecución son los SPs, que ejecutan múltiples threads concurrentemente. Los threads se organizan en bloques de hilos, y existen uno o más bloques ejecutándose concurrentemente en SM. La organización de los threads puede verse por nivel o jerárquicamente: Grid, Bloques y Threads (Fig. 1).

3.1 Cluster de GPU

Un Cluster GPU es una arquitectura heterogénea en la cual, por un lado, se tiene un subsistema conformado por varios cores CPUs y su sistema de memoria y E/S, mientras, por otro lado, se encuentra el subsistema GPU con sus memorias on y off chip. Estos dos sistemas se comunican a través de PCI-E con una velocidad de ancho

de banda baja comparada con las velocidades de comunicación de los sistemas de memoria. Por lo que, el cuello de botella es la comunicación que existe entre dichos subsistemas [14].

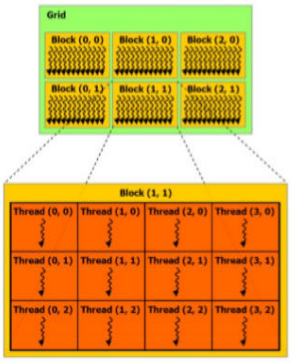


Fig. 1 Jerarquía de Threads de una GPU [13]

3.2 MultiGPU

La arquitectura MultiGPU se refiere al uso de una o más tarjetas gráficas en una misma PC. En este caso las GPUs son fácilmente programables a través de CUDA utilizando SLI (Scalable Link Interface) provista por Nvidia. En el caso de los Cluster de GPU, es necesario utilizar MPI+CUDA [15].

4 N Cuerpos: Problema de Alta Demanda Computacional

El problema de los N Cuerpos es clásico en el ámbito científico, y ha sido estudiado por su adaptabilidad a distintas aplicaciones del mundo real. Carente de una solución analítica, lo convierte en uno de los primeros problemas que se buscó resolver con una computadora [16]. Consiste en la simulación del comportamiento de N Cuerpos en un espacio de trabajo. En particular, el presente estudio se ha basado en el cálculo de la Atracción Gravitacional [17].
Newton planteó la *Ley de Gravitación Universal*, sentando las bases de la *Teoría de la Atracción Gravitacional*: cada cuerpo cuenta con una masa, una posición inicial y una velocidad. La gravedad hace que los cuerpos se aceleren y se muevan provocando la atracción de unos con otros. La magnitud de la fuerza de gravedad entre dos cuerpos se encuentra expresada en la fórmula (1):

$$F = \frac{G * m_i * m_j}{r^2} \quad (1)$$

Siendo m la masa, r la distancia, y G constante gravitacional (6,67 x 10-11)

3.1 Solución del problema en una GPU

Para realizar la simulación se utilizan vectores de tamaño N, que representan las fuerzas, posiciones, velocidades y masas. El algoritmo realiza las siguientes acciones:

1. Calcular la fuerza de atracción del cuerpo i, la cual se determina con los N-i cuerpos. A su vez, los N-i cuerpos se verán influenciados por la fuerza de gravedad de i.
2. Se modifica la posición del cuerpo i, dependiendo de su fuerza de gravedad.
3. Se repite 1 y 2, tantas veces como pasos de simulación se quieran realizar.

La independencia del cómputo del cálculo de la fuerza de atracción de cada cuerpo permite que se pueda ejecutar en una GPU. Básicamente, el algoritmo para una GPU realiza los siguientes pasos:

1. Desde la CPU se reserva memoria en la GPU;
2. Se transfieren los datos desde la CPU a la GPU;
3. Se ejecuta el *kernel*, que realiza el procesamiento del cálculo de las fuerzas de atracción gravitacional;
4. Se transfieren los datos procesados desde la GPU a la CPU.

3.2 Solución del problema en un Cluster de GPU

La ejecución en un Cluster de GPU requiere la utilización de una combinación de MPI con CUDA. Por cada proceso MPI creado se tiene asociada una GPU. Utilizando un esquema master/slave, el proceso master es el encargado de inicializar la información de los cuerpos a distribuir entre los procesos esclavos. Los *N* cuerpos se distribuyen en *P* procesos, por lo que cada proceso debe computar un total de *N/P* elementos. A su vez, cada proceso envia por medio de PCI-E dichos datos a sus respectiva GPU, siendo éstas las que realizan cómputo, ya que los procesos MPI se limitan a comunicar entre ellos los datos. Una vez que la GPU ha calculado la fuerza de atracción de los cuerpos que le corresponden, envía dichos datos a la CPU. Luego, el proceso MPI, comunica dichos datos al resto de los procesos [6] [18].

3.3 Solución del problema en una MultiGPU

En la solución con MultiGPU, se reparte en partes iguales la cantidad de datos a procesar por cada GPU. CUDA provee la función `cudaSetDevice(i)` (donde $i \leq T$, con *T* la cantidad de GPUs en una PC) para setear la afinidad de las GPUs de una PC [17]. El cálculo de la fuerza de atracción gravitacional de cada uno de los cuerpos que componen el espacio de trabajo se realiza del mismo modo que en la solución explicada para una GPU [19].

4 Trabajo Experimental

El entorno de prueba está compuesto por las siguientes máquinas:

- Cluster de multicore con procesadores Quad Core Intel i5-2310 de 2,9 GHz, caché de 6 MB: cada nodo cuenta con una GPU GeForce TX 560TI. CUDA 8.0
- Cluster de multicore con procesadores Quad Core Intel i5-4460 de 3.2 GHz, caché de 6 MB: cada nodo cuenta con una GPU GeForce GTX 960. CUDA 8.0
- Una PC con un procesador Quad Core Intel i5-2310 de 2,9 GHz, con dos GPU Tesla C2075. CUDA 4.2

Todas las pruebas de la simulación se hicieron para uno y tres pasos. La cantidad de cuerpos (N) se varió para 128000, 256.000 y 512.000. En las pruebas se utilizaron bloques de threads de 256 hilos para todas las GPUs. Los datos obtenidos son el resultado de un promedio de diez ejecuciones del algoritmo y mediciones de consumo energético. Se plantó un lote de escenarios conformados por los tres tamaños de entrada, repitiendo la simulación diez veces.
Para realizar las mediciones se utilizaron una pinza amperimétrica (para la medición de corriente, con una sensibilidad 1A/100mv, 1A/10mv y 1A/1mv) y un transformador (que censa la tensión existente en la entrada del suministro de energía). La tensión se midió directamente de la línea eléctrica a la cual se encuentra conectado el Cluster de multicore y la MultiGPU. Ambos dispositivos se conectan a los canales de entrada del Osciloscopio (Rigol DS1074Z, que tiene una tasa de muestreo de un 1GSa/s [20]). Para la medición de consumo realizada se utilizaron 2 canales (uno para la corriente y otro para la tensión), almacenando 12Mpts en memoria.
Para analizar las muestras obtenidas para cada escenario planteado, se modificó una aplicación de análisis de consumo energético denominada energyAnalyser [21].

A partir de los datos de la corriente y la tensión extraídos del osciloscopio, se realiza el cálculo de la potencia instantánea (2).

$$P(t) = V(t) * I(t) \quad (2)$$

siendo V el voltaje e I la corriente
Una vez obtenida la potencia instantánea, se calcula la potencia promedio, y la cantidad de Joules consumidos por la ejecución del algoritmo de los N Cuerpos.

En la Tabla 1, se presentan los tiempos de ejecución para la simulación con tamaño de entrada N: 128000, 256000 y 512000, para uno y tres pasos de la simulación, para una GPU GeForce GTX 960, una Tesla C2075, una GeForce GTX 560, Cluster de GPU y la MultiGPU.

Tabla 1. Tiempos de ejecución de la simulación de los N Cuerpos medidos en segundos utilizando una GPU, Cluster con dos GeForce GTX 960, una MultiGPU con dos Tesla C2075 y un Cluster de dos GeForce GTX 560.

Cantidad de cuerpos	GeForce GTX 960		Tesla C2075		GeForce GTX 560	
	1 Paso	3 Pasos	1 Paso	3 Pasos	1 Paso	3 Pasos
128000	5,23	13,54	5,36	16,1	6,42	18,91
256000	16,16	49,33	19,66	63,9	24,87	75,68
512000	64,64	194,58	78,64	252,87	99,08	297,45
Cantidad de cuerpos	Cluster de dos GeForce GTX 960		MultiGPU con dos Tesla C2075		Cluster de dos GeForce GTX 560	
	1 Paso	3 Pasos	1 Paso	3 Pasos	1 Paso	3 Pasos
128000	3,69	7,52	2,69	8,08	3,46	9,93
256000	9,42	26,83	10,74	32,21	12,48	37,44
512000	35,18	130,59	42,59	85,18	49,75	149,44

Puede notarse que la GeForce GTX 960 es la que ejecuta en un menor tiempo el problema. Sin embargo, cualquiera de las tres GPU, consiguen acelerar el cómputo

del algoritmo desarrollado. En trabajos previos, puede observarse la aceleración significativa que se obtiene utilizando GPU comparada con las versiones en memoria compartida, memoria distribuida y memoria híbrida [3][6].
Los algoritmos paralelos utilizan como medida de performance el *speedup* para medir el comportamiento del aumento de la cantidad de procesadores comparado con utilizar sólo uno. Las arquitecturas empleadas en este trabajo son inherentemente paralelas [22], por lo tanto, esta métrica no es la más adecuada para medir su rendimiento. Por esto se define el concepto de *Aceleración* (3), como métrica para medir performance al utilizar más de una GPU [6].

$$Aceleración = \frac{Tiempo_1_GPU(N)}{Tiempo_ + _GPU(N)} \quad (3)$$

Para el caso de la MultiGPU con dos Tesla C2075, se puede observar en la Fig. 2 que se obtiene una aceleración superlineal. Esto se debe generalmente a la división de los datos: teniendo N datos a procesar con una GPU, cuando se trabaja con dos, la división de los datos es N/2, es decir se tiene una menor cantidad de datos a procesar en cada GPU, obteniéndose un resultado superlineal.
Actualmente, acelerar el cómputo, teniendo en cuenta sólo performance en la escalabilidad de los sistemas no es suficiente. Un sistema que brinde performance pero que consume demasiada energía, no es sostenible en el tiempo ni rentable económicamente. Por esto, eEl estudio del consumo energético de las aplicaciones se ha tornado un punto importante.
Se define la potencia promedio de energía como la cantidad de energía promedio entregada o absorbida por un elemento. Lo que implica que la cantidad de energía necesaria para procesar un dato es la misma independientemente del dato que sea y la cantidad de datos a procesar. La Tabla 2, presenta las potencias promedio de energía medidas en Watts por segundo para el uso de una GPU, los Cluster de GPU y MultiGPU al ejecutar el algoritmo de los N Cuerpos.

Tabla 2. Potencia promedio de energía medida en Watts por segundo para la simulación de los N Cuerpos utilizando una GPU, un Cluster con dos GeForce GTX 960, una MultiGPU con dos Tesla C2075 y un Cluster de dos GeForce GTX 560.

Cantidad de cuerpos	GeForce GTX 960		Tesla C2075		GeForce GTX 560	
	1 Paso	3 Pasos	1 Paso	3 Pasos	1Paso	3 Pasos
128000	160,48	163,75	276,74	279,26	191,87	196,49
256000	160,91	166,07	276,96	279,26	193,01	203,28
512000	162,36	164,49	278,51	278,17	202,77	208,97
Promedio	163,01		278,18		199,39	
Cantidad de cuerpos	Cluster de dos GeForce GTX 960		MultiGPU con dos Tesla C2075		Cluster de dos GeForce GTX 560	
	1 Paso	3 Pasos	1 Paso	3 Pasos	1 Paso	3 Pasos
128000	306,85	307,6	376,4	385,71	382,29	383,78
256000	315,62	320,42	375,94	381,4	390,88	392,33
512000	321,13	322,89	378,39	388,37	394,64	406,63
Promedio	315,76		381,08		391,75	

Como se puede observar en la Tabla 3, la potencia promedio de energía no varía significativamente. Las variaciones dadas pueden deberse a los movimientos de los datos en la jerarquía de memoria de la GPU o a la refrigeración, entre otras. Por lo cual, se puede calcular una media de las potencias promedio de energía medidas y obtener una potencia media de energía que caracterice a cada GPU y Cluster de GPU/MultiGPU.
En este punto, no es posible hacer una valoración precisa en cuanto al consumo energético de cuál arquitectura es conveniente para ejecutar la simulación propuesta. Esto se debe a la forma de medición disponible, que permite medir "desde afuera" de la máquina, por lo que no sólo se está obteniendo el consumo energético de la GPU, sino también del resto de los componentes de la PC. Para poder comparar las arquitecturas, se realizaron mediciones en modo de reposo, para obtener de esta manera la potencia estática de cada una. En la Tabla 3, se pueden observar las potencias estáticas en Watts por segundo para las arquitecturas utilizadas en la experimentación.

Tabla 3. Potencia estática medida en Watts por segundo para una PC o para un Cluster con dos GeForce GTX 960, una MultiGPU con dos Tesla C2075 y un Cluster de dos GeForce GTX 560.

GPU	Una PC	Cluster/MultiGPU
GeForce GTX 960	69,4	120,2
Tesla C2075	132	132
GeForce GTX 560	97,6	122,1

En la Tabla 4, se pueden observar las potencias promedio de energía sin la potencia estática en Watts por segundo para el uso de una GPU, el uso de los Cluster de GPU y MultiGPU.

Tabla 4. Potencia promedio de energía sin la potencia estática en Watts por segundo para la simulación de los N Cuerpos utilizando una GPU, un Cluster con dos GeForce GTX 960, una MultiGPU con dos Tesla C2075 y un Cluster de dos GeForce GTX 560.

Cantidad de cuerpos	GeForce GTX 960		Tesla C2075		GeForce GTX 560	
	1 Paso	3 Pasos	1 Paso	3 Pasos	1 Paso	3 Pasos
128000	91,01	94,28	144,74	147,45	94,22	98,84
256000	91,44	96,6	144,96	147,26	95,36	105,63
512000	92,89	95,02	146,51	146,17	105,12	111,32
Cantidad de cuerpos	Cluster de dos GeForce GTX 960		MultiGPU con dos Tesla C2075		Cluster de dos GeForce GTX 560	
	1 Paso	3 Pasos	1 Paso	3 Pasos	1 Paso	3 Pasos
128000	186,64	187,35	244,4	253,7	260,27	261,76
256000	178,97	180,44	243,94	249,4	268,86	270,31
512000	185,45	183,6	246,39	256,37	272,62	284,61

Finalmente, la Fig. 3 presenta la energía en Joules consumida por la simulación utilizando una GPU.

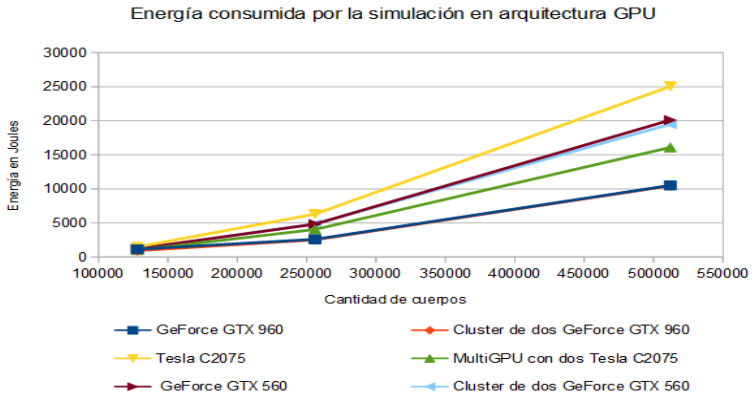


Fig. 3. Energía en Joules consumida por la simulación para cada arquitectura GPU

Puede observarse que para el caso de las GeForce la energía consumida es la misma, pero en menor tiempo de trabajo al agregar un nodo. En cambio en la MultiGPU, trabajar con dos tarjetas gráficas reduce el tiempo de cómputo, pero desde el punto de vista energético, el consumo es mayor.

5 Conclusiones y Trabajos Futuros

En la actualidad, el procesamiento de grandes cantidades de datos, el alcance de la máxima cantidad de transistores por microchip, y la demanda de la reducción de los tiempos de repuesta de las aplicaciones, han hecho que el procesamiento paralelo sea, en muchos casos, la única solución para tratar algunos problemas computables [18].
Las GPU permiten una reducción significativa de los tiempos de procesamiento, además de un menor costo y la facilidad de adaptar algunos problemas a estas arquitecturas, lo que ha hecho que sean una opción altamente viable a la hora de elegir una arquitectura paralela. Sin embargo, la aceleración del cómputo trae aparejada una problemática: la cantidad de energía consumida. El uso de la computadora, sus accesorios y sus recursos está catalogado en la actualidad como uno de los responsables del calentamiento global. El diseño de todos los sistemas de cómputo, desde dispositivos móviles hasta los sistemas de cómputo de altas prestaciones, está siendo impactado por el consumo energético.
En el presente trabajo se ha mostrado el consumo energético de tres diferentes tarjetas gráficas al ejecutar una solución del problema de Alta Demanda Computacional (N Cuerpos). Por un lado, utilizando una GPU de manera individual, y por otro, con dos Cluster de GPU y una MultiGPU.
A partir de las mediciones de performance y consumo energético presentadas en el apartado anterior, se concluye que:

- Referido al tiempo de ejecución, la GPU GeForce GTX 960 es la que ejecuta en un menor tiempo el problema. Sin embargo, cualquiera de las tres GPU, consiguen acelerar el cómputo del algoritmo desarrollado.
- La aceleración obtenida en cualquiera de las tres GPU utilizadas se encuentra cerca de la aceleración óptima.

- Además, se pudo comprobar lo demostrado en [23], que la cantidad de Watts promedios consumidos por una GPU, un Cluster de GPU o una MultiGPU, es independiente del tamaño de la entrada, y la variación que se observa en las mediciones es insignificante. Dicha variación puede deberse a los movimientos de datos y/o a la refrigeración de las placas.

  A partir de los resultados obtenidos, para el problema de los N Cuerpos, y con los escenarios experimentales llevados a cabo, se puede concluir que el uso de una GPU GeForce GTX 960 o de un Cluster de GPU GeForce GTX 960 consigue tanto en performance como en consumo energético las mejores prestaciones. Si se tienen en cuenta los costos de las distintas GPU utilizadas, la mejor opción sigue siendo la GeForce GTX 960. La placa que menos consume en líneas generales es la GeForce 960 para todos los tamaños del problema probados.

Como trabajos futuros se puede mencionar:

- Proponer un modelo de estimación de energía para el problema planteado, que permita predecir la cantidad de energía consumida al agregar GPU o al incrementar el tamaño de la entrada.
- Utilizando el mismo algoritmo de N cuerpos, medir consumo energético en más de dos nodos para los Cluster de GPU con GeForce GTX 560 y GeForce GTX 960.
- Estudiar y utilizar NVIDIA Management Library (NVML), que permite monitorear y administrar varios estados de los dispositivos NVIDIA GPU. Proporciona un acceso directo a las consultas y comandos expuestos a través de nvidia-smi. Entre las consultas que se pueden realizar, se encuentra la administración de la energía en la placa. La misma está disponible para la Tesla C2075 (no así para las otras placas: GeForce GTX 560 y 960). La consulta devuelve como resultado la potencia instantánea consumida por la tarjeta gráfica.
- Realizar análisis de consumo energético referidas a otro tipo de simulación, tales como simulaciones orientadas a individuos.

Referencias

1. Porcelli A.M., Martinez A.N.: La nueva economía del siglo XXI: análisis del impacto de la informática en el ambiente. Tendencias actuales en tecnologías informáticas verdes, un compromiso con la sustentabilidad. Questio Iuris vol. 8 nro 4 pp. 2174-2208. (2015)
2. Francis K., Richardson P.: Green Maturity Model for Virtualization. The Architecture Journal. pp. 9-15 (2008)
3. Montes de Oca E., De Giusti L., De Giusti A., Naiuof M.: Comparación del uso de GPU y cluster de multicore en problemas con alta demanda computacional. XI Workshop de Procesamiento Distribuido y Paralelo. CACIC 2012. Bahía Blanca, Buenos Aires, Argentina. (2012)
4. Baladini J., Morán M., Rozas C., De Giusti A., Suppi R., Rexachs D., Luque E.: Consumo energético de sistemas de cómputo de alta prestaciones. (2016)

5. AlMusbahi, Nahhas, AlMuhammadi, Anderkairi, Hemalatha: Survey on Green Computing: Vision and Challenges. Interenational Journal of Computer Applicatios. Vol 167, nro 10. (2017)
6. Montes de Oca E., De Giusti L., Chichizola F., De Giusti A., Naiouf M., "Utilización de Cluster de GPU en HPC. Un caso de estudio". IVX Workshop de Procesamiento Distribuido y Paralelo. CACIC2014. ISBN 978-987-3806-05-6. La Matanza, Buenos Aires, Argentina (2014)
7. Díaz J., Ambrosi V., Castro N., Candia D., Vega E., Rodriguez A.: Experiencia de la enseñanza de Green IT en la currícula de carreras de Informática de la UNLP. XI Congreso de Educación en Tecnología y Tecnología en Educación (2016)
8. Valdés Castro E.: Tecnologías de información que contribuyen con las prácticas de Green IT. Ingenium, pp. 11-26 (2014)
9. González C., Pérez R., Vásquez Stanescu C., Araujo G.: Eficiencia energíetica. Uso racional de la energía eléctrica en el sector administrativo. Ministerio del Poder Popular para la Energía Eléctrica. Municipio Libertador, Distrito Capital República Bolivariana de Venezuela (2014)
10. Talebi M., WayT.: Methods, Metrics and Motivation for a Green Computer Science Program. SIGCSE'09, Chattanooga, Tennessee, USA. (2009)
11. Silvestein M., Kim S., Huh S., Zhang X., Hu Y., Wated A.,Witchel E., "GPUnet: Networking Abstractions for GPU Programs", Transactions on Computer Systems, Vol. 34, No. 3, Article 9, Publication date: September 2016, ACM 0734-2071/2016/09, 2016
12. www.nvidia.com/object/what-is-gpu-computing.html fecha de acceso febrero 2018
13. www.nvidia.es/object/cuda-parallel-computing-es.html fecha de acceso: febrero 2018.
14. Nvidia Development: https://devtalk.nvidia.com/default/topic/808106/question-abount-cudasetdevice-and-multiple-host-threads (2018)
15. la.nvidia.com/object/tesla-feactures-la.hmtl fecha de acceso febrero de 2018
16. Bruzzone S.: LFN10, LFN10-OMP y el Método de Leapfrog en el Problema de los N Cuerpos. Instituto de Física, Departamento de Astronomia, Universidad de la República y Observatorio Astronómico los Molinos, Uruguy (2011)
17. https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html fecha de acceso: febrero 2018
18. Tsuyoshi H., Keigo N.: 190 TFlops Astrophysical N-body Simulation on cluster of GPUs. Universidad de Nagasaki. IEEE 978-1-4244-7558-2 (2010)
19. Montes de Oca E., De Giusti L., Chichizola F., De Giusti A., Naiouf M.: Análisis de uso de un algoritmo de balanceo de carga estático en un Cluster Multi-GPU Heterogéneo. XV Workshop de Procesamiento Distribuido y Paralelo. CACIC2016. San Luis, San Luis, Argentina (2016)
20. "RIGOL User's Guide MSO1000Z/DS1000Z Series Digital Oscilloscope": http://beyondmeasure.rigoltech.com/acton/attachment/1579/f-050a/1/-/-/-/-/MSO1000Z%26DS1000Z_UserGuide.pdf (2018)
21. Herramienta desarrollada por el Dr. Balladini J. de la Universidad Nacional del Comahue
22. Nvidia GPU Computing: www.nvidia.com/object/what-is-gpu-computing.html (2018)
23. Adrian Pousa, Victoria Sanz, Armando De Giusti. Análisis de rendimiento de un algoritmo de criptografía simétrica sobre GPU y Cluster de GPU. Instituto de Investigación en Informática LIDI, Fac. de Informática, UNLP. HPC La TAM 2013. (2013)

Modelado estadístico de potencia usando contadores de rendimiento sobre Raspberry Pi

Juan Manuel Paniego[1], Leandro Libutti[1], Martin Pi Puig[1], Franco Chichizola[1], Laura De Giusti[1,2], Marcelo Naiouf[1], Armando De Giusti[1,3]

[1]Instituto de Investigación en Informática III-LIDI, CEA-CIC, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina.
[2]Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC), La Plata, Argentina.
[3]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), CABA, Argentina.
{jmpaniego, llibutti, mpipuig, francoch, ldgiusti, mnaiouf, degiusti}@lidi.info.unlp.edu.ar

Resumen. Controlar la disipación de potencia y la temperatura es una gran preocupación en todos los sistemas informáticos modernos. No obstante, obtener información sobre el consumo de energía del procesador y del sistema puede no ser una tarea trivial. Afortunadamente, los procesadores de hoy en día cuentan con una gran cantidad de contadores de hardware para monitorear diferentes eventos en CPU y memoria. En este trabajo, se diseña un nuevo modelo estadístico de estimación de potencia destinado a la placa de desarrollo embebido Raspberry Pi 3. El modelo mapea valores de ciertos contadores de rendimiento a consumo de potencia del dispositivo a través de regresión lineal. Se analizan decenas de aplicaciones correspondientes a benchmarks clásicos, obteniendo un error promedio menor al 6.8% tanto para soluciones secuenciales como para algoritmos paralelos utilizando OpenMP.

Palabras claves: Consumo energético, Estimación de potencia, Raspberry Pi, Contadores de rendimiento, Modelo estadístico**.**

1. Introducción

En la actualidad, el consumo energético se ha convertido en una métrica de gran importancia que juega un papel fundamental a la hora de especificar el diseño de un sistema. Si bien hay muchas situaciones en las que es deseable ceder rendimiento en pos de un menor consumo energético, la performance sigue siendo el objetivo principal, por lo que la energía pasa a ser una preocupación de segundo nivel para los usuarios finales. Sin embargo, la constitución de un equilibrio entre el consumo de energía y los requisitos de rendimiento es de suma importancia para aprovechar de manera eficiente los recursos disponibles en muchos entornos, desde pequeños sistemas embebidos hasta grandes centros de cómputo de alto rendimiento (HPC).

Aunque los procesadores multi-core proporcionan una mejor relación rendimiento/consumo respecto a los single-core, la disipación de potencia continúa siendo un limitador de rendimiento clave para ambas arquitecturas. Recientemente, se han propuesto varias técnicas micro arquitectónicas para minimizar los problemas energéticos, como la adaptación dinámica de recursos del procesador [1][2][3] y el escalado dinámico de tensión y frecuencia (DVFS) [4][5].

La monitorización energética del procesador es importante para generar una correcta administración y, por consiguiente, reducir su consumo cumpliendo con los límites de disipación térmica de potencia del dispositivo. Luego, el control de energía requiere una supervisión en línea de la potencia consumida.

Si bien es posible realizar mediciones por hardware al sistema para implementar un control de potencia de grano fino, tal instrumentación es costosa y no comúnmente disponible. Generalmente, se basa en empotrar instrumentos de medición para monitorear cada nodo. Además, la latencia y los períodos de muestreo de las herramientas utilizadas pueden ser grandes en comparación con las escalas de tiempo en las que las cargas de trabajo pueden variar.

No obstante, los procesadores proporcionan un conjunto de contadores de hardware que se pueden utilizar para monitorear una amplia variedad de eventos relacionados con la performance del sistema con una alta precisión. Dado que cada evento está asociado con un determinado gasto de energía, cualquiera de ellos puede ser utilizado como parámetro en un modelo de rendimiento. Aunque el número de contadores suele ser limitado, los mismos permiten contabilizar una amplia diversidad de eventos. Por lo tanto, el problema de crear un modelo útil es encontrar un conjunto restringido de eventos que describa la mayor parte de la variación en potencia. Aunque algunos eventos representan actividades con poco impacto en la energía, otros exhiben una alta correlación.

Dadas las restricciones que presenta la instrumentación por hardware, construir un modelo estadístico es una alternativa viable para realizar la estimación del consumo energético del sistema en tiempo real. En este trabajo se entrena un modelo de potencia basado en regresión lineal y eventos de rendimiento para la placa de desarrollo Raspberry Pi 3. El modelo es validado a través de un conjunto de aplicaciones con diferentes cargas de trabajo. Esta placa presenta un set limitado de contadores, los cuales a su vez soportan una cantidad reducida de eventos. Desafortunadamente, esta placa no presenta contadores para monitorear el consumo energético del sistema.

Este artículo se organiza de la siguiente manera: en la Sección 2 se recopilan los trabajos relacionados en el ámbito de la predicción del consumo energético. Luego, en la Sección 3 se presenta la metodología para el diseño del modelo estadístico. Seguidamente, en la Sección 4 se detallan los procedimientos realizados para la construcción del modelo. Luego, en la Sección 5 se demuestra la validez de la estimación alcanzada. Finalmente, en la Sección 6 se comentan las conclusiones y las líneas de trabajo futuras.

2. Trabajos relacionados

En esta sección se presentan algunas de las investigaciones previas relacionadas con el presente trabajo. Isci et al. [5][6] construyó una metodología de modelado de potencia en tiempo real basada en el uso de contadores de rendimiento para identificar las diferentes fases de potencia dentro de un determinado código científico. Bellosa analizó el consumo energético total del sistema y lo comparó frente a la información obtenida por los contadores de rendimiento en [7]. Por otro lado, Joseph et al. [8] propuso el uso de los contadores de rendimiento como una base para la estimación de potencia del procesador y sus subcomponentes. Luego, Contreras et al. [9] utilizó contadores de rendimiento para estimar en tiempo real el consumo de potencia sobre un determinado procesador y su memoria. Además, propone que este esquema puede servir como base para la estimación en sistemas embebidos.

Por otra parte, Bircher et al. [10][13][19] exploró el uso de los contadores de rendimiento para predecir el consumo energético de diversos subsistemas como la CPU, memoria, chipset, E/S, disco y GPU. El mismo, fue desarrollado y validado sobre dos plataformas distintas. Asimismo, Rodrigues et al. [18] estudió el uso de los contadores de rendimiento aplicados en la estimación del consumo energético en tiempo real sobre dos arquitecturas diferentes: una orientada a alto rendimiento y la restante basada en bajo consumo.

Singh et al. [11] construyó una estimación del consumo de potencia en tiempo real haciendo uso de los contadores de rendimiento. Los mismos son recolectados a través de micro-benchmarks sobre códigos secuenciales y multihilo. Lee et al. [12] sugirió crear modelos de predicción de potencia y rendimiento usando diferentes parámetros de diseño de hardware y contadores de rendimiento. En el mismo, se realiza una combinación de modelos lineales y no lineales. Se utilizó, además, un framework de simulación para evaluar los errores de predicción.

Pusukuri et al. [14] desarrolló un modelo de potencia lineal basado en solo un par de predictores. Funciona para diversas frecuencias de CPU con una gran precisión. El modelo posee una alta portabilidad sobre arquitecturas x86 ya que los predictores utilizados se encuentran presentes en la mayoría de los nuevos procesadores. También, Jiménez et al. [15] caracterizó el comportamiento térmico y energético de un sistema basado en IBM POWER 6. De esta manera, diseñó el primer modelo analítico de consumo de potencia sobre esta arquitectura, el cual está basado en regresión lineal y contadores de rendimiento. En este sentido, Lim et al. [16] presentó y evaluó un enfoque general para construir modelos de estimación de potencia del sistema basados en contadores de rendimiento. A través de esta técnica, construyó un modelo estadístico de regresión lineal para el Intel Core I7 utilizando diferentes frecuencias, cantidad de núcleos y tamaño de problema. Aunque el artículo presenta resultados interesantes, los procesadores Intel ya disponen de un software de estimación de potencia propio conocido como Intel RAPL (Running Average Power Limit), previamente analizado en trabajos como [20][21][22].

Por otro lado, Lively et al. [17] desarrolló un conjunto de modelos híbridos de estimación de performance y consumo centrados en las aplicaciones. El mismo analiza

un conjunto de códigos científicos en su implementación MPI/OpenMP y genera un procedimiento adecuado para llevar a cabo el modelado y la validación.

Si bien los trabajos anteriores analizaron e implementaron diferentes técnicas de estimación energética, la arquitectura objetivo de los mismos son los procesadores CPU tradicionales. No obstante, este artículo se centra en realizar un estudio de similares características sobre procesadores de arquitectura ARM.

3. Metodología

Generalmente, el diseño de estos modelos estadísticos comprende dos aspectos básicos. En primer lugar, analizar una serie de aplicaciones que incorporen diferentes comportamientos computacionales, con el fin de extraer información importante sobre el rendimiento de una determinada arquitectura. Por otra parte, evaluar la performance del sistema (tiempo de ejecución) y/o el consumo de potencia del mismo.

Luego de la recolección de los datos, se selecciona el método estadístico a utilizar para la estimación objetivo. Los modelos existentes abarcan regresiones lineales, gaussianas, máquinas de soporte vectorial, redes neuronales, entre otros. Particularmente, en este trabajo se utiliza el método matemático de regresión lineal, el cual permite aproximar la relación de dependencia entre la potencia (variable dependiente) y la información que provee el hardware (variables independientes). Este modelo puede ser descrito de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

El conjunto de n observaciones se encuentra denotado por $\{Y_i, X_1, X_2, \ldots, X_n\}$ donde Y_i es la variable dependiente y X_i son las variables predictoras. Además, $\{\beta_0, \beta_1, \ldots, \beta_n\}$ son los coeficientes de las variables predictoras. Cabe destacar que cada uno de los predictores X_i se encuentra normalizo de acuerdo con la cantidad de ciclos ocurridos.

Para los datos de entrada del modelo se utilizan los contadores de rendimiento provistos por la arquitectura de hardware utilizada, los cuales son un conjunto de registros de propósito especial cuya función es almacenar en forma acumulativa ciertos eventos del sistema.

La selección de los contadores se realiza mediante un análisis de correlación individual contra la variable dependiente y, además, una correlación mutua entre cada par de contadores de manera tal de detectar información redundante entre ellos.

Generalmente, para implementar el modelo predictivo se emplean diferentes herramientas de desarrollo como Weka, Matlab, RapidMiner, R Studio, entre otras. En este caso, se utiliza la suite RapidMiner Studio 8.2, constituido como el producto mejor valorado frente a sus competidores por su facilidad de uso, calidad de soporte, rápida administración, entre otras, según la lista de los top 16 software para análisis estadístico y predictivo [28].

4. Desarrollo del modelo de estimación

En este trabajo, el modelo predictivo se construye con el objetivo de estimar el consumo de potencia de la placa de desarrollo Raspberry Pi 3. Es una placa embebida que presenta el microprocesador, la memoria, las comunicaciones inalámbricas y los puertos en un mismo circuito. El modelo utilizado dispone de un procesador Broadcom BCM2837B0, Cortex-A53 (ARMv8) con 4 cores de 64 bits a 1.4GHz. Posee una memoria de 1GB SDRAM y una GPU VideoCore IV a 400 MHz. Además, cuenta con conectividad inalámbrica y cableada, así como también, con un vasto número de pines de E/S de propósito general. La placa utiliza el sistema operativo Raspbian 8 Jessie (kernel 4.4.50), una distribución GNU/Linux basada en Debian.
El procesador ARM Cortex-A53 solo dispone de 6 contadores de rendimiento. Con ellos se puede contabilizar los distintos eventos disponibles en el chip, los cuales se encuentran listados en la Tabla I.

Tabla I. Eventos de rendimiento disponibles en Raspberry Pi 3 (ARM Cortex-A53).

Contador	Descripción
TOT_CYC	Totalidad de ciclos del procesador
TOT_INS	Totalidad de instrucciones completadas
LD_INS	Instrucciones de carga completadas
SR_INS	Instrucciones de almacenamiento completadas
BR_INS	Instrucciones de salto completadas
BR_MSP	Fallos en predicción de saltos
L1_ICM	Fallos en caché de instrucciones L1
L1_DCA	Cantidad de accesos a la caché de datos L1
L1_DCM	Fallos en caché de datos L1
L2_DCA	Cantidad de accesos a la caché de datos L2
L2_DCM	Fallos en caché de datos L2
TLB_IM	Fallos en instrucciones en TLB
TLB_DM	Fallos en datos en TLB
HW_INT	Cantidad de interrupciones de hardware

Para llevar a cabo la recolección de datos, se realizó el conexionado de los diferentes dispositivos involucrados, como se muestra en la Figura 1.

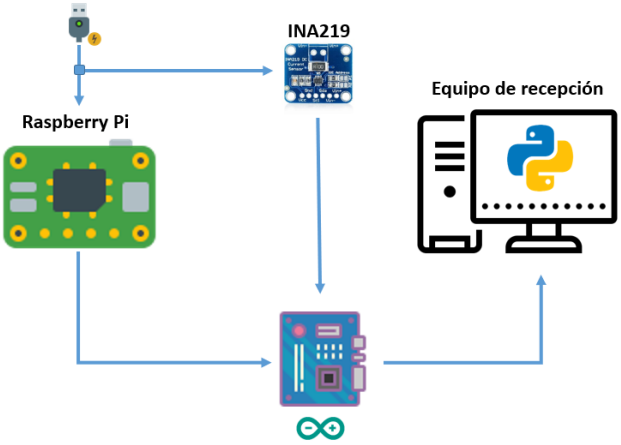


Figura 1. Esquema de conexión.

El sistema de medición está compuesto por una placa Arduino UNO conjunto al sensor INA219. En particular, Arduino UNO es una placa de desarrollo basada en el microcontrolador ATmega328 y muy utilizada para crear dispositivos digitales y objetos interactivos que permiten sensar y manipular diferentes variables del medio en el que se desarrollan. Por otra parte, el dispositivo INA219 permite analizar tensión y corriente continua a través del protocolo I^2C. De esta manera, combinando ambas variables se obtiene la potencia instantánea.

Luego, como se puede apreciar, la Raspberry Pi se alimenta por USB desde la línea eléctrica a través de un transformador convencional. Simultáneamente, el sensor mencionado mide la corriente y la tensión instantánea que consume la placa embebida. En este sentido, al momento de ejecutar cada uno de los benchmarks, la Raspberry Pi genera una señal digital por uno de sus pines de E/S para que la placa Arduino comience a almacenar la información proveniente del sensor INA219. Al mismo tiempo, el equipo de recepción ejecuta un script Python, el cual genera un archivo de muestras con los datos recibidos.

Se analizan las soluciones secuenciales y multihilo (OpenMP) para cada una de las aplicaciones de los benchmarks seleccionados. En primera instancia, se evaluó NAS [23], un conjunto de benchmarks originados por el grupo de investigación NASA Advanced Supercomputing. Además, se analizó RODINIA [24], una suite destinada a computación heterogénea que incluye diferentes aplicaciones según los patrones algorítmicos de Berkeley [25]. Por último, se ejecutaron varias de las aplicaciones que provee la suite LINPACK [26], las cuales permiten medir la potencia de cálculo de punto flotante en el sistema.

En este punto, se lleva a cabo una instrumentación sobre el núcleo de cada uno de los algoritmos, lo que permite supervisar el nivel de rendimiento de una aplicación. Este proceso se realiza a través de PAPI y genera un archivo con la información correspondiente a los contadores de rendimiento evaluados durante la ejecución del algoritmo.

Una vez analizados los diferentes benchmarks, se procede a la carga de los datos de potencia e información de los eventos de rendimiento en la suite RapidMiner. Luego,

se normalizan todos los contadores de acuerdo con el contador de ciclos. Posteriormente, se realiza la correlación de cada contador normalizado con la potencia medida, como se observa en la Figura 2. Además, se analiza la correlación mutua entre contadores. Ambos procesos permiten definir los contadores que dan lugar a la estimación de potencia del sistema. Generalmente, es deseable, que el modelo de estimación involucre el menor número de contadores, de manera de disminuir la cantidad de eventos monitorizados en simultáneo sobre el sistema. Particularmente, TOT_INS_NORM, BR_INS_NORM, SR_INS_NORM, L2_DCA_NORM, L2_DCM_NORM son los contadores seleccionados que, sumados a TOT_CYC (para implementar la normalización), representan las variables de entrada del modelo de regresión lineal.

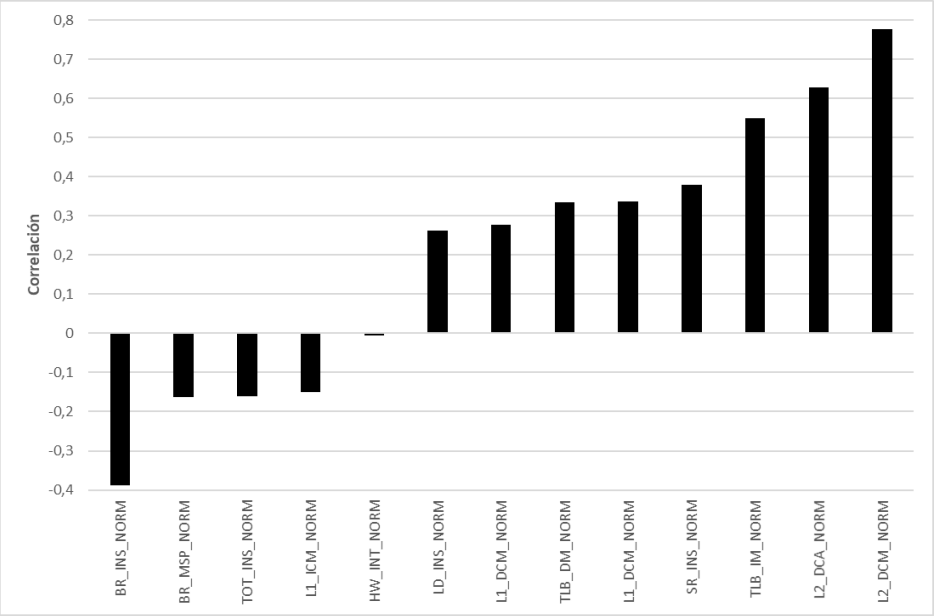


Figura 2. Correlación contador-potencia.

El proceso anterior genera un modelo de regresión lineal que utiliza los contadores seleccionados. Como se pretende modelar tanto soluciones secuenciales como multihilo, se entrenaron dos modelos estadísticos, uno para cada enfoque. Ambos utilizan el mismo set de contadores de rendimiento.

En la Tabla II se muestra el detalle de los pesos asociados a cada contador de rendimiento, según se utilice una solución secuencial u OpenMP.

Tabla II. Pesos obtenidos para cada predictor.

Predictor	Contador	Peso	Algoritmo secuencial	Algoritmo OpenMP
	Intercept	β_0	1.595	1.782
X_1	L2_DCM_NORM	β_1	14.696	79.001
X_2	L2_DCA_NORM	β_2	2.358	8.527
X_3	SR_INS_NORM	β_3	0.108	1.206
X_4	BR_INS_NORM	β_4	0.093	-1.626
X_5	TOT_INS_NORM	β_5	0.058	0.676

De esta manera, ambos modelos de predicción pueden ser descriptos de la siguiente manera:

$$Y_{SEC} = 1.595 + 14.696\,X_1 + 2.358\,X_2 + 0.108\,X_3 + 0.093\,X_4 + 0.058\,X_5$$
$$Y_{OMP} = 1.782 + 79.001\,X_1 + 8.527\,X_2 + 1.206\,X_3 - 1.626\,X_4 + 0.676\,X_5$$

La información reflejada anteriormente puede trasladarse fácilmente a una implementación por software o hardware para permitir que las aplicaciones en ejecución estimen el consumo energético del dispositivo en tiempo real.

5. Validación

Luego de realizado el proceso de entrenamiento para la obtención de los pesos, se contrastaron ambos modelos con los correspondientes valores observados. En este sentido, en la Figura 3 y 4 se detalla la potencia real, obtenida por el proceso de medición, y la potencia estimada, para un conjunto de las aplicaciones analizadas.

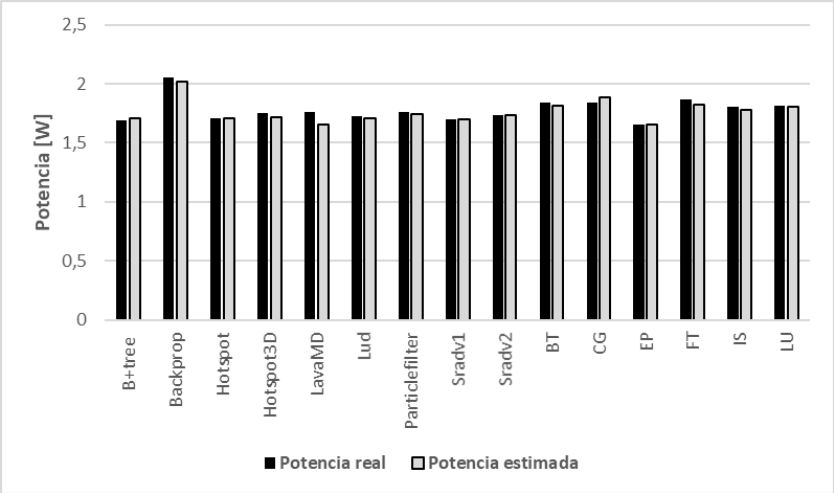


Figura 3. Potencia real vs. estimada para el modelo secuencial.

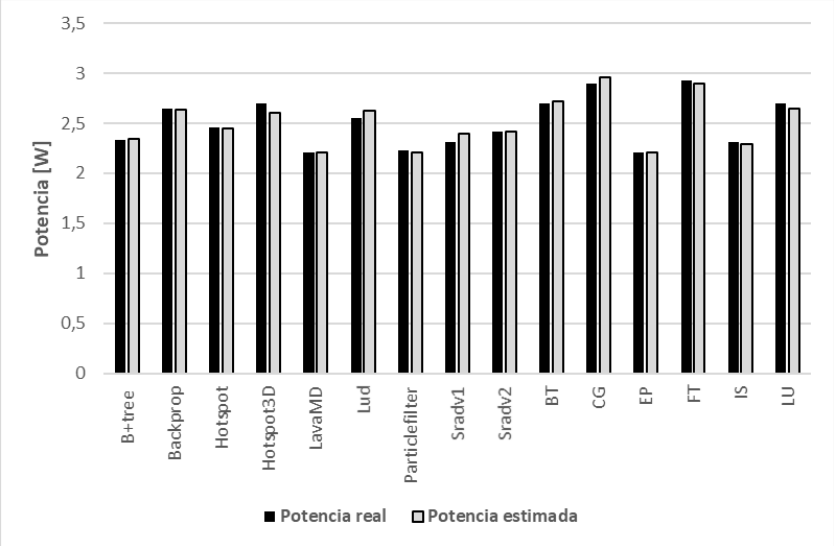


Figura 4. Potencia real vs. estimada para el modelo OpenMP.

Para estimar la precisión del modelo se implementó una validación leave-one-out cross-validation (LOOCV), técnica que permite separar la información de manera tal que, para cada iteración, se destina una única muestra para los datos de prueba mientras que el conjunto restante conforma los datos de entrenamiento del modelo actual.

Como resultado se obtuvo un error promedio de 6,3% para aplicaciones secuenciales y 6,8% para algoritmos OpenMP. Adicionalmente, en pos de clarificar la precisión y robustez de ambos modelos de estimación, se evaluó la raíz del error cuadrático medio (RMSE) [27]. En este caso, el modelo secuencial alcanza un RMSE de 0.033 mientras que la estimación paralela produce un error de 0.042.

Parte del error cometido puede deberse al hecho de que el modelo fue diseñado para funcionar online, razón por la cuál algunos eventos de rendimiento son marginados de la estimación. De esta manera, ciertos comportamientos no habituales de una determinada aplicación generan una desviación mayor entre la predicción y la potencia real.

6. Conclusiones y Líneas de Trabajo Futuro

Este artículo trata la utilización de contadores de rendimiento, específicamente sobre Raspberry Pi 3, con el fin de obtener una estimación confiable del consumo de potencia.

Se construyó un modelo estadístico para aplicaciones secuenciales y otro para aplicaciones paralelas utilizando OpenMP. Ambos basados en regresión lineal y validados frente a diferentes muestras obtenidas por experimentación. En este sentido, los datos generados para los modelos se adquieren a través de la instrumentación de

diferentes benchmarks. Además, la información correspondiente a la potencia real del dispositivo se extrajo a partir de un sistema físico de medición.

Como resultado del proceso de correlación, se seleccionaron los contadores de rendimiento que actúan como predictores en los modelos de regresión. Se restringió la cantidad de predictores seleccionables de acuerdo con la cantidad máxima de contadores de hardware que dispone el dispositivo en estudio. De esta manera, la estimación puede efectuarse en tiempo real.

La validación de ambos modelos estadísticos arrojó un error promedio menor a 6.8% tanto para algoritmos secuenciales como OpenMP.

Finalmente, se plantea como trabajo futuro la creación de un único modelo de predicción que se desligue de la cantidad de hilos empleada por cada aplicación. Además, evaluar la utilización de diferentes esquemas de regresión, los cuales permitirían eventualmente alcanzar errores menores.

Referencias

1. A. Buyuktosunoglu, S. Schuster, D. Brooks, P. Bose, P. Cook, D. Albonesi. An adaptive issue queue for reduced power at high performance. In: Workshop on Power-Aware Computers Systems, (2000).
2. D. Folegnani, A. Gonzalez. Energy-effective issue logic. In: 28th International Symposium on Computer Architecture, (2001).
3. F. D. Rossi, M. Storch, I. de Oliveira, C. A. De Rose. Modeling power consumption for DVFS policies. In: IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 1879-1882), (2015).
4. J. S. Lee, K. Skadron, S. W. Chung. Predictive temperature-aware DVFS. IEEE Transactions on Computers (pp. 127-133), (2010).
5. C. Isci, M. Martonosi. Identifying Program Power Phase Behavior Using Power Vectors. In: Proceedings of the 6th Annual IEEE International Workshop on Workload Characterization, (2003).
6. C. Isci, M. Martonosi. Runtime Power Monitoring in High-End Processors: Methodology and Empirical Data. In: Proceedings of the 36th International Symposium on Microarchitecture, (2003).
7. F. Bellosa. The benefits of event-driven energy accounting in power- sensitive systems. In: Proceedings of 9th ACM SICOPS European Workshop, (2000).
8. R. Joseph, M. Martonosi. Run-time Power Estimation in High Performance Microprocessors. Proceedings of the 2001 international symposium on Low power electronics and Design (ISLPED), (2001).
9. G. Contreras, M. Martonosi, Power prediction for intel XScale processors using performance monitoring unit events. In: Proceeding of ISLPED (pp. 221–226), (2005).
10. W. L. Bircher, L. K. John. Complete system power estimation using processor performance events. IEEE Trans. Comput., vol. 61, no. 4 (pp. 563–577), (2012).
11. K. Singh, M. Bhadauria, S.A. McKee, Real time power estimation and thread scheduling via performance counters. SIGARCH Comput. Archit. News, vol. 37, no. 2 (pp. 46–55), (2009).
12. B. Lee, D. Brooks. Accurate and efficient regression modeling for microarchitectural performance and power prediction. SIGOPS Oper. Syst. Rev. (pp. 185–194), (2006).
13. W. Lloyd Bircher, Lizy K. John. Complete System Power Estimation: A Trickle-Down Approach Based on Performance Events. In: Proc. of ISPASS (pp. 158-168), (2007).

14. K. Pusukuri K., D. Vengerov, A. Fedorova. A methodology for developing simple and robust power models using performance monitoring events. In: Proceedings of WIOSCA, (2009).
15. V. Jiménez, F. J. Cazorla, R. Gioiosa, M. Valero, C. Boneti, E. Kursun, C. Cher, C. Isci, A. Buyuktosunoglu, P. Bose. Power and thermal characterization of POWER6 system. In: Proceedings of the 19th international conference on Parallel architectures and compilation techniques (pp. 7-18), (2010).
16. M. Y. Lim, A. Porterfield, R. Fowler. SoftPower: fine-grain power estimations using performance counters. In: Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing (pp. 308-311), (2010).
17. C. Lively, X. Wu, V. Taylor, S. Moore, H. C Chang, C. Y. Su, K. Cameron. Power-aware predictive models of hybrid (MPI/OpenMP) scientific applications on multicore systems. Computer Science-Research and Development (pp. 245-253), (2012).
18. R. Rodrigues, A. Annamalai, I. Koren, S. Kundu. A study on the use of performance counters to estimate power in microprocessors. IEEE Transactions on Circuits and Systems II: Express Briefs (pp. 882-886), (2013).
19. W. Bircher, J. Law, M. Valluri, L. K. John. Effective use of performance monitoring counters for run-time prediction of power. University of Texas at Austin Technical Report TR-041104, (2004).
20. J. M. Paniego, S. Gallo, M. Pi Puig, F. Chichizola, L. De Giusti, J. Balladini. Analysis of RAPL Energy Prediction Accuracy in a Matrix Multiplication Application on Shared Memory. In: Argentine Congress of Computer Science (pp. 37-46). Springer, Cham, (2017).
21. S. Desrochers, C. Paradis, V. M. Weaver. A validation of DRAM RAPL Power Measurements. In: Second International Symposium on Memory Systems (pp. 445-470), (2016).
22. M. Hahnel, B. Dobel, M. Volp, H. Hartigl. Measuring Energy Consumption for Short Code Paths Using RAPL. Technische Universität Dresden. ACM SIGMETRICS Performance Evaluation Review Volume 40. (pp. 13-17), (2012).
23. D. H. Bailey, E. Barszcz, J. T. Barton, D. S. Browning, R. L. Carter, L. Dagum, R. A. Fatoohi, P. O. Frederickson, T. A. Lasinski, R. S. Schreiber, V. Venkatakrishan, S. K. Weeratunga, H. D. Simon. The NAS parallel benchmarks. The International Journal of Supercomputing Applications, (pp. 63-73), (1991).
24. S. Che, M. Boyer, J. Meng, D. Tarjan, J. W. Sheaffer, S. H. Lee, K. Skadron. Rodinia: A benchmark suite for heterogeneous computing. In: IEEE International Symposium on Workload Characterization (IISWC) (pp. 44-54), (2009).
25. K. Asanovic, R. Bodik, B. C. Catanzaro, J. J. Gebis, P. Husbands, K. Keutzer, D. A. Patterson, W. L. Plishker, J. Shalf, S. W. Williams, K. A. Yelick. The landscape of parallel computing research: A view from berkeley (Vol. 2). Technical Report UCB/EECS-2006-183, EECS Department, University of California, Berkeley, (2006).
26. J. J. Dongarra, P. Luszczek, A. Petitet. The LINPACK benchmark: past, present and future. Concurrency and Computation: practice and experience (pp. 803-820), (2003).
27. T. Chai, R. R. Draxler. Root mean square error (RMSE) or mean absolute error (MAE)?. Geoscientific Model Development Discussions, (pp. 1525-1534), (2014).
28. The Top 16 Predictive Analytics Software, https://www.g2crowd.com/categories/predictive-analytics?utm_content=buffer22c4a&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer#highest_rated.

Análisis de las topologías Edge y Fog para la continuidad del extremo al Cloud

Javier Sillero Ros, Nelson Rodríguez

Departamento de Informática, F. C. E. F. y N, Universidad Nacional de San Juan

Facultad de Ciencias Exactas, Físicas y Naturales, UNSJ, Ig. De la Roza y Meglioli, Rivadavia, San Juan, Argentina

javiersilleroros@gmail.com, nelson@iinfo.unsj.edu.ar

Abstract. Cloud Computing es un modelo que ha demostrado eficiencia y presenta muchas ventajas, por lo que ha sido adoptado por la industria. En los últimos años ha habido un fuerte impulso para mover todo al Cloud centralizado, habilitado por la virtualización y fomentado por la necesidad de reducir costos, acortar el tiempo de comercialización de nuevos servicios y aumentar la flexibilidad. Sin embargo, se perdió de vista la importancia de la ubicación de la funcionalidad para el rendimiento, el uso eficiente de los recursos de red y la experiencia del suscriptor. Con el surgimiento de IoT estas deficiencias se ponen de manifiesto y Cloud Computing presenta serias dificultades para adaptarse. Este trabajo presenta un análisis de la utilización de Fog y Edge Computing para mantener la continuidad del extremo hacia el Cloud, teniendo en cuenta los requisitos que deben cumplirse en redes IoT.

Keywords: Cloud Computing, Edge Computing, Fog Computing, IoT.

1. Introducción

Internet de las cosas (IoT) representa una de las tecnologías más disruptivas de los últimos años y es la base de nuevos desarrollos y tendencias como Smart City, Smart Home y Smart Factory. Se puede definir como una red dinámica de nodos físicos y virtuales, que permite una comunicación fluida entre objetos y la provisión de servicios contextuales [1]. Es necesario que la masiva cantidad de datos producidos por esta nueva tecnología sean procesados de alguna manera, por lo que son enviados inevitablemente al Cloud debido a la gran capacidad que ofrece el mismo.

La expansión del Cloud ha terminado de consolidar este modelo. A la Infraestructura, Plataforma y Software como servicio, se le agregaron: Redes, Almacenamiento, Contenedores, Funciones, Desktop, Base de datos, Seguridad, y Recuperación ante desastres, entre otras funcionalidades. En definitiva, lo que actualmente se conoce como XaaS o Everything as a Service [2].

Sin embargo, la gran cantidad y variedad de tipos de datos que se generan en el extremo y se envían al Cloud está saturando las redes y se presentan problemas como:

alta latencia, baja disponibilidad de ancho de banda, interrupciones momentáneas de Internet y diversos temas de seguridad. Es por ello que surgen los modelos Fog Computing y Edge Computing, con el objetivo de llevar parte del procesamiento, control y almacenamiento al borde de la red.

De esta forma, tanto el extremo como el Cloud deben trabajar en sintonía y se debe mantener la continuidad, pero, para lograr que el extremo y el Cloud se complementen correcta y eficientemente se requiere analizar una vasta cantidad de variables. Este estudio presenta un análisis acerca de las principales topologías de red utilizadas en IoT con el objetivo de determinar aquellas configuraciones de red que vuelven del modelo Fog Computing una solución aplicable a la realidad de modo tal que se propicie la continuidad IoT-Cloud.

El presente trabajo se estructura de la siguiente manera: en la Sección 2 se definen los conceptos de Edge y Fog Computing, en la Sección 3 tiene lugar la explicación de la situación actual y de la problemática planteada, la Sección 4 explicita la metodología de estudio empleada, mientras que la Sección 5 las topologías IoT consideradas para el estudio y la Sección 6 la herramienta de simulación usada junto con los casos de estudio planificados. La Sección 7 exhibe los resultados obtenidos para cada tipo de topología y la Sección 8 los comportamientos y resultados generales del análisis. Por último, en la Sección 10, se detallan las conclusiones.

2. Edge y Fog Computing

Llevar parte de las funciones de los centros de datos al extremo de la red ha sido puesto en práctica anteriormente, entre sus antecedentes se puede nombrar a las redes de distribución de contenido, cuyo objetivo es acelerar el rendimiento web. Las CDN son valiosas para la entrega de video por el ahorro de ancho de banda al almacenar en caché, pero el nuevo modelo Edge Computing generaliza y amplía el concepto de CDN al aprovechar la infraestructura de computación en la nube [3].

En 2009 se desarrolla el concepto de Cloudlet, que en la actualidad es mantenido por la Universidad de Carnegie Mellon y varios sponsors como Intel, Vodafone e IBM, entre otros. Un cloudlet se puede ver como un "*centro de datos en una caja*" cuyo objetivo es "*acercar el Cloud*" [4]. Dicha variante es bastante compleja en cuanto a su funcionamiento, por lo que no se utiliza frecuentemente.

En 2012, Flavio Bonomi y colaboradores presentaron el término Fog Computing para referirse a esta dispersa infraestructura de Cloud [5]. Sin embargo, en este caso, la motivación para la descentralización es la escalabilidad de la infraestructura IoT en lugar del rendimiento interactivo de las aplicaciones móviles.

Se han propuesto diferentes arquitecturas para realizar plataformas de computación de borde. El análisis de estas arquitecturas revela que el borde de la red no está claramente definido y los nodos que se espera que participen pueden variar. Además, la terminología utilizada para describir dicho borde difiere mucho, el mismo término se utiliza para definir arquitecturas y funcionalidades diferentes [6].

Dichas arquitecturas no son un simple traslado de la funcionalidad del Cloud a los nodos intermedios, dado que las aplicaciones IoT son diferentes de las basadas en Web. Ciertas aplicaciones IoT necesitan respuestas en tiempo real o con un límite de

tiempo conocido y además, muchas de estas presentan un gran número de dispositivos conectados que pueden variar, resultando un desafío importante mantener la escalabilidad.

Recientes iniciativas de la industria y la academia en forma conjunta, están generando contribuciones que auguran que estas arquitecturas tengan, en propuestas abiertas, soluciones interoperables y disponibles [7] [8].

Los términos Edge y Fog suelen llevar significados diferentes según el autor que los referencie, lo cual genera confusión. Asimismo, se suele definir Mist Computing como otro término dentro del modelo Edge. En particular, Fog y Mist Computing, usualmente se asocian a IoT en el extremo y ambos son incluidos como parte del Edge Computing. Se considera a Fog como aquella arquitectura que se ubica solamente en los dispositivos de red (extremo medio), mientras que a Mist como aquella que se refiere a los dispositivos finales (extremo final).

3. Situación Actual

En el extremo, los modelos que soporten IoT deben priorizar, entre otras cosas, cuatro cuestiones básicas para un buen funcionamiento: la minimización de la latencia, la conservación del ancho de banda, el abordaje de cuestiones de seguridad y el movimiento de la información a los lugares más indicados para su procesamiento [9]. Dichas cuestiones serían abordables por el modelo del Cloud, pero la dependencia de las redes instaladas no permite una completa satisfacción de dichos principios.

Fog Computing, por su parte, busca reducir el tráfico en la red, logrando una disminución en la latencia, y persigue un acercamiento del cliente con el servidor y una ampliación en el número de servidores, de modo que se aproveche mejor el ancho de banda disponible, se incremente la productividad de las aplicaciones y servidores y se mejore la experiencia de usuario. En este nuevo paradigma todo aquel dispositivo que tenga capacidad de procesamiento y se encuentre conectado a la red (siendo un dispositivo de red) es capaz de pre-procesar información y puede ser considerado un nodo Fog.

Por lo tanto, es necesario analizar las variantes de redes IoT: en qué casos son más convenientes, cuáles y cómo se adaptan mejor, siempre favoreciendo la continuidad extremo-Cloud.

A continuación, se presenta una evaluación de las topologías más usadas en IoT y mediante un análisis de estas, embebidas en entornos Fog Computing, se determina en qué condiciones estructurales las aplicaciones distribuidas en las mismas funcionarán de manera eficiente empleando un modelo Fog. Así es como se espera determinar la influencia del modelo Fog Computing principalmente en la continuidad IoT-Cloud.

4. Metodología de Estudio

En el presente trabajo se llevaron a cabo estudios sobre las topologías IoT más utilizadas, para lo cual se empleó una metodología cuantitativa en el marco de una investigación experimental, haciendo uso de una herramienta de simulación.

Dicha herramienta es específica para simular entornos Fog Computing y la misma entrega como resultado datos acerca de la Garantía de QoS de cada red simulada, utilización de recursos Fog Computing, costo de mantenimiento de cada topología y ubicación de aplicaciones predefinidas (carga en los diferentes nodos Fog). Luego de obtener los resultados de la simulación se evalúan los comportamientos observados bajo los distintos patrones de configuración ingresados.

De esta manera, mediante la simulación de entornos Fog se pretende evaluar el rendimiento de cada una de las configuraciones topológicas, haciendo variar la escala, el tipo de conectividad, el tipo y la cantidad de los dispositivos que componen las redes, y las aplicaciones a desplegar en los mismos. Todo ello es con el objeto de explorar aquellas variantes topológicas IoT que permiten una continuidad IoT-Cloud y determinar la/s más conveniente/s sin degradar la funcionalidad de ninguno de los extremos y contribuyendo a una distribución balanceada de, entre otras cosas, las tareas de cómputo y almacenamiento entre los dispositivos de red que conforman un entorno Fog y el Cloud.

5. Topologías IoT

Existen tres modelos topológicos básicos para IoT, los mismos son: el modelo Estrella, el modelo Peer-to-Peer y el modelo Mesh.

El modelo Estrella se basa en la centralización del cómputo, almacenamiento y control. En una red IoT de tipo Estrella, un dispositivo final está conectado únicamente a un dispositivo de red, el cual pasa a ser, en un entorno Fog, un nodo Fog. Los diferentes nodos Fog en redes Estrella se comunican mediante la conexión al Cloud o a nodos Fog de mayor jerarquía dentro de la topología.

Este tipo de estructura facilita la agregación y remoción de dispositivos finales y la detección de fallos en los mismos. Además, las situaciones complejas siempre son resueltas por el/los dispositivos/s central/es, conformando una red con una performance consistente, predecible, de baja latencia y suficiente ancho de banda. A pesar de todo ello, las redes Estrella presentan, debido a la centralización de las funcionalidades, una importante debilidad: la existencia de un único punto de fallo.

En cuanto a las redes de tipo Peer-to-Peer, su estructura consiste en que todos los nodos existentes en la red están conectados entre sí mediante un link. De esta manera, los nodos Fog de una topología Peer-to-Peer pasan a estar conectados punto a punto con los demás nodos Fog de la red (interconexión de dispositivos de red), mientras que los dispositivos finales se conectan al nodo de red más cercano. Así, este tipo de configuración es adaptativa a la situación en la cual se encuentra la red respecto del tráfico de información.

Si bien la latencia se ve disminuida debido a los tipos de enlaces establecidos, las redes de pares son más complejas de configurar producto de que es necesario establecer los links entre cada uno de los nodos de red agregados y los ya existentes y, luego, conectar los dispositivos finales con los mismos. Además, en términos de opinión de especialistas que trabajan en el mercado, las redes Peer-to-Peer no son muy útiles para ambientes IoT, debido a que no es habitual que dos dispositivos se comuniquen directamente entre ellos.

Por último, las redes Mesh (combinación de los otros dos tipos) consisten básicamente en tres clases de nodos: los gateways o Dispositivos de Funcionalidad Completa (FFD), los sensores/actuadores-routers o Dispositivos de Funcionalidad Reducida (RFD) [10], y los sensores/actuadores simples. Los Dispositivos de Funcionalidad Completa representan, en un ambiente Fog Computing, la salida al Cloud o nodos de mayor jerarquía (y por ende de mayor capacidad), mientras que los Dispositivos de Funcionalidad Reducida son los nodos de red que poseen capacidades para correr algún tipo de aplicación (nodo Fog). Los dispositivos finales son los sensores o actuadores simples, productores y receptores de la información enviada a la red.

Este tipo de topología combina en parte las ventajas de los modelos anteriores: son redes adaptativas (mayor tolerancia a fallos) y el comportamiento de los sectores estrella de la red introduce estabilidad en cuanto a rendimiento. Así como se concatenan las ventajas, las desventajas también se hacen presente producto de la herencia: la redundancia de los links de la configuración Peer-to-Peer continúa acomplejando la configuración y el mantenimiento y, como si fuera poco, continúa existiendo la posibilidad de la conformación de un cuello de botella propio de las redes Estrella.

6. Herramienta de Simulación

La herramienta de trabajo utilizada se denomina Fog Torch Pi, presentada como un prototipo en el proyecto de investigación: "*Through the Fog*", actualmente en ejecución en la Universidad de Pisa (Italia), y publicada junto al paper: "*How to best deploy your Fog applications, probably*" [11]. La misma es una aplicación desarrollada en Java [*https://github.com/di-unipi-socc/FogTorchPI*] y consta de 3 versiones hasta el momento.

Se utilizó su versión Cost Model, la cual "…*permite expresar las capacidades de procesamiento y los atributos promedio de la Calidad de Servicio (QoS) de la infraestructura Fog, junto con los requerimientos de procesamiento y QoS de la aplicación, y determinar los despliegues de la aplicación sobre la infraestructura Fog que satisfacen todos los requisitos preestablecidos*" [11].

El proceso de simulación consta de tres fases: la definición de la infraestructura, la definición de la/s aplicación/es y la simulación propiamente dicha. En la primera de las fases se determinan la topología a simular: se definen los nodos (servidores Clouds, nodos Fog, dispositivos finales) en cuanto a sus características de hardware y de software, y los links entre los mismos. La segunda fase consiste en crear la aplicación a desplegar en la infraestructura, para lo cual se definen los módulos de software y sus requerimientos, que deberán ser mapeados de la mejor manera sobre la infraestructura.

Por último, se lanza la simulación, la cual entrega todas las combinaciones de despliegue posibles de la aplicación ingresada sobre la topología señalada, devolviendo como resultado, para cada uno de los despliegues, el dispositivo donde se aloja cada módulo de software, el porcentaje en que satisface los requisitos de QoS (Garantía de Calidad de Servicio), el promedio de consumición de recursos de los

dispositivos Fog (considerando almacenamiento y memoria RAM) y el costo monetario mensual de montar una infraestructura de esas características.

Luego de diseñar y simular las situaciones necesarias, se procede a evaluar tanto los patrones ingresados como el producto obtenido y, a partir de ello, evaluar la performance de la red en general.

6.1. Casos de prueba

Para cada tipo de topología se definió un conjunto de prueba o archivos de entrada que representan una situación en particular, modificando en ocasiones una sola variable y en otras más de una.

En promedio se constituyeron 20 casos de prueba para cada configuración topológica, de manera que, en los diferentes escenarios, variaron tanto la escalabilidad de los dispositivos finales y los dispositivos de red (nodos Fog), como así también los tipos de enlaces: sobre todo los enlaces desde los nodos Fog hacia los servidores del Cloud, pero también los enlaces entre los dispositivos de red (en el caso de redes que permiten la interconexión de estos).

También alternaron las capacidades de software de las máquinas virtuales requeridas para desplegar las aplicaciones ingresadas. Estas últimas, se alojan en el Cloud y en los dispositivos de red, siempre que el hardware de los mismos tolere dicha carga.

En cuanto a las condiciones de hardware, las mismas no fueron modificadas con respecto al Cloud, sino respecto a los nodos de red, de manera que se observa el comportamiento en situaciones extremas de carga de trabajo.

7. Resultados

La figura presentada a continuación (Figura 1) expresa de manera resumida los resultados obtenidos de todo el proceso de simulación en su conjunto. La misma denota una situación en la que la cantidad de nodos de red, aplicaciones y dispositivos finales conectados a los nodos es igual para todas las topologías. A su vez, el tamaño de las máquinas virtuales y los requerimientos tanto de software como de hardware también son idénticos para todos los casos.

Los elementos variables son los links entre los nodos de red debido a la naturaleza de las diferentes topologías, pero el objetivo es representar un modelo comparativo general bajo el cual se pueden expresar los resultados de manera conjunta.

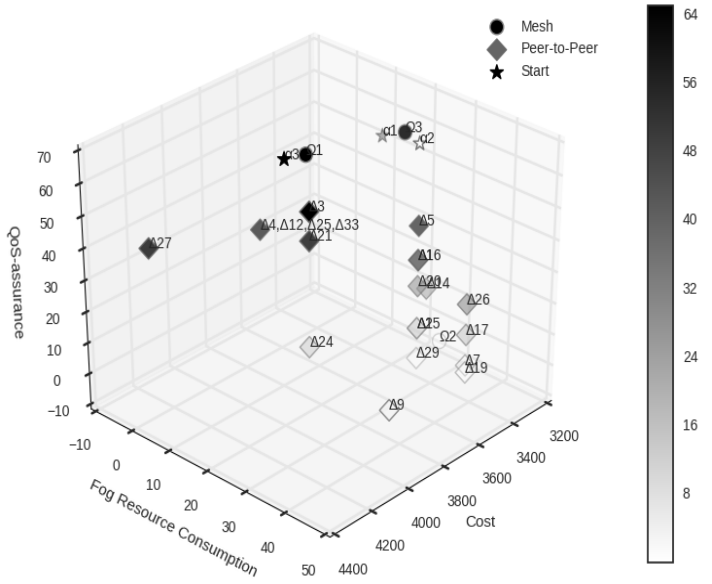


Fig. 1: Resultados comparativos de las 3 variables resultado entregadas por el simulador para las 3 topologías consideradas.

Si bien los números expresados en el gráfico pueden parecer muy diferentes a los a continuación expuestos para cada tipo de topología, se observa mediante este una tendencia a los comportamientos observados particularmente.

7.1. Redes Estrella

En las redes de tipo Estrella pura, en las que diferentes nodos Fog tienen comunicación directa con el Cloud, la QoS de la topología se ve decrementada si se aumenta la cantidad de nodos Fog. Sin embargo, al incrementar estos, se amplía momentáneamente la cantidad de despliegues, siempre dependiendo de: el tipo de link establecido entre los nodos de red y el servidor del Cloud, el número de aplicaciones desplegadas y los dispositivos a los cuales requieren acceso esas aplicaciones.

A su vez, si se incrementan los módulos de software capaces de ser alojados en dispositivos de red, aumentan los despliegues y la Garantía de QoS proporcionada por la red no se ve disminuida respecto de la utilización de únicamente el Cloud.

Los valores de Garantía de Calidad de Servicio arrojados para este tipo de red rondan entre el 50% y el 80%, entregando los mejores valores cuando se considera un solo nodo de red conectando ambos extremos.

En cuanto a la utilización de recursos Fog Computing en topologías con un solo nodo Fog, la misma ha llegado a ser del 84%, mientras que el promedio del resto de los despliegues -en los que el número de estos varía- se mantiene sobre el 20%. En aquellos despliegues donde se aloja una o más de una aplicación o módulo en un Fog, la consumición de recursos tiende a incrementar.

El planteo de topologías Estrella de manera jerárquica ha producido pocos despliegues con Garantías de Calidad de Servicio muy similares a las redes Estrella puras, pero se intensifica la utilización del Cloud.

7.2. Redes Peer-to-Peer

Los resultados de las simulaciones de las redes cuya configuración topológica es de tipo Peer-to-Peer arrojaron como conclusión que: contra más pares (nodos Fog) estén interconectados, mayor es la cantidad de despliegues posibles, y, si se incrementa a su vez la cantidad de módulos de software que requieren acceso a los dispositivos finales, mayor aún es el número de combinaciones. Sin embargo, la consumición de recursos Fog, independientemente de la cantidad de links que existan entre los nodos de red, nunca supera el 30%.

Si bien todo indica que la ubicación de los módulos de software más pesados se realice en los nodos Fog más potentes o capaces de este tipo de topología, la dispersión de estos en los diferentes nodos no contribuye de buena manera a la QoS de la red en general. Existen situaciones en las que, ante la distribución de las aplicaciones en los distintos dispositivos de red más potentes, la performance de la red se ve decrementada comparado con aquellas en donde se sobrecargan algunos nodos.

En cifras concretas, la Garantía de Calidad de Servicio de estas redes no supera el 80%, aunque se plantearon pruebas con variabilidad de links desde los nodos Fog hacia el Cloud, colocando enlaces de mayor latencia y menor ancho de banda, obteniendo resultados muy aislados con QoS superior al 85%, con consumiciones de recursos Fog del 25% en promedio y, claramente, una tendencia a la utilización mayoritaria del Cloud, cargando sobre los nodos de la red intermedios aquellas aplicaciones más livianas.

7.3. Redes Mesh

En general, para este tipo de topología, la Garantía de Calidad de Servicio es menor que las redes Estrella, a excepción de cuando se centraliza la comunicación con el Cloud. Normalmente, cuando toda la información de la red es transferida al Cloud a través de un único punto, la QoS es superior al 60%.

La predominancia de secciones en formato Estrella dentro de una red Mesh asegura, como anteriormente se menciona, la estabilidad en la performance de la misma, lo que se ve plasmado como el incremento en la Garantía de QoS en redes Mesh predominantemente de tipo Estrella.

Este tipo de topología expresa desde sus resultados la tendencia a una mayor cantidad de despliegues cuando la red presenta más secciones de tipo Peer-to-Peer, entregando datos de performance inferiores: nuevamente se introduce el comportamiento típico de las redes de pares (más despliegues posibles, menor QoS).

En la mayoría de los casos, el despliegue de las aplicaciones se da en los nodos raíz de las secciones Estrella y en los nodos más fuertes de las secciones Peer-to-Peer, arrojando como consumición de recursos un promedio no superior al 25%.

8. Resultados Generales

Además de los resultados expresados particularmente para cada tipo de topología, se observaron ciertas características respecto al comportamiento de las diferentes topologías luego de realizar el proceso de simulación de estas.

Respecto a la Garantía de Calidad de Servicio de los despliegues entregados, la misma se ve alterada cuando el tipo de enlace varía su latencia y ancho de banda, de modo que los enlaces de menor ancho de banda y mayor latencia siempre entregan una mejor performance de la red, pero no proporcionan una variedad nutrida de despliegues posibles.

El mencionado porcentaje de Garantía de QoS arrojado por el simulador estima cuán bien un despliegue de las aplicaciones ingresadas sobre la topología diseñada satisface la Calidad de Servicio requerida en su conjunto, por lo tanto, ello indica de qué manera se adapta la configuración topológica dentro de un ambiente Fog a las necesidades del IoT y los usuarios en general.

Respecto al porcentaje de consumición de recursos hardware en los nodos de red producido por la carga de aplicaciones, se presenta una clara influencia sobre estos respecto del tamaño de las máquinas virtuales o recursos hardware requeridos por las aplicaciones a desplegar sobre los mismos.

Por último, en cuanto a la variable de costos, en aquellas topologías donde el uso de recursos Fog es mayor, el costo de utilización de estas es inferior comparado a aquellas en las que se utilizan pocos recursos hardware de los nodos de red, requiriendo más del procesamiento y almacenamiento del Cloud. Este comportamiento se debe a que el empleo de recursos de fuentes mucho más grandes (inicialmente no diseñadas para el procesamiento continuo de funciones relativamente sencillas) es naturalmente más costoso, mientras que la utilización de nodos de red, que de hecho ya existen y están disponibles, resultan ser más accesibles y, en ocasiones, lo suficientemente convenientes en cuanto a sus capacidades para justificar su aprovechamiento.

9. Conclusiones

Luego de introducir las diferentes topologías IoT en entornos Fog Computing, realizar la simulación de estas sobre el prototipo Fog Torch Pi y analizar los resultados producidos, se ha arribado a la conclusión de que las configuraciones topológicas de tipo Estrella son lo más convenientes para promover la continuidad IoT-Cloud.

El descenso de las funcionalidades del Cloud hacia el borde de la red no puede realizarse de manera totalmente federal. Las redes federales, con cantidades de enlaces y nodos superiores entregan diferentes y mayores posibilidades de despliegue de las aplicaciones Fog sobre los distintos nodos de red capaces de soportar las mismas, pero ese tipo de redes añaden diversos problemas de interconectividad, por lo que dichas posibilidades no contribuyen a una red estable y de un rendimiento predecible y constante.

Durante el proceso de este trabajo se ha comprendido que la descentralización de las funcionalidades del Cloud no debe ser sobrecargada solo sobre los dispositivos

finales ni distribuida por la red con criterios de igualdad. La continuidad IoT-Cloud es el balance de cargas entre los dispositivos existentes, es la distribución de dichas tareas a lo largo del camino del extremo al Cloud, pero de manera semi-centralizada y criteriosa, agrupando sobre determinados puntos aquellas capacidades de orquestación y organización que favorezcan dicha prolongación extremo-Cloud.

La escalabilidad en las redes ha demostrado en este estudio que, a mayor cantidad de dispositivos a controlar, mayor es la cantidad necesaria de nodos Fog a implantar en la red, exigiendo una capacidad suficientemente alta de los mismos para poder cargar las aplicaciones necesarias. Esto a su vez significa una mayor utilización de recursos Fog, pero sin garantizar demandas de QoS en un 100%, lo que introduce una nueva discusión acerca de los servicios ofrecidos por el modelo Fog Computing en general.

Si bien las topologías Estrella presentan una de las soluciones más convenientes en términos de rendimiento a la continuidad IoT-Cloud, este trabajo deja en claro que no existe aún un modelo estático excepcional para ser utilizado en todos los casos y a gran escala. La integración extremo-Cloud requiere de un arduo y minucioso estudio, que contemple muchas más variables que afectan o podrían afectar al funcionamiento en armonía de los extremos opuestos.

Referencias

1. Lee G. M. et al: The internet of things: concept and problem statement :01. Dept. Reseaux et Service Multimedia Mobiles; Electronics and Telecommunications Research Institute; China Internet Network Information Center; Services repartis, Architectures, Modelisation, Validation, Administration des Reseaux, Research Report, Mar. 2011, https://hal.archives-ouvertes.fr/hal-00634616 (2011).
2. CBS Interactive Inc: Special Report: The future of Everything as a Service. (2017).
3. Dilley J. et al.: Globally Distributed Content Delivery. IEEE Internet Computing, vol. 6, No.5, 2002, pp.50–58, (2002).
4. Satyanarayanan M., Bahl P., Cáceres R., Davies N.: The Case for VM-Based Cloudlets in Mobile Computing. Pervasive Computing. IEEE (2009).
5. Bonomi F.et al.: Fog Computing and Its Role in the Internet of Things. Proc. 1st Edition MCC Workshop Mobile Cloud Computing (MCC12), pp.13–15 (2012).
6. Premsankar G., Di Francesco M., Taleb T.: Edge Computing for the Internet of Things: A Case Study. IEEE Internet of Things Journal, Vol. 5, No. 2, April (2018).
7. Open Fog Consortium, http://www.openfogconsortium.org
8. Open Edge Computing, http://www.openedgecomputing.org
9. CISCO: Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are". (2015). White Paper: http://www.cisco.com/go/iot.
10. Jekishan K. Parmar, Ankit Desai. "IoT: Networking Technologies and Research Challenges". International Journal of Computer Applications. Vol. 154. November (2016).
11. Brogi A., Forti S., Ibrahim A.: How to best deploy your Fog applications, probably. 1st IEEE International Conference on Fog and Edge Computing (ICFEC 2017), May 14th, 2017, Madrid, Spain, http://pages.di.unipi.it/forti/pdf/icfec17.pdf.

Análisis del Uso de un Cluster de Raspberry Pi para Cómputo de Alto Rendimiento

Sebastián Rodríguez Eguren, Franco Chichizola, and Enzo Rucci

III-LIDI, CEA-CIC, Facultad de Informática, Universidad Nacional de La Plata
La Plata (1900), Buenos Aires, Argentina
Email: {seguren,francoch,erucci}@lidi.info.unlp.edu.ar

Resumen En la actualidad, la mejora de la eficiencia energética se ha vuelto uno de los principales desafíos para el cómputo de alto rendimiento (HPC). La Raspberry Pi (RPi) es una *Single Board Computer* de bajo costo, la cual cuenta con múltiples núcleos, capacidad para conectarse en red y soporte a sistemas operativos y herramientas de programación paralela tradicionales. Aunque sus núcleos no son potentes, el bajo consumo de potencia las vuelven una opción atractiva desde el punto de vista energético. En este trabajo se analiza la viabilidad de usar un cluster de placas RPi para HPC. Mediante la ejecución de diferentes aplicaciones paralelas con alta demanda computacional, se identificaron fortalezas y debilidades del cluster RPi. Adicionalmente, se comparó su rendimiento y eficiencia energética con el de un cluster x86.

Keywords: Raspberry, RPi, x86, HPC, Eficiencia energética

1. Introducción

En la actualidad, los grandes sistemas HPC son dominados por procesadores de propósito general con dos conjuntos de instrucciones (x86 y Power) de tres empresas vendedoras (Intel, AMD e IBM) [5]. Estos procesadores han sido diseñados especialmente para ofrecer un alto rendimiento por núcleo, pero con costo económico (de cientos a miles de dólares) y demanda de potencia acordes (algunos puede requerir más de 150W). En la búsqueda de construir sistemas más grandes que nos permitan llegar a la escala de los Exaflops, resulta fundamental explorar otras estrategias que ofrezcan tasas de rendimiento costo-eficientes y que sean capaces de reducir el consumo energético de estos sistemas [24].

Durante décadas, el desarrollo de plataformas HPC estuvo focalizado casi únicamente en mejorar el rendimiento de las mismas. Esto provocó un crecimiento exponencial en los requerimientos de potencia de estos sistemas (no sólo para alimentarlos sino también para refrigerarlos), lo que a su vez repercutió en el costo económico de los mismos. Es por eso que, hoy en día, la reducción del consumo energético se ha vuelto uno de los principales desafíos para la comunidad HPC.

Entre las diferentes alternativas actualmente exploradas, se encuentra el uso de procesadores de bajo consumo como los ARM. Esta clase de procesadores ha sido especialmente diseñada para el mercado de dispositivos móviles y los sistemas embebidos. Recientemente, han entrado al mercado de PCs y servidores y se espera que en el futuro incrementen significativamente su poder de cómputo manteniendo el bajo consumo de potencia [18]. Para la comunidad HPC, este escenario presenta una oportunidad de emplearlos para construir sistemas de alto rendimiento. En los últimos años, el uso de procesadores ARM se ha vuelto popular entre desarrolladores y entusiastas de la programación debido al surgimiento de *Single Board Computer*'s (SBC) como la Beagleboard [4], la Raspberry Pi [7] y la Pandaboard [3]. Esto ha llevado a la formación de varias comunidades de desarrollo que invierten tiempo y esfuerzo en portar una amplia gama de sistemas operativos, aplicaciones, librerías y herramientas a los sistemas ARM más populares [9].

La RPi es una SBC de bajo costo desarrollada en el Reino Unido por la Fundación Raspberry Pi, con el objetivo de estimular la enseñanza de las ciencias de la computación en las escuelas. Desde su surgimiento, la RPi ha sido adoptada masivamente para una amplia gama de aplicaciones debido a su alto cociente rendimiento/precio, su tamaño reducido (similar al de una tarjeta de crédito) y su versatilidad [11]. En particular, desde el punto de vista del procesamiento paralelo, las RPi tienen múltiples núcleos, pueden conectarse en red y dan soporte a sistemas operativos y herramientas de programación tradicionales de HPC. Aunque sus núcleos no son potentes, el bajo consumo de potencia de las RPi (˜3.5W) las vuelven una opción atractiva desde el punto de vista energético. Es por eso que resulta interesante analizar la viabilidad del uso de un cluster de RPi para HPC.

El resto del documento se organiza de la siguiente forma: la Sección 2 describe las características del cluster de RPi utilizado en este estudio mientras que la Sección 3 detalla las aplicaciones benchmark seleccionadas. La Sección 4 analiza los resultados obtenidos, seguido de la Sección 5 que estudio los trabajos relacionados. Finalmente, la Sección 6 presenta las conclusiones y las posibles líneas de trabajo futuro.

2. Cluster de RPi

Existe en el mercado una amplia gama de variantes de las placas RPi. La familia original tiene 3 versiones. La primera RPi (versión 1) tuvo varios modelos: A, B y B+, siendo B el más utilizado. La versión 2 sólo tuvo un modelo (B). Por su parte, la versión 3 también tuvo durante un largo tiempo un único modelo (B), siendo presentado recientemente su sucesor (B+) en marzo de 2018.

Para esta investigación se ha desplegado un cluster de 20 placas RPi 3 modelo B, las cuales se encuentran interconectadas por un switch 3COM 3C16475BS y alimentadas por una fuente estándar para PC de 250W. Como software de base, las placas cuentan con el sistema operativo Raspbian Stretch Lite. En la Figura 1 se presenta una RPi 3 B mientras que en la Figura 2 se muestra una foto del cluster utilizado en este trabajo.

La RPi 3 está conformada por un *System-on-Chip* (SoC) especialmente desarrollado para este modelo (Broadcom BCM2837), el cual incluye 4 núcleos de procesamiento ARM Cortex-A53 de alto rendimiento que funcionan a 1.2GHz con 32Kb de cache de nivel 1 y 512Kb de nivel 2. Además, cuenta con un procesador gráfico VideoCore IV que se encuentra conectado a un módulo de memoria LPDDR2 de 1 GB en la parte posterior de la placa.

Este modelo comparte el mismo chip SMSC LAN9514 que su predecesor, la RPi 2, agregando conectividad Ethernet 10/100 y cuatro canales USB a la placa. El chip SMSC se conecta al SoC a través de un único canal USB, que funciona como un adaptador USB-a-Ethernet y un concentrador USB. También se puede encontrar el chip Broadcom BCM43438 que proporciona conectividad LAN inalámbrica de 2.4GHz 802.11n y Bluetooth 4.1 de bajo consumo.

3. Aplicaciones Benchmark

Para poder realizar el análisis de rendimiento y eficiencia energética, se seleccionaron 3 aplicaciones que se pueden considerar clásicas y representativas de aplicaciones científicas, que poseen alta demanda computacional pero con diferentes características entre sí. Las implementaciones fueron desarrolladas en lenguaje C con la librería MPI. A continuación, se describen las características distintivas de cada una de las aplicaciones junto a las técnicas de programación y optimización realizadas en cada una de las aplicaciones paralelas.

Multiplicación de Matrices (MM). La implementación paralela sigue el modelo *master-worker*, donde uno de los procesos es responsable de la distribución del trabajo y la recuperación de los resultados parciales. Para computar $C = A \times B$, el *master* envía a cada proceso (inclusive a sí mismo) un bloque de filas de la matriz A y la matriz B en forma completa. Luego, cada proceso será responsable de calcular un bloque de filas de C computando las celdas de a bloques cuadrados para maximizar la localidad de datos. Por último, resulta importante mencionar que los procesos emplean operaciones de comunicación colectiva para el intercambio de datos (`MPI_Scatter` para distribuir A, `MPI_Bcast` para enviar B y `MPI_Gather` para recuperar los resultados parciales de C).

Sintéticamente, MM es una aplicación que demanda gran cantidad de cómputo regular y que requiere poca cantidad de comunicaciones al inicio y al final de cada tarea, pero de gran tamaño.

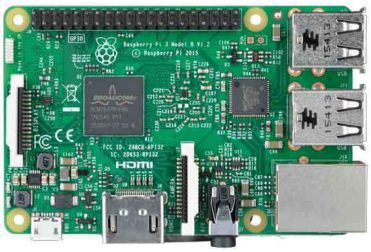


Figura 1: RPi 3 B

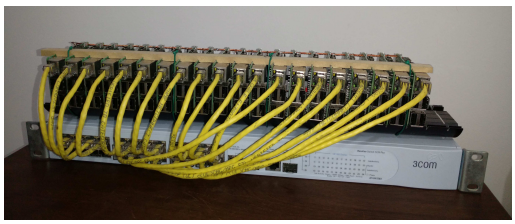


Figura 2: Cluster de RPi 3 B

Método de Jacobi. Jacobi es un método iterativo para resolver sistemas de ecuaciones lineales. Durante un número determinado de pasos, se actualizan dos matrices (*M1* y *M2*) en forma alternada. Esto significa que en el paso *i*, *M1* se actualiza leyendo los valores de *M2* mientras que, en el paso *i+1*, *M2* se actualiza a partir de los valores de *M1*. El nuevo valor de una celda está definido por el promedio de cuatro vecinos (el de arriba, el de abajo, el de la izquierda y el de la derecha).

La implementación paralela sigue el modelo *Single Program Multiple Data* (SPMD). Cada proceso maneja un bloque de filas de cada matriz y necesita de espacio adicional para mantener las filas de los límites de sus vecinos. En cada paso, cada proceso actualiza el bloque de filas correspondiente y, a continuación, intercambia con sus vecinos las filas de los límites de sus bloques.

En resumen, Jacobi es una aplicación regular con alto grado de paralelismo y de dependencia a nivel de datos entre las distintas tareas y etapas de ejecución. Esta gran cantidad de sincronización entre las tareas genera una considerable cantidad de comunicaciones de moderado tamaño a lo largo de toda la ejecución.

N-Reinas. Este problema se basa en determinar la cantidad de formas diferentes de ubicar N reinas en un tablero de N×N sin que se amenacen (no pueden convivir dos reinas en una misma fila, columna o diagonal). En la implementación paralela se generan todos los posibles tableros iniciales (tableros con las 4 primeras reinas ubicadas) y se los distribuye bajo demanda entre los procesos utilizando el modelo *master-worker*. El proceso *master* se encarga únicamente de distribuir estos tableros y recibir los resultados. Cuando un *worker* recibe un tablero inicial para procesar, calcula todos las formas válidas de completar las N reinas en ese tablero (el trabajo para realizar ese cálculo varía considerablemente dependiendo del tablero inicial) y le responde al *master*.

En síntesis, N-Reinas es una aplicación con alto grado de paralelismo, pero donde las tareas son irregulares y de gran tamaño, lo que obliga a realizar la distribución dinámica de las mismas entre los procesos. Esto genera una gran cantidad de comunicaciones no estructuradas de tamaño pequeño.

4. Resultados Experimentales

4.1. Diseño Experimental

Para realizar la experimentación se utilizó el cluster de RPi descrito en la Sección 2. Además, también se usó un cluster de computadoras x86, donde cada nodo cuenta con un procesador Intel Core i5-4460 3.20Ghz, 8 GB RAM y conexión de red Gigabit Ethernet. Las pruebas fueron realizadas en tres etapas:

- El objetivo de la primera etapa fue determinar los tiempos de ejecución de los algoritmos secuenciales, variando el tamaño de problema. En el caso particular de MM, además se probaron diferentes tamaños de bloque a fin de obtener el óptimo, el cual se utilizó posteriormente para correr las pruebas paralelas.

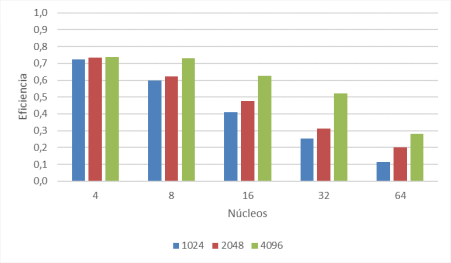


Figura 3: Valores de Eficiencia obtenidos para MM en el cluster RPi

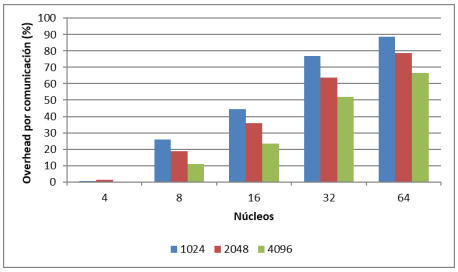


Figura 4: Valores de Overhead obtenidos para MM en el cluster RPi

- La segunda etapa de pruebas se basó en determinar los tiempos de ejecución de los algoritmos paralelos en el cluster de RPi.
- En la tercera etapa se ejecutaron los algoritmos paralelos en el cluster de computadoras x86.

Cada prueba fue repetida 10 veces y los valores reportados corresponden a la media de los datos obtenidos.

4.2. Resultados en el cluster RPi

Para evaluar el comportamiento de las aplicaciones benchmark al escalar el problema (N) y/o la arquitectura (P) se analiza la Eficiencia. En la Figura 3 se grafican los valores de Eficiencia obtenidos para MM. Se puede apreciar como la Eficiencia se incrementa a medida que crece el tamaño del problema para un determinado número de núcleos y se degrada al incrementar el numero de núcleos para un tamaño de problema dado. Este último decremento se debe principalmente al overhead que generan las comunicaciones, ya que a medida que aumenta P, se incrementan las interacciones entre procesos.

Resulta interesante analizar la incidencia de las comunicaciones en el rendimiento final considerando que la interacción entre procesos limita la Eficiencia alcanzada en las pruebas realizadas. En la Figura 4 se ilustran los porcentajes de tiempo de ejecución dedicados a comunicación para cada conjunto de pruebas realizadas de la aplicación MM. Se puede notar que, al aumentar P y mantener fijo N, el porcentaje de comunicación se incrementa. A la inversa, el porcentaje de comunicación disminuye cuando N aumenta y P se mantiene. También se observan porcentajes elevados de overhead que obviamente penalizan el rendimiento. Esto se debe a la velocidad de la interfaz de red de la RPi (100Mbit), lo cual impone una fuerte limitación al rendimiento final de la aplicación considerando el volumen de las comunicaciones. La excepción es cuando $P = 4$, ya que no se genera comunicación *real* por la red (se utiliza una única RPi).

Resulta importante mencionar que, aunque se probó con tamaños de matrices más grandes, no fue posible reportar sus resultados debido a que el requerimiento de memoria supera ampliamente el espacio disponible en cada nodo del cluster, lo que produce fallas en la ejecución.

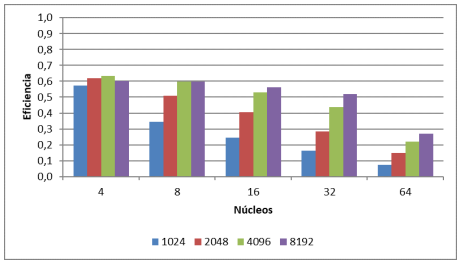


Figura 5: Valores de Eficiencia obtenidos para Jacobi en el cluster RPi

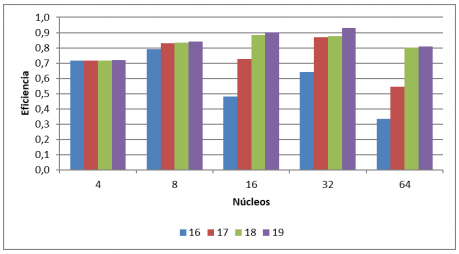


Figura 6: Valores de Eficiencia obtenidos para N-Reinas en el cluster RPi

En la Figura 5 grafican los valores de Eficiencia obtenidos para Jacobi. De los resultados se aprecia un comportamiento similar al caso MM. Al mantener fijo el tamaño de la matriz y aumentar la cantidad de procesadores, la Eficiencia disminuye debido al overhead de comunicación que se al comunicar los bordes de la matriz a cada proceso y, en menor medida, a la distribucin inicial de datos. Al escalar N manteniendo fijo P, la Eficiencia mejora en general, dado que el overhead mencionado anteriormente tiene menor peso en el tiempo total de ejecución.

Se debe notar que los valores de Eficiencia obtenidos con Jacobi son inferiores a los correspondiente a MM. Esto es debido a que, aunque manejen estructuras de datos más pequeñas, deben realizarse más comunicaciones (de pocos datos) y la relación cómputo-comunicación es mucho mas baja. Además, también es importante mencionar que existen muchos puntos de sincronización que producen overhead (espera ociosa).

En la Figura 6 se presentan los valores de Eficiencia obtenidos para N-Reinas. Es posible notar un incremento de la Eficiencia a medida que aumenta N, lo cual se debe a que el componente paralelo del algoritmo crece a un ritmo mayor que el componente serie al aumentar el tamaño del tablero. También se logra un incremento en la Eficiencia -más suave- al incrementar el número de núcleos hasta 32, que luego se reduce (P= 64). Este comportamiento no se da en todos los casos debido a la irregularidad propia de la aplicación.

Cabe recalcar que de las pruebas realizadas, N-Reinas fue la que mejores resultados arrojó (valores de Eficiencia cercanos a 1) y sus causas son múltiples. En primer lugar, la aplicación trabaja con un conjunto reducido de datos, de tamaño mucho menor al del resto de los benchmarks empleados. En segundo lugar, si bien requiere una gran cantidad de comunicaciones, las mismas son de pocos datos y no representan un porcentaje de tiempo significativo respecto al tiempo total de la aplicación. Por último, el algoritmo trabaja con números enteros a diferencia de los anteriores que computan en punto flotante.

4.3. Comparación con cluster x86

Para poder comparar el rendimiento entre los diferentes clusters se requiere una métrica común. Es por eso que se seleccionó la métrica GFLOPS para MM y

P	N		
	1024	2048	4096
4	0,43	0,51	0,51
8	0,72	0,87	1,01
16	0,98	1,32	1,74
32	1,21	1,73	2,88
64	1,09	2,24	3,12
GFLOPS			

(a) MM

P	N			
	1024	2048	4096	8192
4	0,74	0,68	0,67	0,64
8	0,89	1,12	1,27	1,27
16	1,27	1,79	2,23	2,39
32	1,70	2,51	3,71	4,41
64	1,53	2,64	3,71	4,61
GFLOPS				

(b) Jacobi

P	N			
	16	17	18	19
4	69,95	70,06	70,22	70,66
8	154,63	162,51	162,88	164,92
16	188,12	284,56	345,50	353,21
32	501,28	680,00	687,74	728,97
64	522,59	852,10	1257,04	1266,97
MQUEENS				

(c) N-Reinas

P	N		
	1024	2048	4096
4	7,85	7,99	7,99
8	7,19	12,09	13,86
16	7,17	13,73	17,43
32	7,78	11,32	23,58
64	5,02	12,95	32,62
GFLOPS			

(d) MM

P	N			
	1024	2048	4096	8192
4	6,88	3,76	3,84	3,84
8	6,88	7,48	7,61	7,67
16	6,61	20,01	23,79	23,83
32	5,36	27,13	45,18	47,96
64	5,73	27,92	75,07	86,71
GFLOPS				

(e) Jacobi

P	N			
	16	17	18	19
4	327,88	327,27	329,12	331,14
8	679,39	750,89	766,08	770,29
16	888,43	1284,37	1347,85	1358,97
32	1035,69	2259,81	2697,68	2782,20
64	554,17	2753,42	5014,06	5623,87
MQUEENS				

(f) N-Reinas

Tabla 1: Rendimiento de las aplicaciones benchmark en el cluster RPi (a-c) y en el x86 (d-f).

Jacobi. Sin embargo, en el caso de N-Reinas, al ser una aplicación que computa con números enteros, se empleó una métrica *ad-hoc*. Esta métrica considera el número total de reinas posicionadas y se la denominó MQUEENS (millones de reinas posicionadas).

Como se puede apreciar en la Tabla 1, en ambos cluster y para todas las aplicaciones, el rendimiento global del sistema mejora al aumentar tanto el tamaño del problema como de la arquitectura. Como era de esperar, el cluster x86 obtiene rendimientos superiores en las 3 aplicaciones. Esto se debe a que sus procesadores son más potentes y sofisticados, sus memorias RAM son más rápidas, su interfaz de red es más veloz, entre otras características más avanzadas de este cluster. Por otra parte, a nivel de aplicación, la mayor diferencia se encuentra en MM con una diferencia promedio de 10.4× y una máxima de 18.2×, seguido por Jacobi con promedio de 9× y máxima de 18.8×; y por último, N-Reinas donde la diferencia es mucho menor, promedio de 4× y máxima de 4.7×.

En forma similar al análisis anterior, para poder evaluar la eficiencia energética se empleó la métrica FLOPS/Watt para MM y Jacobi, mientras que MQUEENS/Watt fue utilizada para N-Reinas. Para estimar los Watt consumidos por el cluster x86, se empleó el valor observado de consumo de su procesador en diferentes estudios realizados [10,17], el cual a su vez coincide con la Potencia de Diseño Térmico (84W). Por otro lado, en el caso del cluster de RPi, se optó por emplear el valor observado como máximo en diferentes reportes disponibles (3.7W) [22,16,21].

La Tabla 2 muestra la eficiencia energética de las aplicaciones benchmark en ambos clusters. A diferencia del análisis de rendimiento, es el cluster de RPi quien obtiene los mejores cocientes de eficiencia energética para todos las combinaciones (N,P). Esto se debe en gran medida a su bajo consumo de potencia. En particular, obtiene una diferencia promedio de 2.5× y una máxima de 5× para MM, 3.1× con una máxima de 7.2× para Jacobi y 6.5× con máxima de 21.4× para N-Reinas. Es claro que la mayor diferencia se da en N-Reinas, ya que justamente es la aplicación en la que el cluster de RPi obtiene el mejor rendimiento y la menor diferencia con el cluster x86.

Resulta importante mencionar 2 cuestiones que podrían incidir en los resultados encontrados. La primera es desfavorable para el cluster RPi debido a que existen procesadores x86 con mejores tasas de eficiencia energética (por ejemplo, familia Xeon de Intel). Por el contrario, la segunda cuestión es beneficiosa para el cluster de RPi, ya que en el análisis realizado se consideró el consumo completo de la placa, mientras que en el de x86, sólo el del procesador.

5. Trabajos Relacionados

Existen varios trabajos que realizan análisis de rendimiento y/o eficiencia energética de cluster de RPi's, además de comparaciones con otras arquitecturas.

Kiepert [19] describe el armado de un clúster de 32 RPi's. Para evaluar el rendimiento del mismo, ejecutó una aplicación MPI que estima el valor del número Pi. Al ser una aplicación *embarazosamente paralela*, el cluster obtiene aceleraciones muy cercanas al número de nodos empleados.

Die [15] desplegó un cluster de 8 RPi's, ejecutando OpenFOAM [2] y High-Performance Linpack (HPL) [1] para el análisis de rendimiento. El cluster armado mostró buenos rendimiento para OpenFOAM en los casos en que el volumen de comunicaciones era bastante menor que el de cómputo. Respecto a HPL, Die encontró niveles de prestaciones aceptables mientras los requerimientos de memoria no superaron la memoria RAM disponible. En este estudio se observaron resultados similares, aunque al contar con más memoria RAM en cada placa, fue posible procesar problemas de mayor tamaño.

Cox et al [14] armaron un cluster de 64 RPi modelo B (usando uno de ellas como *frontend*). Al igual que el caso anterior, emplearon HPL para evaluar el rendimiento alcanzando 1.14 GFLOPS con el sistema completo. En forma similar a este trabajo, observaron que la capacidad de memoria de las placas, impiden computar problemas más grandes.

P	N		
	1024	2048	4096
4	0,12	0,14	0,14
8	0,10	0,12	0,14
16	0,07	0,09	0,12
32	0,04	0,06	0,10
64	0,02	0,04	0,05
GFLOPS/Watt			

(a) MM

P	N			
	1024	2048	4096	8192
4	0,20	0,18	0,18	0,17
8	0,12	0,15	0,17	0,17
16	0,09	0,12	0,15	0,16
32	0,06	0,08	0,13	0,15
64	0,03	0,04	0,06	0,08
GFLOPS/Watt				

(b) Jacobi

P	N			
	16	17	18	19
4	18,90	18,94	18,98	19,10
8	20,90	21,96	22,01	22,29
16	12,71	19,23	23,34	23,87
32	16,94	22,97	23,23	24,63
64	8,83	14,39	21,23	21,40
MQUEENS/Watt				

(c) N-Reinas

P	N		
	1024	2048	4096
4	0,09	0,10	0,10
8	0,04	0,07	0,08
16	0,02	0,04	0,05
32	0,01	0,02	0,04
64	0,00	0,01	0,02
GFLOPS/Watt			

(d) MM

P	N			
	1024	2048	4096	8192
4	0,08	0,04	0,05	0,05
8	0,04	0,04	0,05	0,05
16	0,02	0,06	0,07	0,07
32	0,01	0,04	0,07	0,07
64	0,00	0,02	0,06	0,06
GFLOPS/Watt				

(e) Jacobi

P	N			
	16	17	18	19
4	3,90	3,90	3,92	3,94
8	4,04	4,47	4,56	4,59
16	2,64	3,82	4,01	4,04
32	1,54	3,36	4,01	4,14
64	0,41	2,05	3,73	4,18
MQUEENS/Watt				

(f) N-Reinas

Tabla 2: Eficiencia energética de las aplicaciones benchmark en el cluster RPi (a-c) y en el x86 (d-f).

Un cluster de 25 RPi modelo B fue construido por Pfalzgraf y Driscoll [23]. Usando 3 aplicaciones de álgebra lineal (triada de vector, multiplicación de matriz-vector y multiplicación de matrices) analizaron el rendimiento del cluster armado. En el caso de la multiplicación de matrices, logran un pico de 1.1 GFLOPS con 24 placas. En esta investigación, se alcanza un pico de 3.12 GFLOPS con 16 placas, aunque obviamente se debe tener en cuenta la mayor potencia de las RPi 3 comparado a la RPi modelo B.

Cloutier et al [13] armaron un cluster de 25 RPi 2 modelo B. Una vez más, HPL fue empleado para el análisis de rendimiento. En este caso además, evaluaron la eficiencia energética del cluster midiendo el consumo a través de un medidor WattsUp Pro. En la comparación con un sistema basado en un Intel Haswell EP de 16 núcleos, observan que el cluster de RPi se ve superado por el sistema x86 en forma opuesta a lo observado al momento. Los autores indican que entre las principales causas se encuentra el profundo nivel de optimización de HPL como benchmark para medir rendimiento y la alta tasa de operaciones en punto flotante, lo que favorece a los procesadores x86 tradicionales.

Cicirello [12] construyó un cluster de RPi 3 modelo B, seleccionando como casos de estudio la estimación del número Pi y la multiplicación de matrices. Lamentablemente, la ausencia de tiempos de ejecución o FLOPS impide la comparación directa con los resultados de este estudio. Más allá de eso, y en sentido similar a este trabajo, Cicirello detecta que el rendimiento del cluster está significativamente limitado por la velocidad de la interfaz de red en aquellas aplicaciones que tienen comunicaciones de gran volumen. Además, encuentra aceptables tasas de aceleraciones en aplicaciones donde la cantidad y el volumen de comunicaciones tienen mucho menor peso que la proporción de cómputo.

Por último, los mismos autores de [13] desplegaron un cluster de RPi 3 B en [20]. Al igual que en su trabajo previo, el uso de HPL como caso de estudio los lleva a encontrar que los sistemas x86 resultan superiores no sólo en rendimiento pico sino también en rendimiento/watt.

6. Conclusiones y Líneas de Trabajo Futuro

En la actualidad, la reducción del consumo energético se ha vuelto uno de los principales desafíos para la comunidad HPC. Entre las diferentes alternativas actualmente exploradas, se encuentra el uso de procesadores ARM. En este trabajo se estudia la viabilidad del uso de un cluster de RPi. Del análisis realizado sobre los resultados obtenidos se puede concluir que:

- Aquellas aplicaciones que demandan comunicaciones de gran volumen, probablemente sufran una degradación importante en su rendimiento debido a la limitada velocidad de la interfaz de red que tiene la RPi, como se observó en MM.
- Aquellas aplicaciones que requieran una gran cantidad de comunicaciones de tamaño moderado, también penalizarán su rendimiento por la misma causa que el caso anterior, como se apreció en Jacobi.
- Aplicaciones que trabajen con volúmenes de datos muy superiores a la capacidad disponible de la memoria RAM de las RPi, pueden sufrir errores en ejecución, tal como se vio en MM.
- Aplicaciones que demanden mucho cómputo pero con pequeños datos, que no tengan una incidencia significativa de comunicaciones, probablemente tengan muy buen rendimiento. Esto se ve reflejado en el caso de N-Reinas, donde sus resultados muestran niveles altos de eficiencia.

Como era de esperar, desde el punto de vista del rendimiento, el cluster RPi presenta tasas de prestaciones más bajas en comparación a un cluster x86 (entre 4 y 10.4× de diferencia promedio). Sin embargo, desde la perspectiva de la eficiencia energética, el cluster de RPi alcanza mejores resultados (entre 2.5 y 6.5× de diferencia promedio). Considerando lo analizado, es posible decir que los cluster de RPi ofrecen capacidades para cómputo paralelo y pueden obtener tasas de eficiencia energética superiores a los clusters x86. Sin embargo, debido a la gran diferencia de rendimiento pico que tiene con estos últimos, no representan hoy una opción viable para HPC si el objetivo principal es la reducción del tiempo de ejecución.

Algunas de las líneas de trabajo futuro que se desprenden de esta investigación pueden ser:

- Considerando que existen placas más avanzadas que la RPi 3 (como pueden ser la Banana Pi M3 [6] o la ROCK64 [8]), interesa replicar el estudio realizado para continuar la búsqueda de alternativas viables basadas en procesadores ARM.
- Como el consumo real puede diferir del estimado, sería interesante medir realmente el consumo en ambos clusters para hacer un estudio más preciso del balance rendimiento/energía.

Referencias

1. Hpl. http://www.netlib.org/benchmark/hpl
2. Openfoam. https://www.openfoam.com
3. PandaBoard. https://en.wikipedia.org/wiki/PandaBoard (2017)
4. Sitio Web Beagleboard. https://beagleboard.org/ (2017)
5. TOP500. https://www.top500.org (2017)
6. Banana Pi M3. http://www.banana-pi.org/m3.html (2018)
7. Raspberry. https://www.raspberrypi.org (2018)
8. ROCK64. https://www.pine64.org/?page_id=7147 (2018)
9. Balakrishnan, N.: Building and benchmarking a low power ARM cluster. Ph.D. thesis, University of Edinburgh (2012)
10. Benchmark-tests: Intel core i5-4460. https://tinyurl.com/ybduwtu2
11. Chaudhari, H.: Raspberry pi technology: A review. International Journal of Innovative and Emerging Research in Engineering (IJIERE) pp. 83–87 (2015)
12. Cicirello, V.A.: Design, configuration, implementation, and performance of a simple 32 core raspberry pi cluster (08 2017)
13. Cloutier, M.F., Paradis, C., Weaver, V.M.: Design and analysis of a 32-bit embedded high-performance cluster optimized for energy and performance. In: 2014 Hardware-Software Co-Design for HPC. pp. 1–8 (Nov 2014)
14. Cox, S., Cox, J., Boardman, R., et al: Iridis-pi: a low-cost, compact demonstration cluster. Cluster Computing 17(2), 349358 (Jun 2014)
15. Die, B.: Distributed computing with the Raspberry Pi. Ph.D. thesis, Kansas State University (2014)
16. Dramble, R.P.: Power consumption benchmarks. https://tinyurl.com/jy7fvzc
17. Hardware.Info: Amd vs intel: 57 processor megatest. https://goo.gl/hxZkTX
18. Jarus, M., Varrette, S., Bouvry, A.O..P.: Performance evaluation and energy efficiency of high-density hpc platforms based on intel, amd and arm processors. Lecture Notes in Computer Science 8046, 182–200 (2013)
19. Kiepert, J.: Creating a raspberry pi-based beowulf cluster. https://tinyurl.com/y778qkvx
20. M. F. Cloutier, C.P., Weaver, V.M.: A raspberry pi cluster instrumented for fine-grained power measurement. Electronics 5(4) (2016)
21. Magazine, T.M.: Raspberry pi 3 is out now! specs, benchmarks & more. https://www.raspberrypi.org/magpi/raspberry-pi-3-specs-benchmarks
22. McDonald, S.: Raspberry pi 3 a first look. https://tinyurl.com/y9z4roje
23. Pfalzgraf, A.M., Driscoll, J.A.: A low-cost computer cluster for high-performance computing education. In: IEEE International Conference on Electro/Information Technology. pp. 362–366 (June 2014)
24. Wolfe, M.: Arm yourselves for exascale. https://tinyurl.com/y7gus9ru (2011)

Workshop Tecnología Informática Aplicada en Educación (WTIAE)

Coordinadores

Gladys Gorja (UNLP) - Alejandra Malberti (UNSJ) - Claudia Russo (UNNOBA)

Temas

Diseños y desarrollos de TICs orientadas a la Educación
Análisis de nuevas TICs para aplicar en la educación
Entornos y ambientes de soporte para Educación a Distancia
Hipermedias educativas
Laboratorios Virtuales
Simulación, Visualización y Realidad Virtual aplicadas en Educación
Aplicaciones educativas de la Inteligencia Artificial
Laboratorios Remotos
Métricas de Calidad de productos TICs aplicados a educación
Tecnología para ambientes e-learning
Tecnología para ambientes m-learning

Validación de contenido de instrumentos para medir el nivel de integración tecnológica en el aula y el nivel de impacto en los estudiantes

Marisa Cecilia Tumino, Juan Manuel Bournissen y Franco Forneron
Universidad Adventista del Plata, Libertador San Martín, Entre Ríos, Argentina
{marisa.tumino, juan.bournissen}@uap.edu.ar

Resumen. El trabajo de investigación tiene como propósito incentivar el uso de aplicaciones informáticas por parte de los docentes universitarios. Para ello en esta primera etapa se pretende diseñar y validar un instrumento que permita medir el nivel con que los docentes integran las TIC en el aula, de acuerdo al modelo de Adell [1]. Por su parte, se busca interpretar el impacto que el nivel de integración de las TIC en el aula produce en los estudiantes. En este sentido se trabaja en una escala de evaluación de impacto desde la propuesta de Abdala [2], quien entiende la evaluación de impacto como "el proceso evaluativo orientado a medir los resultados de las intervenciones, en cantidad, calidad y extensión según las reglas preestablecidas" (p. 28-29). La participación de expertos evaluadores en las dos fases de validación de contenido de las escalas, hacen posible el avance del proyecto. Se espera en su etapa final, que el trabajo permita descubrir los beneficios de la aplicación de tecnologías por partes de los docentes para todos los actores involucrados en el proceso de enseñanza y aprendizaje.

Palabras clave: TIC, integración tecnológica, evaluación de impacto.

1. Introducción

Desde sus orígenes, las universidades se encuentran involucradas en la transformación de los procesos de enseñanza y aprendizaje. Como bien lo señala Unigarro [3], la educación se encuentra sometida al devenir histórico y, consecuentemente, a los cambios continuos derivados de la configuración que va adoptando la humanidad con el paso del tiempo… "va sufriendo adaptaciones sucesivas que le permiten mantenerse vigente". Como lo declara Pontes Pedrajas [4], las tecnologías de la información y la comunicación ejercen una influencia cada vez mayor en la educación científica, no sólo en la mejora del aprendizaje de la ciencia, sino también en la formación permanente del docente.

Las instituciones educativas deben asumir que la tiza y la pizarra no son las únicas herramientas didácticas y comprender que "incorporar otras alternativas seguramente permitirá establecer relaciones educativas que superen la unidireccionalidad que se da tradicionalmente entre maestro y estudiante" Unigarro [3].

En respuesta a las demandas sociales de oportunidades de capacitación o formación profesional adaptadas a sus necesidades, el estudio propone no solo ofrecer información relevante que conduzca a la capacitación de los docentes en la aplicación

de nuevas TIC, sino especialmente extraer información de la evaluación del impacto que la implementación de estas TIC tiene sobre el desarrollo de las clases.

1.1. Impacto de las tecnologías en la educación

El desarrollo de todas las áreas involucradas con la vivencia humana ha embebido los beneficios que brindan las Tecnologías de la Información y de las Comunicaciones (TIC). Como lo declaran Prieto Díaz y colaboradores [5], en la educación "se descubre un universo ilimitado de posibilidades como recurso para el aprendizaje con la posibilidad de expandirse a un mayor número de usuarios de diferentes escenarios con la capacidad de socializar el conocimiento".
Luego de un estudio de corte cuantitativo, López de la Madrid [6], concluye que "con la integración de las TIC en la educación superior, los docentes han venido generando nuevas estrategias y actividades que hasta hace pocos años no se tenían contempladas".
Prieto Díaz y colaboradores [5] reconocen los impactos del desarrollo y la implementación de las TIC sobre la universalización de la educación superior, como "fuente de información, canal de comunicación, instrumento cognitivo y de procesamiento de la información". Esto implica la adopción de nuevos roles para los actores del proceso educativo y retos no solo para formadores y estudiantes sino también para las instituciones educativas que "han de lidiar con aspectos técnicos, formación especializada, seguridad informática y otros elementos que determinan la expansión de las TIC´s". Los autores concluyen que las nuevas TIC brindan herramientas que, apropiadamente orientadas por entornos de aprendizaje, pueden potenciar la formación de los profesionales competentes que demanda la sociedad.
Domingo Coscollola y Fuentes Agustó [7], luego de un estudio que procuraba identificar las ventajas e inconvenientes del uso de las TIC en los procesos de enseñanza y aprendizaje, un 90% de los participantes señala que las TIC representan una "renovación metodológica innovadora que propicia un aumento de la motivación y participación de los estudiantes, que facilita su comprensión y el aprendizaje en general, que proporciona nuevos recursos educativos, y que aumenta la satisfacción, motivación y autoestima del docente." (p. 175)
Por su parte, Peinado Rocamora y Navarro Ardoy [8], en un estudio que pretendió analizar si ciertas pautas de comportamiento en el uso de estas tecnologías, junto con la inmersión en redes sociales, mejoraba el rendimiento académico y motivación en una muestra de 46 estudiantes, concluyen que el uso de las TIC incrementa la motivación lo que sugiere una mejora en su rendimiento académico.
Abarca Amador [9] en un trabajo que tuvo la intención de indagar lo que motiva a los docentes a utilizar TIC, señala que la motivación de los estudiantes es el motor que hace que los docentes utilicen las TIC. Asimismo los docentes mostraron su motivación al respecto, lo que conlleva su deseo de actualizarse e innovar las clases. Concluye que el uso de las TIC es un recurso valioso para promover el desarrollo integral de los estudiantes universitarios.
En virtud de las ventajas que evidencian los estudios respecto de la implementación de las TIC, se considera oportuno describir algunos de los recursos que estas tecnologías ponen al servicio de la educación.

1.2. Integración tecnológica en docentes

Emprender procesos de alfabetización digital pertinentes espacial y temporalmente, con el objetivo de impartir una formación integral, resulta para Alcántara Blas [10] un reto para la educación que permite la inserción en la sociedad de la información y del conocimiento. Varios autores han propuesto modelos descriptivos de la incorporación de la información, y del conocimiento, con la intención de explicar los cambios en el rol docente. Alcántara Blas [10] ensambla modelos, reconociendo que el conocimiento que requiere el docente para enseñar con, y sobre TIC, considera el diseño instruccional, pedagógico y tecnológico, "además de procesos de integración de las TIC y dominio de plataformas digitales".
El término de apropiación tecnológica comúnmente se relaciona con los cambios que se producen en los docentes como resultado de la implementación de tecnologías en sus clases, convirtiendo el trabajo docente en algo diferente (:494). Para Colás Bravo, Rodríguez López y Jiménez Cortés [11] la apropiación plantea la forma en que las herramientas culturales tecnológicas son asumidas por los sujetos y estructuran sus maneras de interpretar la realidad como la base de su aprendizaje. El proceso de apropiación de la cultura del e-learning implica el traslado del control del uso de herramientas tecnológicas desde los contextos hacia los individuos. La apropiación se proyecta y expresa empíricamente en iniciativas de índole pedagógica emprendidas por el profesorado Colás Bravo y Jiménez Cortés [11]. Según Cisneros Gallart [12], las investigaciones apuntan a la falta de confianza del profesorado en el uso de tecnologías, como barrera para la integración educativa de las TIC.
Riascos-Erazo, Quintero-Calvache y Ávila-Fajardo [13] advierten que los docentes no solo debe contar con los recursos tecnológicos sino también con las competencias que demanda la aplicación eficiente de las TIC, adaptada a los diferentes escenarios que deben abordar en el aula.
Dada la existencia de variados modelos de apropiación o de integración de las TIC al desarrollo formativo, es pertinente adoptar un modelo a los efectos de clarificar las etapas que transitan los educadores en el proceso de implementación de las tecnologías.

1.2.1. Modelo de integración de las TIC a la educación

Dentro de los modelos propuestos por diferentes autores, y a los efectos de identificar las fases de apropiación de las TIC en las aulas, se adopta el modelo de Adell [1] para quien el proceso de apropiación de las TIC tarda de 3 a 5 años, considerando que las fases por las que pasan los docentes son las siguientes:
1. Acceso: aprender el uso básico de la tecnología; proceso que insume un año.
2. Adopción: los docentes usan las PC para hacer lo mismo que hacían sin ellas, sin innovación didáctica.
3. Adaptación: se integra la nueva tecnología en prácticas tradicionales pero aumentando la productividad, el ritmo y la cantidad de trabajo. Los estudiantes comienzan a utilizar aplicaciones.

4. Apropiación: experimentan nuevas maneras de trabajar didácticamente; abriéndose a posibilidades que no serían posibles sin la tecnología. Incorporan naturalmente tecnologías en las experiencias.
5. Innovación: es la fase a la que no alcanzan todos y se presenta cuando el docente utiliza la tecnología de manera que nadie lo hizo antes. Crea experiencias nuevas con las prestaciones que ofrecen las tecnologías.
Para la medición del nivel de integración de las TIC en el aula por parte del docente, se propuso una rúbrica que pretende conducir las entrevistas que los investigadores mantendrán con los docentes, una vez finalizadas las observaciones en la clase donde se apliquen las tecnologías seleccionadas por los docentes.

1.3. Evaluación de impacto

Riascos-Erazo, Quintero-Calvache y Ávila-Fajardo [13], mediante la aplicación de evaluación de impacto, concluyen que las TIC "fortalecen considerablemente la aprehensión de conocimiento y el enriquecimiento cognitivo como el resultado de la percepción y el grado de utilización de las TIC en el proceso de enseñanza-aprendizaje". (p. 154)
"Por evaluación de impacto se entiende el proceso evaluativo orientado a medir los resultados de las intervenciones, en cantidad, calidad y extensión según las reglas preestablecidas." Abdala [2].
Según este autor, la medida de las características principales de la evaluación de impacto, centrada en los resultados, permite comparar el grado de cumplimiento alcanzado con el logro deseado. Mediante la evaluación de impacto se compara la planeación con el resultado de la ejecución del plan. Esta modalidad "abarca todos los efectos secundarios a la planeación y a la ejecución: específicos y globales, buscados (según los objetivos) o no; positivos, negativos o neutros; directos o indirectos" Abdala [2], con lo que la evaluación cuantitativa se robustece al incorporar técnicas cualitativas.
Para la evaluación del nivel de impacto que la implementación de las TIC tiene sobre el proceso de aprendizaje de los estudiantes en el aula, se adecuaron modelos como los implementados por Riascos-Erazo, Quintero-Calvache y Ávila-Fajardo [13], o Balas-Nakash, Rodríguez-Cano, Muñoz-Manrique, Vásquez-Peña y Perichart-Perera [14] y se consideran los indicadores detallados en el Anexo 2.

2. Objetivos del estudio

Los objetivos del presente proyecto se traducen en construir y obtener evidencia de validación de contenido de: (a) una rúbrica que pretende medir el nivel de integración tecnológica de las aplicaciones informáticas en el aula y (b) una escala que intenta medir el nivel de impacto de la integración de las TIC en los estudiantes.

3. Metodología

En primero lugar, la rúbrica propuesta para valorar el nivel con que el docente integra las TIC en el aula, partió del modelo de Adell [1] y que consta de cinco fases denominadas, tal como se describió previamente, (a) acceso, (b) adopción, (c) adaptación, (d) apropiación y (e) innovación. Se adoptaron diez criterios que abarcaran aspectos involucrados con la integración de las TIC en el aula. Posteriormente se propusieron, para cada uno de los criterios, los indicadores ajustados al nivel de integración descrito por Adell [1]. La rúbrica en su primera versión se envió a expertos, quienes fueron invitados a evaluar en primera instancia los reactivos o indicadores de la rúbrica y sugerir las modificaciones que consideraran oportunas (modificación, eliminación o inclusión de criterios y/o indicadores). Una vez concebida la versión final de la rúbrica, luego de atender a las sugerencias de los expertos, en una segunda instancia se les solicitó la valoración de la claridad y pertinencia de los indicadores a fin de obtener evidencias de la validación de contenido. Actualmente la investigación se encuentra a la espera de la recepción de estas valoraciones para finalmente verificar el grado de acuerdo entre los expertos mediante la V de Aiken, como una de las técnicas que permite cuantificar la claridad y pertinencia de cada ítem respecto de un dominio de contenido formulado por jueces. Su valor oscila desde 0 a 1, siendo el valor 1 el indicativo de un acuerdo perfecto entre los jueces. La interpretación del coeficiente se centra en la magnitud calculada y en el nivel de significación estadística obtenido ($p<0,05$). Como criterio para mantener un ítem se asumirá que el índice no sea inferior a 0,80. Cabe señalar que el diseño de este instrumento se adecua para orientar al evaluador de la clase tanto durante las observaciones como durante la entrevista mantenida con el docente al finalizar la observación.

En segundo lugar, y a los efectos de construir la escala propuesta para evaluar el impacto de la integración de las TIC en los estudiantes, se confeccionaron ítems, teniendo en cuanta las ideas de Abdala [2], juntamente con modelos propuestos por Riascos-Erazo, Quintero-Calvache y Ávila-Fajardo [13], o Balas-Nakash, Rodríguez-Cano, Muñoz-Manrique, Vásquez-Peña y Perichart-Perera [14]. El borrador de la encuesta así obtenida en su primera versión se envió a expertos, quienes fueron invitados a evaluar en primera instancia los ítems y sugerir las modificaciones que consideraran oportunas (modificación, eliminación o inclusión de ítems). Una vez obtenida la versión final de la escala, en una segunda instancia, se solicitó la valoración de la claridad y pertinencia de cada ítem a fin de obtener evidencias de la validación de contenido. Actualmente la investigación se encuentra a la espera de la recepción de estas valoraciones para finalmente verificar el grado de acuerdo entre los expertos mediante la V de Aiken, como una de las técnicas que permite cuantificar la claridad y pertinencia de cada ítem respecto de un dominio de contenido formulado por jueces. Es importante destacar que mientras la rúbrica sería operada por evaluadores externos a la clase, la escala de impacto es diseñada para ser respondida por los estudiantes de la misma clase.

4. Resultados descriptivos

Hasta el momento, los diez expertos que participaron del estudio han respondido sugiriendo modificaciones en todos los ítems del cuestionario y la inclusión de cinco nuevos ítems, quedando conformado en esta primera etapa por 22 ítems. Por su parte, los mismos expertos sugirieron modificaciones en algunos indicadores de la rúbrica sin agregar ni eliminar criterios a la misma. La segunda fase del proceso permitirá obtener la versión final de ambos instrumentos; sin embargo en este informe se anexan las versiones actuales a los efectos de proveer el estado de avance de la investigación.

5. Discusión y conclusión

En esta primera etapa se pudo apreciar el grado de disposición de los expertos, quienes se sumaron al proyecto con gran expectativa, reflejando en cierto sentido la relevancia de la propuesta para la implementación futura de los instrumentos en proceso de construcción. Se estima que la propuesta permitirá advertir no solo el potencial de las TIC en los procesos de enseñanza y aprendizaje sino también el amplio camino de ensayo y prueba que los docentes pueden recorrer en bien de las experiencias de formación profesional de los estudiantes.

6. Referencias

1. ADELL, Jordi. Fases en la apropiación de las TIC por los docentes. Presentación magistral dptomatesbg (28/05/2008). Recuperado de: https://www.youtube.com/watch?v =URskJlRgd5w
2. ABDALA, Ernesto. Manual para la evaluación de impacto en programas de formación para jóvenes. OIT, CINTERFOR, 2004.
3. GUTIÉRREZ, Manuel Antonio Unigarro. Educación Virtual: Encuentro Formativo en el Ciberespacio. UNAB, 2004.
4. PONTES-PEDRAJAS, Alfonso. Aplicaciones de las Tecnologías de la Información y de la Comunicación en la educación científica. Primera parte: funciones y recursos. Revista Eureka sobre Enseñanza y divulgación de las Ciencias, 2005, vol. 2, no 1.
5. PRIETODÍAZ, Vicente, et al. Impacto de las tecnologías de la información y las comunicaciones en la educación y nuevos paradigmas del enfoque educativo. Educación Médica Superior, 2011, vol. 25, no 1, p. 95-102.
6. LÓPEZ DE LA MADRID, María Cristina. Impacto de las tecnologías de la información y la comunicación (TIC) en el docente universitario. El caso de la Universidad de Guadalajara. Perspectiva Educacional, Formación de Profesores, 2013, vol. 52, no 2.
7. DOMINGO COSCOLLOLA, María; FUENTES AGUSTÍ, Marta. Innovación educativa: experimentar con las TIC y reflexionar sobre su uso. Pixel-Bit. Revista de Medios y Educación, 2010, vol. 36, p. 171-180. Recuperado de: https://idus.us.es/xmlui/bitstream /handle/11441/22612/file_1.pdf?sequence=1&isAllowed=y
8. PEINADO ROCAMORA, Pedro; NAVARRO ARDOY, Daniel. Aumento de la motivación mediante el uso de redes sociales. Didáctica, innovación y multimedia, 2014, no 29, p. 0001-15.

9. AMADOR, Yohanna Abarca. El uso de las TIC en la educación universitaria: motivación que incide en su uso y frecuencia. Revista de Lenguas Modernas, 2015, no 22, p. 335-349.
10. ALCÁNTARA BLAS, Lizbeth. Procesos de apropiación tecnológica en docentes que integran plataformas educativas como apoyo a sus cursos presenciales, en ambientes de aprendizaje en el nivel medio superior.
11. COLÁS BRAVO, María Pilar; JIMÉNEZ CORTÉS, Rocío. Evaluación del impacto de la formación (online) en TIC en el profesorado. Una perspectiva sociocultural. Revista de educación, 2008, vol. 346, p. 187-215.
12. RIASCOS-ERAZO, Sandra Cristina; QUINTERO-CALVACHE, Diana María; ÁVILA-FAJARDO, Gloria Patricia. Las TIC en el aula: percepciones de los profesores universitarios. Educación y educadores, 2009, vol. 12, no 3.
13. RIASCOS-ERAZO, Sandra Cristina; QUINTERO-CALVACHE, Diana María; ÁVILA-FAJARDO, Gloria Patricia. Las TIC en el aula: percepciones de los profesores universitarios. *Educación y educadores*, 2009, vol. 12, no 3.
14. BALAS-NAKASH, Margie, et al. Tres métodos para medir la adherencia a un programa de terapia médica y nutricia en mujeres embarazadas con diabetes y su asociación con el control glucémico. Rev Invest Clin, 2010, vol. 62, no 3, p. 235-243.

Rúbrica: Nivel de Integración de las TIC en el Aula

Criterios	Acceso (1)	Adopción (2)	Adaptación (3)	Apropiación (4)	Innovación (5)
¿Con qué nivel de destreza opera las tecnologías aplicadas en las clases?	Conoce el uso básico de la tecnología aplicada	Adopta las tecnologías para hacer lo mismo que hacía en las clases sin ellas	Implementa las tecnologías para facilitar el cumplimiento de las prácticas tradicionales pero aumentando la productividad del trabajo	Con la implementación de las tecnologías en las clases experimentan nuevas maneras de trabajar didácticamente, creando nuevas posibilidades que serían imposibles sin tecnologías. Son capaces de seleccionar, diseñar y fundamentar apropiadamente tareas con TIC en el momento oportuno y grado necesario	Utiliza las tecnologías de manera innovadora, es decir, de un modo que nadie lo hizo antes
¿Con qué propósito usa las TIC en los procesos de enseñanza y aprendizaje	Cumplir con las demandas académicas	Generar interés en los estudiantes	Agilizar las sesiones de clase	Para hacer más lúdica y didáctica las sesiones de clase	Innovar las formas de enseñar y aprender con el objetivo de generar aprendizajes significativos
En qué aplica las tecnologías	Navegación en internet. Comunicación con los estudiantes (correo electrónico)	Compartir materiales de estudio (subir o bajar materiales)	Discusión en foros o elaboración de proyectos colaborativos	Procesamiento de datos masivos. Elaboración de contenidos en Internet	En creación de formas innovadoras de aprender
Elaboración de recursos didácticos	El docente elabora recursos didácticos en Power Point o algún otro software de presentación	Elaboración de recursos didácticos digitales en Texto plano	Elaboración de recursos didácticos Hipertextuales	Elaboración de recursos didácticos Multimediales e Hipermediales, seleccionando, diseñando y fundamentando apropiadamente los recursos con TIC	Elaboración de recursos didácticos de forma innovadora mediantes las tecnologías
Integración TIC en los Programas de Asignaturas	Incluyen en el programa alguna actividad donde se requiere el uso de tecnología básica	En el programa incluye el uso frecuente de las tecnologías básicas	Desde el programa promueve el uso transparente de las tecnologías en la clase	Aplicaciones especializadas de la disciplina (simuladores, video conferencias u otras aplicaciones que faciliten la interacción entre pares) en el momento oportuno y de acuerdo a las necesidades	Propone en el programa la utilización Innovadora de tecnologías
Promueve en los estudiantes el uso de tecnologías para la entrega de trabajos	Promueve la entrega de trabajos elaborados mediante procesadores de texto	Promueve la presentación de los trabajos con el uso de ofimática	Promueve la elaboración de trabajos digitales basados en la búsqueda analítica de información en Internet	Promueve la creatividad en la entrega de los trabajos bajo diferentes formatos (videos, proyectos informáticos)	Promueve la creatividad en la entrega de los trabajos bajo formatos innovadores
Tutoría	El docente no utiliza TIC en las tutorías	El docente utiliza su dispositivo móvil para las entrevistas tutoriales con los estudiantes	El docente utiliza una plataforma virtual para las tutorías	El docente utiliza software especializados en tutorías	El docente administra las tutorías de forma innovadora mediante las tecnologías
Creación de redes de aprendizaje	Utiliza las tecnologías para el trabajo individual de los estudiantes	Utiliza las tecnologías para el trabajo grupal de los estudiantes	Promueve la creación de comunidades virtuales de aprendizaje entre sus estudiantes	Promueve la creación de comunidades virtuales de aprendizaje entre estudiantes de diferentes instituciones	Promueve la creación de comunidades virtuales de aprendizaje entre estudiantes de diferentes países
Publicación de materiales educativos digitales	Comparte su material digital con sus estudiantes	Comparte su material digital con los miembros de su institución	Publica su material didáctico en Internet	Elabora sus MOOC o Recursos Educativos Abiertos	Publica sus materiales utilizando recursos tecnológicos innovadores
El rol de las tecnologías en la evaluación	El docente utiliza las TIC en la enseñanza pero no al evaluar. La evaluación es presencial, oral o en papel	Las evaluaciones se realizan digitalmente utilizando las tecnologías básicas.	Se evalúa en línea utilizando alguna plataforma virtual con exámenes de corrección automática	Se evalúa en forma colaborativa mediante foros, talleres, wikis	Se evalúa de forma innovadora o evalúa trabajos que muestren innovación tecnológica
Totales					

Escalas 10 – 15 fase 1
16 – 25 fase 2
26 – 35 fase 3
36 – 45 fase 4
45 – 50 fase 5

Adell, J. dptomatesbg (28/05/2008). Fases en la apropiación de las TIC por los docentes. Recuperado de: https://www.youtube.com/watch?v=sLLlwJcQ--Y

Nivel de impacto de las Tecnologías de la Información y de las Comunicaciones (TIC) en el aula

(Adherencia al programa de actividad curricular)

"Por evaluación de impacto se entiende el proceso evaluativo orientado a medir los resultados de las intervenciones, en cantidad, calidad y extensión según las reglas preestablecidas." (Abdala, 2004: 28-29)
Con esta herramienta se pretende medir el nivel de impacto que la implementación de las TIC tiene sobre el proceso de aprendizaje de los estudiantes en el aula.

Para cada consigna del cuestionario, marca una de las celdas según la siguiente escala:
1=Muy en desacuerdo; 2=En desacuerdo; 3=Neutro; 4=De acuerdo; 5=Muy de acuerdo.

Consignas	1	2	3	4	5
1. La utilización de herramientas tecnológicas me ha despertado un mayor interés por la clase					
2. Cuando utilizamos tecnologías en esta asignatura finalicé la experiencia con la sensación de haber aprendido temas que tienen significado en mi formación					
3. La aplicación de herramientas tecnológicas me facilitó responder a las consignas establecidas por el docente					
4. Considero que la implementación de las TIC en el aula mejora mi comprensión de los contenidos					
5. El uso de las TIC en esta asignatura promovió mi participación en las actividades propuestas					
6. La implementación de herramientas TIC me facilitó la comunicación con los docentes y pares					
7. La implementación de las TIC disminuye los riesgos propios del desempeño profesional en un escenario real					
8. La utilización de las TIC me facilitó el abordaje de los contenidos de la asignatura					
9. La forma en que se implementaron las TIC en el aula me ayudó a integrar los contenidos de la asignatura					
10. La forma en que se implementaron las TIC potenció mi autonomía en el proceso de aprendizaje (búsqueda en revisas o tutoriales, formulación de preguntas a compañeros, consultas con expertos, búsqueda de información en base de datos científicas...)					
11. La implementación de las TIC promovió el aprendizaje colaborativo con los compañeros y docentes (responsabilidad ante el otro, respeto por sus opiniones, tolerancia a mejores ideas)					
12. La implementación de TIC en el aula favorece el proceso de aprendizaje					
13. La implementación de las TIC ayuda a asimilar la complejidad de los contenidos					
14. Las TIC me ayudan en el desenvolvimiento profesional					
15. Las TIC constituyen un recurso innovador en el ámbito académico en el que me desarrollo					
16. La unidad curricular impartida se correlaciona apropiadamente con el uso de las TIC					
17. Como estudiante puedo seleccionar las TIC que creo más interesantes o que se adapten mejor a la temática tratada					
18. Las redes sociales permiten reforzar los contenidos de las clases					
19. La implementación de las TIC en el aula me ayudó a demostrar los conocimientos adquiridos en la asignatura					
20. La implementación de las TIC en el aula me ofreció diferentes contextos de aprendizaje					
21. La forma de evaluación propuesta cuando se utilizan TIC me pareció apropiada					
22. La implementación de herramientas TIC me ofreció diferentes contextos de tutoría					

Abdala, E. (2004). Manual para la evaluación de impacto en programas de formación para jóvenes. Montevideo: CINTERFOR.

Relato de experiencia de operación a distancia de un Reactor Nuclear para capacitación y entrenamiento

Walter Miguel Keil[1,1], Norma Adriana Chautemps[1], Carlos Murúa[1]

[1]Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de Córdoba.
Av. Vélez Sarsfield 1611, Córdoba, Argentina.
{Walter Miguel Keil, Norma Adriana Chautemps, Carlos Murúa}walter.keil@unc.edu.ar

Resumen. El objetivo de esta presentación consiste en difundir los resultados de las fases propias de las experiencias en formación y capacitación a distancia, en entrenamiento sobre operación de reactores nucleares y física de reactores. Al ser la investigación cualitativa esencialmente descriptiva y analítica, permitió abordar la compleja realidad de las prácticas haciendo hincapié en las perspectivas de los participantes. El trabajo consistió en realizar una de las prácticas previstas en reactores, con un sistema de video-conferencia para observar la operación, y con un sistema electrónico de adquisición de datos para trasmitir las señales de la instrumentación del reactor. A través de éste se obtuvieron los parámetros utilizados para realizar los cálculos de reactividad, tiempo de respuesta y aproximaciones a crítico del reactor. La experiencia se llevó a cabo con un grupo de profesionales que realizaban entrenamiento para obtener licencia de operación de reactores.

Palabras claves: Educación, Innovación, TIC, Laboratorio de Energía Nuclear.

1 Introducción

La Universidad Nacional de Córdoba cuenta con un reactor de investigación de baja potencia denominado RA-0 (Reactor Argentino número cero) destinado a la formación de recursos humanos en el ámbito de las radiaciones ionizantes y ciencias nucleares. El reactor RA-0 está equipado con dispositivos experimentales educativos y de formación específicos, siendo único en su tipo en la región para su uso en educación, investigación y entrenamiento. La sala de control está provista de un espacio interactivo y además cuenta con instalaciones conexas como lo son: un laboratorio de radionucleídos, un laboratorio de electrónica y un laboratorio en proyecto aún, para el análisis de activación de neutrones; todos ellos constituyen los elementos indispensables para las actividades educativas y de formación específica que se ofrecen a los interesados en el campo de la medicina y de la energía nuclear.

A esta instalación asisten para su formación específica personal ingresante y en entrenamiento de la Central Nuclear Embalse, estudiantes universitarios de las áreas de ingeniería mecánica, ingeniería eléctrica, ingeniería biomédica y también profesionales médicos donde se capacitan en la metodología y aplicación de radionucleídos para su uso en el diagnóstico en el área de la medicina nuclear.

La instrucción y formación se basa en unidades teóricas y prácticas. Su duración depende de las necesidades de los temas de los usuarios o destinatarios y cubren aspectos de la física nuclear, protección radiológica y teoría de reactores nucleares.

En un reactor de investigación la formación académica es una disciplina específica diferente al trabajo de investigación que se lleva a cabo en forma rutinaria en este tipo de instalaciones.

Aplicando la concepción de Weber [1]"….se buscaba en términos generales, que sean los profesores los que seleccionaran una experiencia de enseñanza propia que hubiera resultado significativa, interesante y pertinente al campo de la física de reactores por alguna razón desde el punto de vista de la enseñanza", se utilizó el siguiente práctico con moción de contextualizar el marco de trabajo en una experiencia específica.

2 Metodología

Para realizar el práctico de operación del reactor a distancias se dispone del siguiente equipamiento:

a) Reactor Nuclear
b) Sistema de adquisición de Datos del Reactor
c) Sistemas de Video conferencias
d) Pizarra electrónica
e) Software "Sistema RA0 Consult"

La experiencia en que nos basamos se refiere a la construcción de una clase de laboratorio por parte de los participantes de una práctica a distancia en la operación de un Reactor Nuclear. Para ello tomamos como referencia lo realizado por Weber en el 2005 [2].

En diciembre de 2016 personal de las centrales nucleares Atucha I y Atucha II utilizaron el Reactor Nuclear RA-0 en forma remota para realizar prácticas de Física de Reactores. Dichas prácticas se desarrollaron en el marco del curso "Capacitación Complementaria Inicial – Nivel 1", algunos de cuyos módulos (Física Nuclear, Física de Reactores, Seguridad Radiológica y Nuclear) son dictados mediante videoconferencia por personal del RA-0.

Con la asistencia de un docente en el aula, físicamente ubicada en el predio de la Central Nuclear Atucha, los alumnos realizaron, a través del manejo del reactor por parte de sus operadores en la consola del Reactor en la sede de la Ciudad Universitaria de Córdoba, prácticas de puesta a crítico por barras de control, y comprobaron la evolución de la población neutrónica con el reactor en estado subcrítico, crítico y supercrítico, con y sin fuente de neutrones insertada, utilizando para ello el sistema "RA0 Consult" en su versión síncrona.

A continuación se muestran imágenes del trabajo en el aula:



Fig.1- Los alumnos en la C.N. Atucha

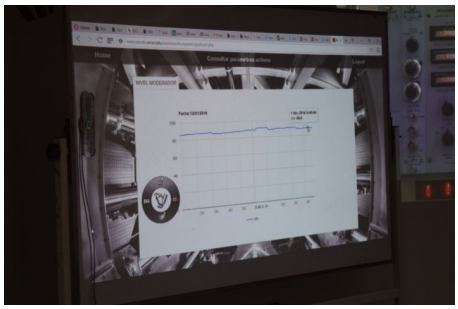


Fig.2- Usando el RA-0 Consult

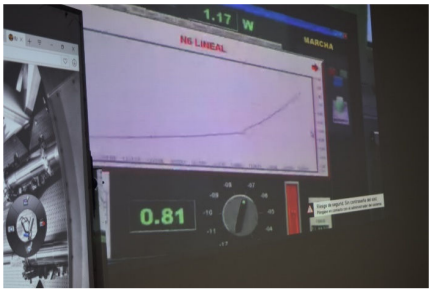


Fig.3- Visión aumentada de la pantalla en aula parámetros

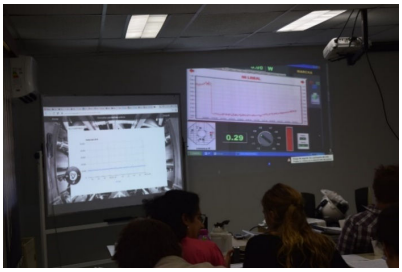


Fig. 4. Visión de otros



Fig4 - Vista desde la Consola de Operación del RA-0

3 Desarrollo de la experiencia

Título de la Experiencia: Configuraciones de Núcleos

Paso Nº1: Núcleo Nº 1

Con todas las barras de control extraídas y máximo nivel de agua, el Núcleo 1 propuesto por el jefe del reactor debe ser subcrítico. A partir de éste núcleo y hasta llegar al núcleo de primera criticidad, no debe cambiarse la posición de los detectores.

Paso Nº2: Núcleo Nº 2

Con el tanque inicialmente vacío y la fuente extraída, se irá agregando al núcleo anterior los elementos combustibles en las posiciones fijadas por el Jefe de Reactor, para llegar al núcleo 2.

Con los valores de contaje correspondientes a todas las barras extraídas y altura máxima de agua, se graficará para los núcleos medidos, la inversa de contaje normalizado en función del número de elementos combustibles.

Paso Nº3: Núcleo Nº 3 y siguientes

Alcanzada la condición de criticidad, se quita la fuente externa. Cumpliendo con las normas de diseño de conjuntos críticos y el procedimiento de operación, se elevará la potencia y se pasará al canal de marcha. Se medirá el exceso de reactividad mediante el método de período siguiendo el procedimiento para la calibración de barras en el núcleo de arranque del conjunto crítico RA-0 extrayendo la barra A o subiendo la altura de agua y compensando con la barra B. El núcleo resultante se denomina "núcleo de primera criticidad".

Nota: Se verificará que para este núcleo no se exceda el valor de 0.4$ para el exceso de reactividad. De no cumplirse esta condición, el Jefe del Reactor definirá un nuevo núcleo quitando elementos combustibles. Se procederá a medir el exceso de reactividad en forma similar al núcleo de primera criticidad.

Se colocará la tapa superior al reactor y se repetirán los pasos anteriores con el objeto de medir la influencia de la misma en el exceso de reactividad.

En la configuración, sin o con la tapa superior al reactor, que presente mayor exceso de reactividad, el Jefe de Reactor, propondrá el agregado o quitado de elementos combustibles para definir el núcleo de arranque cuyo exceso de reactividad no debe superar los 0.4$. Para medir este exceso de reactividad, se procederá en forma similar al núcleo de primera criticidad.

.

4 Evaluación de logros alcanzados

Para que se pueda interpretar el trabajo desarrollado presentamos sintéticamente las expectativas que traía el docente al curso y los aspectos que más nos marcaron de sus comentarios y que impactan en nuestro supuesto. Se espera que la didáctica provea de métodos o técnicas que puedan ayudar a construir mejores clases.

4.1 Actividades propuestas para la evaluación

Propiciamos un trabajo progresivo entre docente y alumnos cuya secuencia presentamos a continuación:
1. Se pide a los participantes que identifiquen la experiencia realizada y que reconozcan lo que fue interesante desde el punto de vista de la enseñanza. Esta actividad trajo sus complicaciones. Algunos alumnos, por ejemplo, no recordaban parte de la experiencia o les resultaba difícil recordar los sucesos. En otros casos, se hacía difícil que los alumnos se comprometieran con la actividad. Esto obstáculos pudieron sortearse en la mayor parte de los casos, porque fuimos aprovechando las construcciones de los colegas y ejemplificando con diferentes casos
Fue importante el trabajo que se hizo en el sentido de definir cuáles podrían ser las instancias adecuadas para su relato. Esto es, revisar las ideas de enseñanza, apropiarse de nuestra perspectiva de la didáctica, analizar qué resulta significativo y no caer en la polaridad de entender los relatos como "todo bueno" o "todo malo" Este tema nos acompañó a lo largo del laboratorio y consideramos interesante y adecuado que así haya sido.
2. Se solicita que se relate la experiencia implementada y realizada, intentando contextualizar su desarrollo. Para ello se elabora un guion. Esta actividad no ofreció en general dificultades, aunque como se destacó en el punto anterior, en algunos casos funcionó como promotora del trabajo anterior. Es decir se aprovechaban los avances de algunos colegas para que otros pudieran "comprender", interesarse, valorar la propuesta y se sintieran invitados a desarrollar la propia.
Esta actividad resultó en general muy enriquecedora. Los alumnos/ profesores se involucraban en la tarea. Construían interesantes relatos, aunque en algunos casos se caía en el relato de anécdotas estereotipadas que distaban de las que nos interesaba rescatar.
3. Se promueve la discusión respecto del caso relatado oralmente en grupos pequeños de pares. Esta discusión se organizó en torno del relato en tanto texto por una parte y de la experiencia relatada por otro. Si bien hubo diferencias entre los diferentes grupos esta actividad se desarrolló sin mayores dificultades. Cabe destacar aquí que hay un tiempo que lleva a los alumnos/ profesores a entender los aportes como tales y no como críticas infundadas ni como valoración de la persona. Se trabajó bastante sobre este punto, relacionándolo con el tema de la evaluación. Se planteó la importancia y el interés que los comentarios críticos pueden tener desde la perspectiva del aprovechamiento que pueden hacer de ellos los autores, en este caso, de los relatos, etc.
4. Se propicia la reflexión individual, a partir de la discusión y aporte de los colegas. En esta instancia se busca justamente aprovechar los aportes del apartado anterior, es decir, que cada uno pueda reflexionar respecto de las discusiones previas. Se promueve la reflexión acerca de los acuerdos y desacuerdos con los comentarios de los colegas, revisar incluso las posibilidades que desde el trabajo con la bibliografía, con otros casos, se puede hacer para la construcción de la narración escrita.
5. Se propone un tiempo para la escritura (un par de semanas). Se construyen algunas pautas que permitan no sólo redactar los acontecimientos, sino identificar puntos cruciales para el análisis En este punto resulta importante, trabajar con relatos de otros colegas, con casos extraídos de la bibliografía. Suelen valorar mucho las pautas de

escritura que se proponen en la bibliografía y se trabaja en clase. A lo largo de la experiencia descubrimos que era importante pautar muy bien los tiempos de entrega, porque esto actuaba como organizador de la propuesta a diferencia de las oportunidades en que lo dejamos más librado a las necesidades y posibilidades de cada uno.

6. Se desarrollan luego nuevas instancias de análisis del relato escrito, que permiten avanzar en la construcción del relato, en dotar de sentido el trabajo, de profundizar en aquellos aspectos conflictivos, etc. La posibilidad de discutir en esta instancia los escritos, pero sin que el autor pudiera modificar el relato. Otro colega leía el relato y en esta posición, la oportunidad de escuchar el relato leído por otro, llevaba al autor a distanciarse tanto de su obra que se le hacía posible encontrar nuevas lecturas críticas respecto de la experiencia, de escritura y del análisis que podía hacer sobre ambos aspectos.

Esta es una de las actividades que más valoraron en la marcha del proceso. A esta altura el docente hace una devolución escrita al trabajo. Con comentarios críticos respecto de la experiencia, el relato, siempre desde el punto de vista de la didáctica y proporcionando recursos para avanzar en la construcción desarrollada hasta el momento en cada uno de los planos indicados.

4.2 Resultados de la evaluación

Presentamos hasta aquí, las actividades propuestas y fuimos incluyendo comentarios, opiniones, valoraciones y ajustes en la actividad a partir de su puesta en práctica. Como recoge Pendlebury [3] de un trabajo de Confery: “Entro en ese contexto con mi programa preparado más o menos explícitamente y, como resultado de mi interpretación de las señales, permito, promuevo, desconozco, entiendo o me confundo. Llamo a esto una “práctica”, y las características que la definen son: su ritmo implacable, sus interacciones humanas, sus dimensiones normativas, sus objetivos contradictorios y su actividad.” (pp. 99).

Entendemos que es imprescindible generar un espacio en el cual los docentes puedan “mirarse” en su práctica. La necesidad de buscar estrategias que promovieran este tipo de reflexión surgió de la experiencia previa del profesor en el dictado del práctico.

Nos centramos en el análisis de la práctica referida desde algunas dimensiones teóricas que creemos pueden contribuir al desarrollo de este trabajo.

El modelo tecnicista de concebir la práctica ha logrado amplia difusión en el medio educativo, y de allí que se entienda la práctica como teoría aplicada. Es aquí donde surge la necesidad de rescatar la relación teoría – práctica, haciendo hincapié en que toda práctica tiene un referente teórico. Este referente teórico puede estar explicitado e identificado con diferentes líneas del pensamiento, o hacer parte de las creencias del docente, conformadas por las teorías implícitas, de las cuales no siempre se tiene conocimiento. En realidad, en toda práctica el referente teórico tiene un poco de cada uno de estos componentes, lo explícito y lo implícito, pero también tiene mucho de intuitivo. La relación teoría – práctica, según Cañal (1987), podemos entenderla en

términos de pensamiento y acción, con un enfoque individualista, que presupone una visión unilateral y racionalista, según la cual, las ideas orientan y dirigen la práctica.
Para desarrollar este análisis recurriremos a las cuatro dimensiones propuestas por Gudmundsdottir [4] para analizar los relatos y las narrativas de los docentes para abordar el concepto del saber pedagógico sobre los contenidos. Las dimensiones que elegimos teniendo en cuenta su pertinencia e interés para analizar nuestro caso son:

a) La experiencia práctica, b) la interpretación, c) la reflexión, d) la transformación.

5 Conclusiones

Los prácticos de laboratorio representan una figura destacada en las prácticas de enseñanza de los docentes, más allá de que para ellos se presente como una prescripción curricular. Los prácticos de laboratorio hacen parte de la propuesta programática, allí se indican cuáles y cuándo se deben realizar para cumplir con el desarrollo del práctico. La propuesta curricular genera un divorcio entre teoría y práctica, originando una ambivalencia que se percibe, tanto en el discurso de los docentes como de los alumnos. En la totalidad de los cuestionarios de los alumnos, se registró que la actividad práctica se realiza para aplicar lo visto en el curso teórico. El docente argumentaba las siguientes finalidades para la realización de actividades prácticas: motivación, aplicación o demostración de algún principio, recolección de datos para comprobar algo. También mencionan que para ellos las actividades prácticas más interesantes son aquellas que sorprenden al alumno, que despiertan la curiosidad. Los docentes seducen a los alumnos a través del discurso y la actividad práctica propuesta. El conocimiento práctico del docente, que sustenta todas sus decisiones de enseñanza, no siempre llega al plano de la verbalización, por eso tal vez surgen las aparentes contradicciones cuando hablan de guías que se aplicarían como recetas, pero que no condicen con lo observado. Es decir, la demanda de la propia enseñanza por ser una actividad práctica, hace que el docente actualice su conocimiento práctico, que es implícito, y cuyos enunciados formales aún no se han concretado. Todo este proceso se enriquece en situaciones de intercambio y socialización de las diferentes propuestas de enseñanza.

A lo largo del trabajo fuimos exponiendo diferentes aspectos que entendemos centrales en la construcción de nuestra propuesta. Intentamos no sólo tomar referentes teóricos sino también presentar una experiencia propia, para rescatar desde allí el valor que otorgamos a la estrategia presentada. Somos conscientes de que en la construcción de esta propuesta incluimos elementos de diversas perspectivas y esta característica trae aparejadas algunas dificultades y complicaciones. Sin embargo optamos por este modo de trabajo porque entendemos que pueden enriquecer nuestra propuesta y creemos que pueden ser de utilidad para revisar nuestra práctica.

Esperamos que se puedan encontrar en el trabajo elementos de interés para su discusión y análisis que permitan a otros y a nosotros seguir construyendo una propuesta, que valoramos junto con los colegas destinatarios de la misma. Incentivamos el uso del relato como la posibilidad de dar sentido a lo pensado y

programado y a lo desarrollado y encontrar en la didáctica aportes para construir mejores prácticas a partir de esa práctica reflexiva.

La buena enseñanza no refiere a la enseñanza exitosa en términos de alcanzar los objetivos sino a aquella que tenga sentido moral y epistemológico. A partir del aspecto moral, se indaga acerca de la justificación de las acciones docentes a partir de principios morales. Desde la perspectiva epistemológica la buena enseñanza supone justificar lo que se enseña de manera racional y que sea digno de que el estudiante lo conozca, lo crea, lo entienda. [5]

Referencias

1. Weber, V. El relato de experiencias como estrategia para la formación docente. CIVE Congreso Internacional Virtual de Educación. UBA. Universidad del Salvador. Buenos Aires. (2005)

2. Weber, V., (en colaboración con) Litwin, E.; Maggio, M.; Lipsman, M. (Coord.) Tecnologías en las aulas. Análisis de casos. Buenos Aires: Amorrortu. -En prensa. (2005).

3. Pendlebury, S. "Razón y relato en la buena práctica docente". En H. Mc.Ewan y K. Egan (comps) La narrativa en la enseñanza, el aprendizaje y la investigación. Buenos Aires: Amorrortu. (1998)

4. Gudmundsdottir, S. "La naturaleza narrativa el saber pedagógico". En H. Mc.Ewan y K. Egan (comps) La narrativa en la enseñanza, el aprendizaje y la investigación. Buenos Aires: Amorrortu. (1998)

5. Fenstermacher, G. "Tres aspectos de la filosofía de la investigación sobre la enseñanza". En M. Wittrock (comp), La investigación en la enseñanza. Enfoques, teorías y métodos I. México: Paidós. (1989)

Use Case: OpenSource Virtualization for Educational Computer Laboratories

Lucas Barboza and Cristian Damián Pacífico

Universidad Nacional de Entre Rios, Facultad de Ciencias de la Administracion, Msñor. Tavella 1424, 3202 Concordia, Entre Ríos, Argentina
{lucas_barboza@outlook.com,cripac@fcad.uner.edu.ar}

Abstract. The usual case with computing laboratories in schools and other public educational institutions is that computers are not adequately set up and maintained in order to provide an homogeneous user experience. They are configured and managed individually and it is almost impossible for managers to keep up with the latest software updates in every single PC. In other cases, some institutions cannot afford having any computers at all. By implementing a central powerful virtual server and a group of lightweight hardware computers in a server-client configuration, it is possible to offer the same user experience one gets with a desktop computer. Consequently, is it possible to reduce the equipment's maintenance effort, simplify software updates and have all computers equally configured. In this work, we define the requirements for an actual educational computer laboratory and present a possible solution to overtake this limitations through the use of a virtualization platform, which provides virtual desktops environments, along with any kind of computational device that can be used as a *thin client*.

Keywords: Computer Laboratories, Virtualization, Thin Clients, SBC, LTSP

1 Introduction

One of the advantages of using virtualization technologies is that they reduce the required infrastructure to run different environments and applications by using the same underlying hardware more efficiently. This makes it possible for organizations to implement consolidation and deploy one or more isolated environments into a single physical machine that hosts a virtual server per scenario.

Implementing virtualization also helps reduce the effort required to maintain and deploy virtual machines, since this is done centrally from the server, which frees the manager from individually configuring each physical device.

This way, a central and more powerful machine can be configured to handle the workload of a group of individually connected workstations that rely on the server for their processing work to be handled. In this schema, all that those computers do is presenting the user with a graphical interaction interface and displaying the output that comes from the server as a result of a performed

task. As consequence, user workstations needn't be resourceful anymore since their processing work is transfered to and handled by the server, what turns them into a sort of dumb terminals.

So the use of *thin* and *fat clients* comes really in handy in a server-client architecture since their main job is to transfer the user input to the server, which performs the requested tasks and returns the graphical results back to the user, all through the LAN. Like this, it is possible to help reduce the cost of implementing a computer laboratory since the price of a *thin client* device is considerably lower than that of an average PC. Also, old machines that cannot handle newer Operating Systems or whose hardware is quite outdated, can be recycled to be used as thin client terminals and become usable again.

We believe this is a feasible solution for those educational institutions which cannot afford having computer laboratories due to its cost and for those who have it but find it difficult to keep it updated and configured according to their users' needs.

In this work, we first characterize computer laboratories at the Facultad de Ciencias de la Administración of the Universidad Nacional de Entre Ríos (hereafter FCAD-UNER), in terms of didactic requirements. Then, we focus on the possible virtualization techniques and the available technologies that serve as potential solutions for the issues encountered at the computer laboratories. We did not describe the general concepts related to virtualization in depth, instead we included some clarifications of the most relevant terms. (For more detailed information, we suggest reading [6, 7, 13, 18, 19]). Next, we present a practical use case where the found solution was implemented at the FCAD-UNER laboratories. Finally, we conclude this work by presenting advantages of deploying such a solution in an educational environment and some future lines of work.

2 Computer Laboratories at FCAD-UNER

Since Computer Laboratories' PCs present their own uniqueness, their analysis and better comprehension allows for a better understanding prior to decide whether to implement virtualization solutions or not.

Next, we present a brief description of the computer laboratories available at the FCAD-UNER, their given use and the major issues pointed out by main users and administrators. This is based both on physical inspections performed on the laboratories and some ad-hoc interviews made to teachers, students and administrators from that same institution.

1. For most classes that make use of computer laboratories, prepared activities rely on the computers all being configured in the same way and having identical or at least similar software running on them. This is quite often an issue because computers are not the same in terms of purchasing methods, manufacturer, hardware and/or installed software, and their configurations differ between laboratories. In consequence, the user experience students get is not equal and depends on the computer they get.

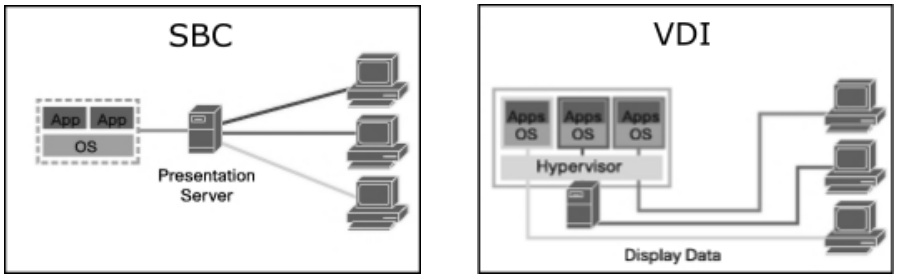
2. Some user groups make a heavy use of the computational power of the PCs, mostly related to Computer Science or Data Analysis practices, while other groups use them for office related work and web surfing. Hence, there are situations in which computers are more stressed and require better performance while in others their processing capabilities aren't used at its maximum capacity. This leads to having different laboratories where computers' hardware and software is set according to their expected use. The result of this decision is that the total computational power is distributed into many individual machines, making it impossible to make the most out of it, as it would be if it were consolidated.
3. One of the most frequent issues for administrators are configuration changes made by students. Despite imposing restrictions, passwords and privileged permissions to standard user accounts, these control measures do not appropriately prevent some situations from arising. For example, there's the usual case where, purposely or not, users infect computers via USB sticks that contain different kinds of malware.

In short, it is important for both teachers and students to have identical or equal environment where to develop their computer laboratory practices. Also, given that not all users utilize computers for the same purpose, their software and hardware varies and most of the time these resources are not efficiently employed. Finally, keeping computers up-to-date and free of unwanted configurations demands administrators constant attention, specially where individual physical configurations are required.

3 Desktop Virtualization

The main purpose of **Desktop Virtualization** is to migrate the workstation's local processing workload to a central server, turning the workstation into a graphical user interface from where to interact with and benefit from the server's computing power. By doing so, the workstation is exempted from managing both hardware and software resources -excluding some minimum required components-, because it relies on the server for all its computational power. This virtualization scenario is possible through **VDI** (Virtual Desktop Infrastructure) or **SBC** (Server Based Computing). While VDI provides a dedicated Virtual Machine (VM) for each individual user [1], SBC consists of a central virtual machine that hosts an Operating System that accepts multiple user connections at the same time [21], giving each user a separate virtual desktop environment. Figure 1 shows the differences between both virtualization options.

Although VDI permits many customized VMs according to every user needs, in terms of management it is still necessary to supervise and configure one environment per user. With SBC, on the other hand, the server hosts one or more VMs, which in turn host a virtualized Operating System that delivers virtual desktops to a range of connected clients. Therefore, it is possible to economize the number of VMs to deploy and manage when using SBC in contrast to VDI,



(a) SBC scheme. (b) VDI scheme.

Fig. 1: SBC and VDI different schemes.

since we only need a certain group of running VMs to provide each user with a desktop, instead of requiring dedicated VMs as is the case with VDI.

When considering the FCAD-UNER computer laboratories, we believe SBC is preferable to VDI as a possible virtualization strategy. SBC works with a single terminal server that delivers a desktop to each connected user (a private user space with its own `/home` directory); instead of having to deploy a dedicated VM per user, as it is the case with VDI. Consequently, there are fewer environments when working with SBC than with VDI, making deployment and management tasks simpler to conduct.

3.1 Thin and Fat Clients

As a result of migrating the computational power from PCs to the server, the user workstation's hardware can be drastically reduced. Or, if it is already resource-limited, it is possible to re-purpose it.

Thin clients [2, 22] are computers with minimum hardware resources that are just good enough to perform basic tasks, such as establishing a connection with a remote server; getting input from the user and sending it to the server for it to process it (getting then the results); storing limited amount of data; or displaying graphics to an output.

Fat clients [8] are similar to thin clients in that they also count on the server to perform heavier processing workload, but they are better equipped in terms of hardware, thus enabling them to handle some tasks locally; turning to the server only for the most demanding processes. While it is rare for a thin client to have more than a 1 GB of RAM or any storing devices at all, this is not the case with fat clients, which hold more hardware resources in comparison.

These devices are well-suited as client terminals in a server-client virtual desktop architecture [8], whether using VDI or SBC, since the client workstation no longer requires to have as many resources as when locally performing the processing workload.

3.2 Terminal Server Solution

As stated before, we chose SBC as a potential virtualization solution reference for the FCAD-UNER computer laboratories. Next, for the terminal server, we determined that it should meet the following requirements:

- That it is possible for the client machines to boot the terminal server session from the network.
- That the server handles the client machine's workload.
- That is not necessary to have any operating system locally installed on the client machine for it to work.
- That the terminal server solution is Open Source in order to avoid license fees.
- That the terminal server solution is compatible with Open Source virtualization platforms, such as XenServer.

After some exploratory research was done to the available technologies, we found out that the Linux Terminal Server Project (LTSP) met the requisites and so we decided to use it as the terminal server solution. LTSP is a collection of open source software that turns a normal GNU/Linux installation into a terminal server, thus enabling the use of thin and fat clients as user terminals. LTSP requires that the client device has a network card that is able to perform network booting (netboot) and at least 300 MB of RAM, for it to properly work [5]. Every connected user gets its own private space, known as chroot jail, that resides in the `/opt/ltsp` directory of the LTSP server. In order to provide a virtual desktop suitable for all devices, it is possible to configure a dedicated directory for the i386, amd64 or armhf architectures -the latter being necessary when working with Raspberry Pis as client machines [17].

4 Implementation Case

To conduct the implementation case, we used a R&D virtualization platform installed at the FCAD-UNER. The virtualization platform is made up of 2 HP ProLiant DL360 Gen 9 8SFF servers, 128 GB of RAM and 2 SSD of 300 GB capacity each; 1 HP MSA 2024" storage array, totalizing 7200 GB of storage capacity. Although this virtualization platform characteristics may seem too bulky for an average educational institution to replicate in its own computer laboratory, this is not the case since LTSP requirements are indeed very limited [5], making it accessible and affordable. We decided to use the aforementioned platform to conduct the implementation because of the ease of access to it and practical reasons solely. Nevertheless, it is totally possible to reproduce this scenario using any computer that meets LTSP server minimum characteristics.

In order to implement a LTSP server, we used XenServer to set up a VM with Ubuntu 16.04 as its GNU/Linux OS distribution. As LTSP works with Debian, Fedora, Gentoo, openSUSE and Ubuntu, we decided to use the latter because of being more used to working with it. The VM was set up with 8 GB

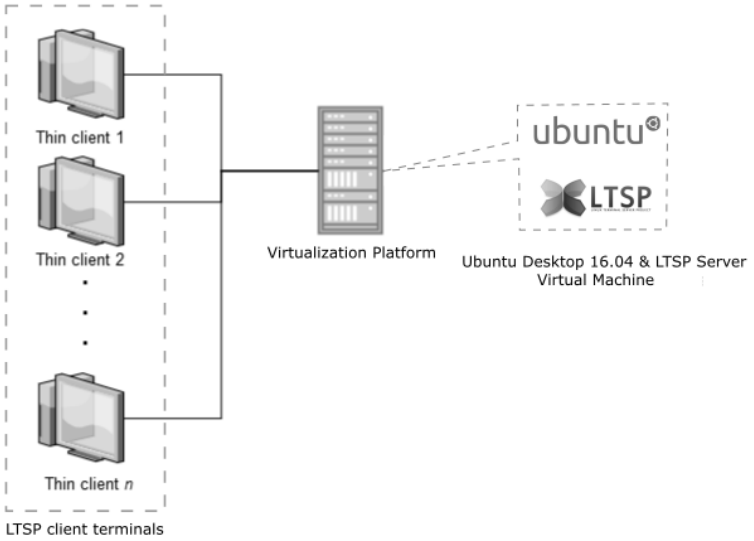


Fig. 2: Diagram of LTSP server-client implementation.



Fig. 3: LTSP working test at one of the FCAD-UNER computer laboratories.

of RAM, 6 CPU cores, 1 NIC with Internet access and a 100 GB of disk space. After installing the LTSP server and making the necessary adjustments and configurations, we connected the VM to the computer laboratory LAN. Figure 2 shows the infrastructure diagram.

We decided to conduct the test using the 16 PCs at one of the laboratories of the FCAD-UNER. The reason for that was to determine if they would all cope with the LTSP server-client architecture and could be used as thin/fat clients in a virtualized computer laboratory scenario, given that they all have different hardware characteristics and configurations. First, in order to boot the thin clients from the LTSP server, we had to enable the networking booting from each of the computer BIOS. Then, after restarting the PCs, they connected to the server and displayed the corresponding login screen, as shown in Fig. 3. We then executed some applications so as to simulate a real use case situation: Firefox Browser, OpenOffice Writer and Calc, and tried out Epoptes software

too (a computer lab management and monitoring tool[1]). We didn't notice any malfunctioning while using it so we concluded it worked correctly.

5 Benefits

We believe that educational institutions could benefit from implementing their computer laboratories with virtualization solutions, like the one we presented here, because of the following reasons:

1. **Cost Reduction**: by using free solutions such the ones proposed here, there are no license fee costs and it possible to execute any number of terminals without worrying about renewing or purchasing new licenses. Moreover, the use of thin and fat clients as user terminals means that the required investment to mount a computer laboratory is less compared to that of installing average PCs. Besides, recycled computers can be used as thin clients, obviating new acquisition costs.
2. **No Proprietary Licenses**: the GNU/Linux distributions are not only free for everyone to use but also they allow and encourage modifications, which means that changes to the software can be done with no restrictions at all.
3. **Simplified Maintenance**: by centralizing and consolidating the whole computer laboratory system into a server, or groups of servers, the management is simplified since the maintaining of the operating system and installed applications is done from a single location. What is more, it is easier to migrate, change, update, install and remove software because doing it once is enough for the whole system to adopt the new configuration. Also, since it is not necessary to have any installed operating system on the client device for it to work, by installing more than one OS in the terminal server gives the user the chance to choose what OS to launch at booting time. Besides, in case a rollback to a previous configuration is needed, using the built-in snapshot tools from XenServer (or any other virtualization server) comes really in handy. Finally, as LTSP clients don't get root access to the terminal server, all user actions are restricted to their given `/home` directory, preventing them from modifying other users desktops and the main system.
4. **Customization**: professors and teachers could freely choose their own customized desktop environment according to their lessons and needs. This way, students of a given class would all have access to the same version of applications, a certain operating system, a special set of programs and/or particular documentation; all this considering that didactic activities present their own requirements.
5. **Machine Independence**: since the workstation no longer stores any data, a user can use any available terminal (even one of his own) to connect to the server and perform their desired tasks. This facilitates the replacement of defective machines and favors the BYOD (bring your own device) movement [15].

[1] http://www.epoptes.org/

6. **Improved Security**: it is possible to implement SSH tunnels between the LTSP server and the connected terminals, and also restricting the communication to a specific LAN. What is more, the use of private chroot jails prevents the propagation of frequently distributed malware, since user sessions are kept isolated from each other and even from the server operating system.
7. **Green Computing** [8]: virtualization contributes to reduce the ecological footprint as it consumes less energy per user while making a better and more efficient use of the underlying hardware. In addition, the use of thin and fat clients demands less electricity compared to PCs, which not only benefits the environment but also reduces the energy expenses. Finally, thin and fat clients have a longer lifespan than traditional PCs, and considering that older computers can be recycled an turned into thin clients, this solution helps reduce E-waste too.

6 Conclusion

After successfully implementing a virtual desktop environment at the FCAD-UNER computer laboratories, based on SBC server-client architecture, using Ubuntu, LTSP and thin/fat clients; we believe it represents a convenient option for those educational institutions that find it difficult to deal with their computers' systems heterogeneity, management and updating; and also for those organizations that cannot afford installing a computer laboratory because of high costs of purchasing new PCs, but could benefit from using thin clients instead.

7 Future lines of work

Perform a comparative between different available OpenSource solutions that provide virtual desktops trough a terminal server (different to LTSP). Analyze the possibility to implement the solution using other devices apart from thin clients (smartphones and tablets). Formally perform and record stress tests in order to quantify the system QoS. Document the required training to implement and maintain such a system, as well as the necessary manuals and procedures to operate it.

References

1. VDI: A New Desktop Strategy, Dec 2006, Daniel Bev eridge, `https://www.vmware.com/pdf/vdi_strategy.pdf`
2. Thin Clients 2011 - Ecological and economical aspects of virtual desktops, Christian Knermann et al., `http://cid448u1296.hd52.hosting.punkt.de/fileadmin/user/upload/documents/PDF_files/White_Paper_EN/thinclients2011-en.pdf`
3. Install LTSP to run a thin client setup, Feb 2016, `http://www.havetheknowhow.com/Configure-the-server/Install-LTSP.html`
4. Concepts - LTSPedia, Sep 2017, `http://wiki.ltsp.org/wiki/Concepts`

5. LTSP Installation & requirements, http://wiki.ltsp.org/wiki/Installation
6. Operating-system-level virtualization - Davoud Teimouri, 2018, https://www.teimouri.net/operating-system-level-virtualization
7. Virtualización de aplicaciones, 2018, https://inlab.fib.upc.edu/es/virtualizacion-de-aplicaciones
8. Shalabh Agarwal, Rana Biswas, and Asoke Nath. Virtual desktop infrastructure in higher education institution: Energy efficiency as an application of green computing. In *Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on*, pages 601–605. IEEE, 2014.
9. Yassine Benkirane. *Clients légers sous Linux: LTSP (French Edition)*. eBook, 2014.
10. Ka Ching Chan and Mary Martin. An integrated virtual and physical network infrastructure for a networking laboratory. In *Computer Science & Education (ICCSE), 2012 7th International Conference on*, pages 1433–1436. IEEE, 2012.
11. Pawel Chrobak. Implementation of virtual desktop infrastructure in academic laboratories. In *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*, pages 1139–1146. IEEE, 2014.
12. Chang-tseh Hsieh. Strategies for successfully implementing a virtualization project: A case with vmware. *Communications of the IIMA*, 8(3):1, 2008.
13. Michael Jorgensen, Michael Fairbanks, Jason Allen Sabin, Nathaniel Brent Kranendonk, and Kal A Larsen. System and method for providing virtual desktop extensions on a client desktop, June 18 2013. US Patent 8,468,455.
14. Cao Le Thanh Man and Makoto Kayashima. Virtual machine placement algorithm for virtualized desktop infrastructure. In *Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE International Conference on*, pages 333–337. IEEE, 2011.
15. Afreen, Rahat Bring your own device (BYOD) in higher education: opportunities and challenges In *International Journal of Emerging Trends & Technology in Computer Science*, volume 3, pages 233–236. 2014.
16. Harry Mangalam. Setting up an LTSP Thin client system, Nov 2010.
17. UbuntuLTSP/RaspberryPi https://help.ubuntu.com/community/UbuntuLTSP/RaspberryPi
18. Daniel A Menascé. Virtualization: Concepts, applications, and performance modeling. In *Int. CMG Conference*, pages 407–414, 2005.
19. Tomislav Petrović and Krešimir Fertalj. Demystifying desktop virtualization. In *Proceedings of the 9th WSEAS international conference on Applied computer science*, pages 241–246. World Scientific and Engineering Academy and Society (WSEAS), 2009.
20. Stephen Soltesz, Herbert Pötzl, Marc E Fiuczynski, Andy Bavier, and Larry Peterson. Container-based operating system virtualization: a scalable, high-performance alternative to hypervisors. In *ACM SIGOPS Operating Systems Review*, volume 41, pages 275–287. ACM, 2007.
21. Volchkov, Andrej Server-based computing opportunities In *IT professional*, volume 4, number 2, pages 18–23. IEEE, 2002.
22. Tolia, Niraj and Andersen, David G and Satyanarayanan, Mahadev, Quantifying interactive user experience on thin clients In *Computer*, number 3, pages 46–52. IEEE, 2006.

Diseño de una plataforma remota para desarrollo de prácticas de laboratorio

Marcelo Zabaljauregui[1,1], Omar Rodríguez[1], Hugo Mazzeo[1] y José Rapallini[1]

[1] Universidad Tecnológica Nacional – Facultad Regional La Plata
1900 La Plata, Buenos Aires, Argentina
Marcelo Zabaljauregui, Hugo Mazzeo, Omar Rodríguez, José Rapallini.
mzabaljauregui@gmail.com; hhmgvm@yahoo.com; rodriguezomarlp@gmail.com; josrap@gmail.com

Abstract. En muchas ocasiones, el contenido de las clases prácticas en asignaturas de carreras como Ingeniería se ve limitado por los recursos disponibles en los laboratorios. Estas limitaciones, que habitualmente tienen que ver con las capacidades de los recursos experimentales -generalmente escasos- pueden ser remediadas en gran medida mediante la utilización de Internet y realidad virtual aplicados al desarrollo de prácticas de laboratorio.
Las aplicaciones existentes en el mercado suelen ser costosas o difíciles de utilizar y en la mayoría de los casos no se adaptan a todas las necesidades requeridas en la enseñanza universitaria.
Surgió entonces la idea de formular un PID para desarrollar un laboratorio de ensayo de aplicaciones de hardware-software sobre el cual pudiera experimentarse con sistemas reales a distancia, sin necesidad de concurrir físicamente al mismo, facilitando el acceso de los estudiantes a las labores relacionadas con la formación práctica con libertad de horarios y tiempos.

Keywords: laboratorio, remoto, virtual, codiseño, microcontrolador, embebido.

1 Introducción.

El objetivo principal del proyecto consistió en brindar a los usuarios la posibilidad de realizar prácticas en entornos controlados, simulando ambientes de trabajo en donde se pueda interactuar en forma remota con los dispositivos o sistemas físicos, químicos, electrónicos, etc.

Para cada caso particular, sólo es necesario modificar el software y hardware propio de la aplicación a automatizar, adaptar sus entradas y salidas y reconfigurar algunos parámetros del sistema para acondicionarlo y vincularlo apropiadamente con el sistema principal.

Se comenzó apuntando a la investigación y desarrollo de proyectos de codiseño hardware/software relacionados con los sistemas de tiempo real, como ser:

- aplicaciones industriales en desarrollos de sistemas de control de supervisión y adquisición de datos con variables de proceso a través de dispositivos sensores y actuadores de distinto tipo;

- aplicaciones de robótica: seguidores de línea, robots resuelve laberintos, brazos robóticos manipuladores de objetos, robots para aplicaciones didácticas, etc.;
- aplicaciones de domótica: control de "edificios inteligentes", sistemas de riego automatizado, sistemas para mejoramiento del confort y ahorro de energía, etc.;
- utilización de sistemas operativos de tiempo real para el diseño en tiempo real estricto.

En disciplinas técnicas y científicas la realización de experimentos con plantas reales es fundamental para consolidar los conceptos adquiridos en las clases teóricas. Sin embargo, debido a diferentes razones, los laboratorios reales no siempre están disponibles, lo cual impone restricciones en el aprendizaje. Afortunadamente, las nuevas tecnologías basadas en Internet pueden ser utilizadas para mejorar la accesibilidad a los experimentos.

Actualmente la educación a distancia se presenta como la solución ideal para un conjunto de trabajos prácticos que exigen disponer de sistemas de enseñanza mucho más flexibles, accesibles y adaptativos (sin limitaciones espaciales ni temporales) que los tradicionales.

Por consiguiente, trasladando este entorno práctico a la enseñanza a distancia, el elemento necesario para abordar la realización de prácticas sobre diversas materias de Ingeniería en Sistemas es la existencia de un sistema de apoyo a la enseñanza consistente en un laboratorio virtual y de telepresencia accesible a través de una red interna o Internet que permita al alumno practicar de una forma lo más similar posible a como si estuviese en las dependencias del laboratorio, dándole la posibilidad de manejar las simulaciones o interactuar con los sistemas reales [1], [2], [3].

Es precisamente de la mano de las TICs desde donde se pueden aportar algunas soluciones para ampliar el acceso a la experimentación en forma de laboratorios remotos y virtuales. Así se conseguirán simultáneamente dos objetivos didácticos: (a) realizar prácticas relacionadas con la asignatura ampliando la disponibilidad de los laboratorios y (b) formar a los alumnos en el uso de las TICs [4].

2 Metodología

A pesar de que nada puede compararse con la interacción real con las plantas físicas (laboratorios reales y presenciales), afortunadamente existen otras opciones que proporcionan a los alumnos la impresión de que están interactuando con plantas reales. Así, normalmente se entiende por laboratorio remoto aquél que existe y puede ser manipulado de forma remota a través de Internet, haciendo uso de webcams, hardware específico para la adquisición local de datos y software para dar una sensación de proximidad con el equipamiento y por laboratorio virtual (local o remoto), o utilizando software informático genérico o específico para recrear el comportamiento de plantas de experimentación que sólo existen en ordenadores usados para la simulación [5].

En los entornos educativos a distancia, resulta esencial disponer de un mecanismo que permita analizar, gestionar y compartir los recursos entre los usuarios. Cuando se trabaja en un entorno colaborativo con experimentos reales ubicados en un laboratorio

teleoperador, resulta obligatorio la implementación de un "gestor de acceso" mediante reservas.

En el presente artículo se presentan los primeros detalles del diseño y utilización del sistema de laboratorio remoto, ya que el proyecto continúa en desarrollo, quedando aún pendiente la etapa correspondiente a la virtualización del sistema. Actualmente se encuentra en etapa de “testeo”, buscándose su puesta a punto y optimización de funcionamiento.

Para la implementación metodológica del proyecto, se consideran las tareas a desarrollar en los ámbitos de trabajo de educación e industria.

En el ámbito educativo, el desarrollo de una plataforma robótica multipropósito, permitirá ensayar y realizar prácticas reales con los alumnos de diversas materias de Ingeniería en Sistemas.

En la industria, en general, la conexión de elementos físicos a una computadora con un software de control adecuado, forma el sistema real en que se basan las prácticas de laboratorio. Para mejorar la interacción con este recurso local se debe diseñar un instrumento virtual que permita ejecutar un sistema simulado. Este instrumento virtual debe implementar un modelo matemático completo que describa adecuadamente el comportamiento del sistema. Los controles de la interfaz virtual, ligados a las variables y ecuaciones expresadas mediante un lenguaje de programación, deben tener el aspecto necesario para interactuar con el usuario de una forma didáctica e intuitiva. Es por ello que la programación de este tipo de aplicaciones suele hacerse usando un software gráfico de alto nivel como LabView, Matlab/Simulink o applets de Java, a modo de sistema de supervisión y control o SCADA.

Comúnmente, el desarrollo de SCADAs y aplicaciones de supervisión de procesos captan el enfoque de aplicaciones de aprendizaje de mayor peso, tales como las desarrolladas en la industria. Por ello, es deseable preparar a los estudiantes en el uso de estas herramientas. Para la realización de las prácticas virtuales, tanto locales como remotas, se dispondría de un aula dotada con ordenadores, alguno de los cuales funcionaría como IGSS (Interactive Graphical SCADA System) permitiendo el control y supervisión de todo el sistema. Se propone la construcción de una plataforma multipropósito que podría, por ejemplo, interactuar con un brazo robótico elemental con tres grados de libertad como mínimo y la posibilidad de accionamiento de una pinza para la manipulación de distintos tipos de objetos.

3 Desarrollo

Uno de los objetivos que se planteó en el proyecto fue que el usuario del laboratorio disponga de una plataforma remota con la cual poner en práctica los conocimientos adquiridos en las materias que cursa sin necesidad de recurrir a la adquisición de material extracurricular [6], [7].

Como se observa en la figura 1, el sistema está compuesto por un conjunto de máquinas conectadas a un servidor que se encarga de gestionar las reservas de los laboratorios y permite a los usuarios acceder al mismo por medio de un navegador web con soporte Java instalado.

De esta manera, distintos usuarios pueden acceder a diferentes aplicaciones remotas en cualquier momento, sólo será necesario seleccionar y reservar la práctica de laboratorio correspondiente mediante el sistema de reservas.

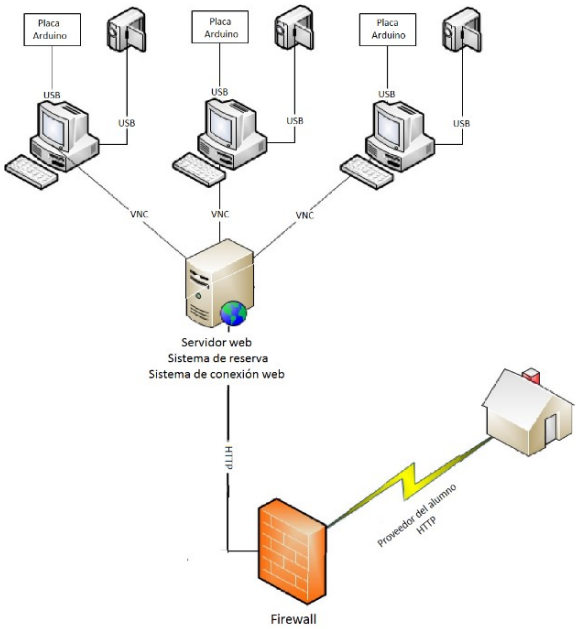


Fig. 1. Diagrama esquemático del laboratorio remoto

En cuanto a las máquinas locales, cada una de ellas dispone de los siguientes elementos:

- Una cámara web y una placa microcontrolada con entradas y salidas analógicas y digitales, ambas conectadas a la computadora local mediante un puerto USB. La placa cuenta con un entorno de desarrollo que permite implementar un sistema embebido con escaso o nulo hardware adicional, facilitando el uso de la electrónica en proyectos multidisciplinares.
- Hardware y software específico para el desarrollo de la práctica de laboratorio, dependiendo de las necesidades de cada caso particular. Además de la placa con microcontrolador adecuada, se requerirán sensores y/o actuadores, interfaces o drivers de potencia para comandar los dispositivos electromecánicos a controlar y una aplicación en software encargada de controlar localmente el sistema físico a experimentar y que a su vez se comunicará con el servidor.
- La aplicación de software principal encargada de automatizar las tareas de la práctica, gestionar las conexiones de datos, turnos y horarios de utilización, etc.

Con respecto al servidor que corre la aplicación de acceso remoto, la utilización de Guacamole demostró que requiere importantes recursos en ciclos de cómputo y memoria RAM. Dado que Guacamole corre en una plataforma Java y permite concentrar muchos accesos de usuarios, hace necesario usar procesadores Intel I5 de segunda generación para cubrir esa demanda.

El proyecto se dividió en tres etapas:

1. Puesta en funcionamiento de las PC locales con las cámaras IP y las placas conectadas y configuradas adecuadamente como también del servidor que permite conectarse a las mismas desde la web. Para tal fin se instalaron y configuraron las siguientes aplicaciones:

   a) Para las máquinas locales:

   - *Sistema Operativo*: Debian Stretch, optimizado para que funcione con la menor cantidad de recursos de hardware, CPU y RAM.
   - *Aplicación propia para el desarrollo de la práctica de laboratorio*: en cada caso habrá que definirla dependiendo del tipo de práctica a realizar, el hardware y software de control, cantidad de sensores y actuadores necesarios, etc.
   - Servidor VNC: necesario para la visualización de la experimentación remota.

   b) Para el servidor de acceso:

   - Sistema Operativo: Debian Stretch
   - HTML: se instaló un servidor Apache.
   - Bootstrap: se usa para darle un aspecto más moderno a los portales de acceso.
   - JQuery: sirve de apoyo al Bootstrap.
   - MySQL: se utiliza como base de datos para los usuarios y laboratorios.
   - PHP: para poder interactuar con MySQL.
   - Tomcat: requerimiento para poder instalar Guacamole.
   - Guacamole: servicio de conexión remota para HTML5.

2. Diseño de un sistema de reservas para gestionar los laboratorios y adecuación del sistema de acceso web para que pueda convivir con el sistema de reservas.
   Ya que el sistema permite seleccionar distintos laboratorios para realizar, el usuario deberá ingresar primero en el laboratorio correspondiente y luego ir a la opción de reserva. Una vez allí podrá seleccionar el día y horario de realización de la práctica, siempre y cuando estén disponibles. Luego de verificar los datos que aparecen en la pantalla e ingresar la dirección de email del usuario (a la cual se le enviará un mensaje con los datos necesarios para ingresar al sistema) y la contraseña, aparecerá la confirmación validando la operación.
   Se utilizaron distintas tecnologías: HTML, PHP, JQuery y MySQL. La conjunción de las mismas permitió diseñar un portal ágil para la gestión de los laboratorios, tanto para los usuarios como para los administradores de la plataforma. Además el uso combinado de Tomcat y Guacamole provee el acceso remoto a través de HTML 5.
3. Apertura de los protocolos necesarios a Internet para que el laboratorio pueda ser accedido desde cualquier lugar por el usuario remoto.

Se cuenta con dos tipos de accesos: uno dedicado a la reserva y activación del laboratorio y otro para el acceso de los administradores a fin de visualizar las reservas hechas y realizar estadísticas del uso del mismo o algún deslogueo forzado en el caso de que surja algún problema o situación imprevista [8], [9].
Existe además un deslogueo automático en caso que después de una reserva se genere otra. Como una posibilidad adicional, está contemplado el caso en el cuál si un usuario está usando el sistema y agota su tiempo -no teniendo una reserva posterior para su uso- se le permita seguir utilizándolo.

El procedimiento para la reserva de laboratorios es el siguiente:

a) Ingresar a la URL asignada para el laboratorio. Tomaremos como ejemplo el caso de una práctica para la materia Aplicaciones en Tiempo Real. En la pantalla de ingreso seleccionamos la solapa “Laboratorios” en la barra de opciones de la parte superior (v. fig. 2).

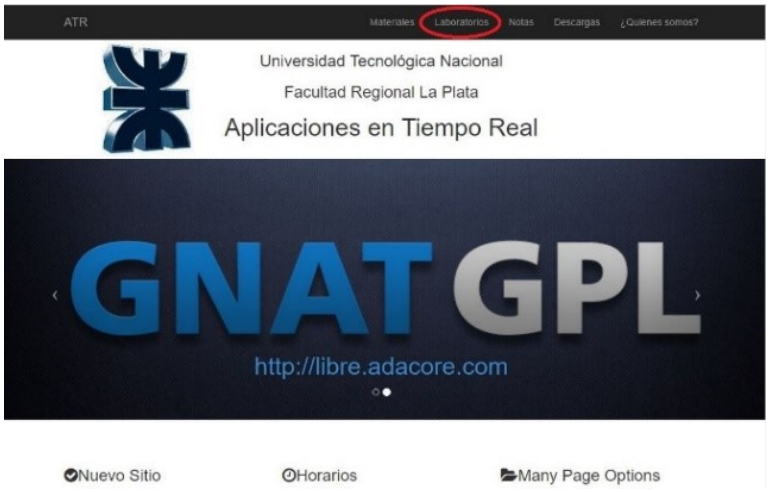


Fig. 2. Pantalla de ingreso al laboratorio remoto

b) Una vez seleccionado el laboratorio a realizar, habrá que ir a la opción “Reserva de laboratorios”, donde se podrá reservar día y horario de realización (v. fig. 3).



Fig. 3. Pantalla de selección y reserva de horario

c) Luego habrá que verificar los datos que aparecen en la pantalla e ingresar una dirección de e-mail válida (que será utilizada para el envío de la confirmación de la reserva, usuario y contraseña para ingresar al sistema -v. fig. 4-).



Fig. 4. Confirmación de reserva recibida por e-mail

Con el usuario y contraseña recibidos, ya será posible acceder al sistema. El procedimiento para el ingreso al laboratorio se realizará de la siguiente manera:

Ingresar al link que fue enviado por email: http://[URL asignada]/guacamole (v. fig. 5).



Fig. 5. Pantalla de ingreso al sistema

Ingresar el nombre de usuario y contraseña recibidos en el e-mail de confirmación de la reserva. Al ingresar se verá el login del sistema operativo remoto (al cual se accede con usr: usuario, psw: usuario -v. fig. 6-):

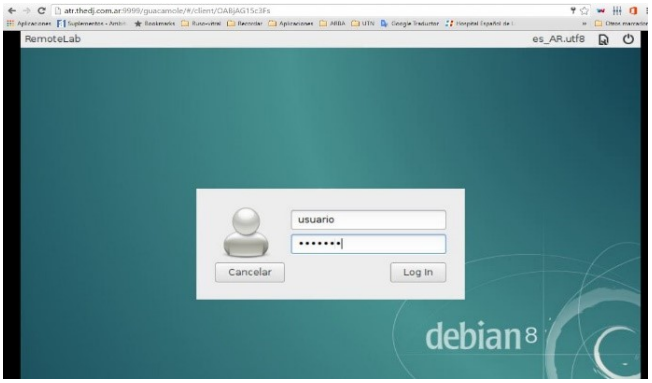


Fig. 6. Pantalla de acceso al laboratorio remoto

Ya se está en condiciones de operar en el laboratorio remoto. Una vez ingresado al entorno de la aplicación, el usuario dispondrá de un menú de aplicaciones que podrá seleccionar. Por ejemplo, dentro del menú Aplicaciones/ATR se encontrarán las distintas aplicaciones para la cátedra correspondiente, en este caso ATR (Aplicaciones en Tiempo Real). La aplicación “Cheese” del menú permite abrir la cámara que apunta al proyecto de laboratorio con el cual se experimentará (v. fig. 7).

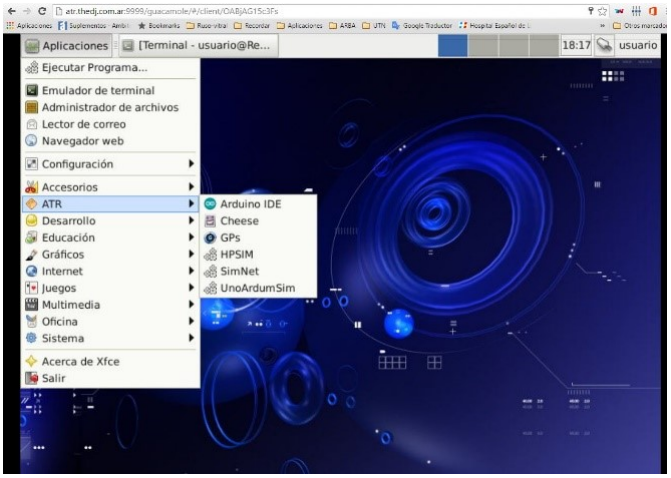


Fig. 7. Aplicaciones disponibles ya dentro del entorno

4 Resultados

La plataforma se puso en funcionamiento recientemente por lo que se puede decir que se encuentra en etapa de prueba. Resta aún continuar con el mejoramiento y actualización del software y la creación de nuevas prácticas de laboratorio que se adapten a la temática de las distintas materias que deseen utilizarlo. Asimismo, falta desarrollar la parte de virtualización de laboratorios, lo que permitiría también mejorar la utilización en situaciones de bajo ancho de banda en las conexiones a Internet.

En las primeras pruebas realizadas se evaluó el sistema en forma local, es decir desde la red interna de la Facultad, logrando resultados satisfactorios. Al probar el acceso remoto desde el exterior, se generaba un delay intolerable para la visualización del video en tiempo real. Para reducirlo, se gestionaron permisos para obtener direcciones IP estáticas para acceder vía Internet utilizando un enlace de mayor velocidad. También se mejoraron las capacidades del servidor de acceso, tanto del hardware como del software. Con esto se logró disminuir el retardo a un valor aceptable. Actualmente se continúa trabajando en la puesta a punto.

Para conocer la opinión y el grado de satisfacción en la utilización del sistema desde el punto de vista del usuario, se prevé realizar una evaluación del funcionamiento del sistema a través de estadísticas, encuestas de opiniones, etc., lo que contribuirá a encontrar fallas o problemas que requieran nuevas soluciones o mejoras en distintos aspectos.

A modo de ejemplo, en la fig. 8 puede verse una captura de pantalla donde se observa una ventana abierta de la aplicación “Cheese” con la imagen de una placa de hardware utilizada para el ensayo y experimentación con circuitos digitales. La aplicación remota permite hacer cambios en el código y luego visualizar los resultados de la experimentación mediante el encendido de leds e indicadores numéricos luminosos.

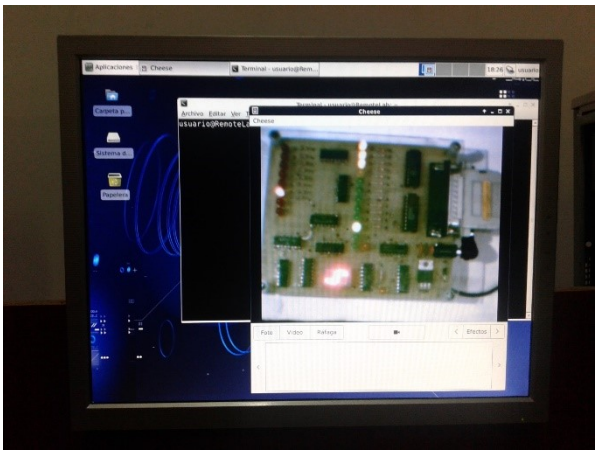


Fig. 8. Ejemplo mostrando una aplicación operada remotamente y visualizada a través de la webcam ubicada en la cámara de la máquina local

5 Conclusiones

En este artículo se ha presentado el diseño de un laboratorio remoto para la realización de prácticas en asignaturas de Ingeniería en Sistemas de Información. El uso de estas plataformas remotas en Internet es recibido con beneplácito tanto por los docentes como por los alumnos, por sus ventajas en lo que respecta a la flexibilidad de horarios de trabajo, manejo de tiempos y posibilidad de repetición de prácticas tanto como se desee.

Otro aspecto que es valorado por muchos estudiantes es la posibilidad de acceder a distintos elementos de hardware, un recurso en general bastante limitado, a los que de otra manera no podría tener acceso todo el alumnado.

Finalmente, como tareas complementarias del proyecto, hay dos objetivos que se persiguen desde el punto de vista del conocimiento y la docencia:

- formación de recursos humanos: desarrollo de un ámbito para la generación de recursos humanos altamente especializados. Estudio y resolución de problemáticas industriales (diseño innovador de sistemas). Transferencia de los conocimientos al sistema académico (alumnos, profesionales, cátedras y grupos de trabajo);
- transferencia de conocimientos a los alumnos que requieran una formación adicional a la recibida en las cátedras de la especialidad. En el plantel básico del grupo se incorporarán graduados noveles y estudiantes del último año. La propuesta a mediano y largo plazo es el dictado de seminarios y cursos de grado y de posgrado para aumentar la capacitación en estos temas. Incluso es intención tender a la implementación de alguna materia electiva y de un magister en la especialidad. Toda propuesta desarrollada e implementada se tratará de transferir al medio productivo.

Referencias

1. Computers & Education. Revista de Elsevier dedicada principalmente a aspectos pedagógicos del uso del computador y las comunicaciones en la enseñanza, incluyendo la enseñanza a distancia.
2. Dormido, S.; Sánchez, J.; Morilla, F. "Laboratorios virtuales y remotos para la práctica a distancia de la automática", Dpto. de Informática y Automática, UNED, Avda. Senda del Rey s/n, 28040 Madrid.
3. Dormido, S. (2004), "Control learning: Present and future", Annual Reviews in Control, vol. 28, pp. 115-136.
4. F. Torres, F. Candelas, S. Puente, F. Ortiz, J. Pomares, P. Gil, M. Baquero, A. Belmonte, "Laboratorios virtuales remotos para el aprendizaje práctico de asignaturas de ingeniería", Departamento de Física, Ingeniería de Sistemas y Teoría de la Señal. Universidad de Alicante.
5. Domínguez, M., Fuertes, J.J., Reguera P. (2009), "Desarrollo de un laboratorio remoto para la formación vía Internet en Automática".
6. Isidro Calvo, Ekaitz Zulueta, Unai Gangoiti, José Manuel López, "Laboratorios remotos y virtuales en enseñanzas técnicas y científicas", Dpto. Ingeniería de Sistemas y Automática Escuela Universitaria de Ingeniería de Vitoria-Gasteiz UPV/EHU.
7. IEEE Transactions on Education. Métodos y tecnologías de educación, materiales para la enseñanza, y programas de desarrollo educacionales y profesionales sobre disciplinas de ingeniería eléctrica.
8. Laboratorio de Control Automático de la Escuela de Politécnica Federal de Lausana: http://emersion.epfl.ch.
9. Laboratorio Remoto de Automática de la Universidad de León: http://lra.unileon.es

Descubrimiento de las áreas de investigación seleccionadas por los tesistas de las carreras de informática de la UM mediante árboles de decisión

Iris Sattolo[1], Gaston Alvarez[1], Matias Garcia[1], Javier Lafont[1,] Lucila Mira[1] , Gabriel Mariuz[1],Nicolás Armilla[1], Marisa Panizzi[1]
[1] Universidad de Morón. Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales. Instituto de Investigación, Desarrollo e Innovación de Sistemas de Información.
Cabildo 134. Morón. Argentina
iris.sattolo@gmail.com; gaston_alvarez19@hotmail.com; matias@clustersistemas.com; lafontjavier@hotmail.com; gmariuz91@gmail.com; lucilamira@gmail.com; nicolasarmilla@hotmail.com; marisapanizzi@outlook.com

Resumen. Las cátedras de las materias de tesis, en las carreras de informática de la Universidad de Morón, cuentan con diferentes herramientas, que brindan apoyo académico a los que comienzan el cursado. Se ha observado que los alumnos tienen dificultad para seleccionar el tema de tesis y, en algunos casos, motiva el abandono o retraso en su carrera. Contar con un registro de tesis, de los graduados, resultó ser útil para obtener patrones sobre la relación entre: el área de investigación abordada en dicha tesis y características que definen al tesista. Para nuestro trabajo se utilizó "el proceso de descubrimiento de conocimiento en base de datos". Finalmente, se presentan los resultados de la utilización de los algoritmos J4.8 en WEKA, ID3 en RapidMiner y CART en Knime y los distintos modelos generados, con los datos recopilados.

Palabras clave: árboles de decisión, WEKA, RapidMiner, Knime, áreas de investigación de tesistas, Carreras de informática UM.

1 Introducción

Las carreras de grado de Informática en la Universidad de Morón (en adelante UM), se encuentran organizadas de manera diferente en sus asignaturas de tesis. La carrera Ingeniería en Informática consta de una asignatura anual, denominada "Tesis de Grado – Proyecto Final Integrador" [1]. La carrera Licenciatura en Sistemas posee dos asignaturas cuatrimestrales, la primera de ellas denominada "Trabajo de Diploma" y la segunda, "Tesis de Grado" [2]. Las cátedras de tesis, de ambas carreras, no cuentan con un instrumento formal que permita a los docentes identificar los saberes que poseen los alumnos, su experiencia laboral, sus tiempos de dedicación a la academia, sus características personales, las líneas de investigación, desarrollo e innovación preferidas, tiempos de permanencia para el desarrollo de sus trabajos de tesis, entre otros. En la actualidad, las cátedras de tesis poseen una planilla de cálculo que contiene la información sobre datos del alumno (apellido, nombre, matrícula y carrera, fecha de defensa de la tesis, docente-tutor o docente-director del trabajo,

título del trabajo, línea de investigación en la cual se inserta el trabajo, resumen, objetivos y futuros trabajos). Este archivo permitió a las cátedras llevar un registro sistemático de un total de 290 tesis, desde el año 2005 hasta de marzo del 2018 [3].

Durante el transcurso de los últimos años, los docentes de las cátedras de las dos carreras observaron que el mayor inconveniente que manifiesta el alumno al comenzar la asignatura es la definición del tema ocasionando un retraso en la finalización de sus estudios y en algunos casos el abandono de la carrera.

En este contexto es que se plantearon las siguientes preguntas de investigación, las cuales impulsaron este trabajo. ¿Son suficientes los datos recolectados de los graduados para determinar patrones de elección del tema de tesis? ¿Se puede correlacionar el área de trabajo de tesis seleccionada, con características tales como área de trabajo, tiempo que disponen para el desarrollo de la tesis, edad al comenzar la tesis, entre otros?

La respuesta a la primera pregunta dio origen a la construcción de un instrumento de recolección de datos denominado TESISTAS-UM y así se completó la información de la cátedra. Se probó su calidad y, por último, se logró una muestra de 114 graduados, compuesta por 46 de la carrera Licenciatura en Sistemas y 68 de la carrera Ingeniería en Informática [4].

Para la segunda pregunta, se decidió utilizar el proceso de "descubrimiento de conocimiento en bases de datos" (en inglés, Knowledge Discovery in Databases, KDD) considerado como "el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos" [5]. Este proceso comprende diversas etapas, reflejadas en la Figura 1, que van desde la obtención de los datos hasta la aplicación del conocimiento adquirido en la toma de decisiones. Entre esas etapas, se encuentra la que puede considerarse como el núcleo del proceso KDD y que se denomina Minería de Datos (en inglés, Data Mining) [6].



Fig. 1. Fases del Proceso KDD [6]

Los problemas tratados en este trabajo se enmarcan en el esquema general del KDD aplicando tareas de clasificación, para obtener un modelo descriptivo, incluidas en la fase de Minería de Datos. En la misma se aplican métodos de aprendizaje para la obtención de modelos y patrones. El aprendizaje siempre será entendido como supervisado, donde los casos pertenecientes al conjunto de datos tienen a priori asignada una clase o categoría, siendo el objetivo encontrar patrones o tendencias de los casos pertenecientes a una misma clase.

La siguiente sección presenta el desarrollo del trabajo, en la cual se describe la metodología empleada y se enuncian los materiales utilizados. Posteriormente, se

presentan los resultados obtenidos de las pruebas realizadas con los algoritmos de árbol de decisión utilizados en diferentes herramientas de trabajo. Se comparten las conclusiones obtenidas y se enuncian futuras líneas de trabajo.

2 Desarrollo

Para el descubrimiento de la información se aplicaron las fases propuestas en Hernández Orallo et al. [6], las cuales se desarrollan en las secciones siguientes.

2.1 Fase de Integración y recopilación

Para la realización de esta fase, se utilizó la planilla de cálculo que la cátedra posee, la cual permitió un registro sistemático de las tesis defendidas y el instrumento de recolección de datos construido [4]. En este estadio del trabajo, se contó con una muestra compuesta por 46 de graduados de la carrera Licenciatura en Sistemas y 68 graduados de la carrera Ingeniería en Informática, totalizando una muestra de 114.

El grupo de investigación reconoce que la muestra no es la aconsejada para el proceso de minería de datos, pero es el resultado de la recolección realizada en un Universidad de gestión privada.

2.2 Fase de selección, limpieza y transformación

Esta fase, conlleva el problema de identificar el conjunto representativo de características adecuadas para construir el modelo. Para la definición de los atributos relevantes a utilizar en la construcción del modelo, se decidió utilizar los resultados de la evaluación realizada en nuestro trabajo anterior y expuesto en TE&ET 2018. Los atributos seleccionados fueron: Carrera, Área de trabajo, Grupo familiar y Edad [4]. Para el atributo clase, se utilizó: área de tesis.

En una primera instancia, para la definición de las áreas de trabajo de las tesis, se realizó un catálogo basado en las áreas de investigación propuestas en el Congreso Argentino de Ciencias de la Computación (CACIC), el cual es organizado por la Red de Universidades con Carreras en Informática (RedUNCI) [7].

Del análisis de datos recolectados, se observó que en el área de Computación Gráfica, Imágenes y Visualización había solamente dos tesis. Dada la escasa elección del área, estas tesis se reagruparon en el área de Innovación en Sistemas de Software por las temáticas abordadas en los mismos. Los temas de tesis que pertenecían a las áreas Innovación en Educación en Informática y Tecnología Informática aplicada a la Educación se reagruparon en un área a la que se denominó Tecnología y Educación por tener las mismas incumbencias propuestas para el Congreso Nacional de Tecnología en Educación y Educación en Tecnología (TE&ET), el cual es organizado por la Red de Universidades con Carreras en Informática (RedUNCI) [7].

El atributo edad originalmente contaba con cuatro rangos (Menor de 25, de 25 a 30, de 31 a 35, Mayor de 35), se reagrupó considerando los siguientes rangos

(Menores de 25 años, Entre 25 y 30 años, Mayores a 30 años) para lograr una distribución más balanceada.

En la recolección de datos, el atributo grupo familiar se definió con cuatro valores (vive-con-padres, solo, hijos y pareja), del análisis de este atributo se logró una reagrupación en dos valores (Con compromiso, Sin compromiso). Esta decisión se fundamenta en la dedicación que el tesista cuenta para con la tesis, el tesista que vive solo (sin hijos) o con sus padres tiene más tiempo respecto a los que son casados o viven en pareja y/o tienen hijos.

La tabla 1, presenta los atributos a utilizar en la fase de DM con sus valores asociados luego de la transformación y limpieza de los datos.

Tabla 1. Atributos y valores para ser utilizados en la fase de Minería de Datos

Atributos	Valores
Área de tesis	Agentes y sistemas inteligentes/ Ingeniería de Software/ Base de Datos y Minería de datos/ Innovación en Sistemas de Software/ Arquitectura, Redes y Sistemas Operativos/ Seguridad Informática/ Tecnología y Educación/ Procesamiento de señales y sistemas en tiempo real
Carrera	Licenciatura en Sistemas/ Ingeniería en Informática
Área de trabajo	Análisis funcional y requerimientos/ Bases de datos y minería de datos/ Desarrollo/Infraestructura/ Procesos de negocio/ Seguridad informática/Testing/Varios/ No trabaja
Edad	Menores de 25 años/ Entre 25 y 30 años/ Mayores a 30 años
Grupo familiar	Con compromiso/ Sin compromiso

2.3 Fase de minería de datos

En este trabajo, se construyó un modelo preliminar de carácter descriptivo, en el cual el objetivo no es predecir nuevos datos, sino describir los existentes [6]. Este modelo debe permitir identificar las áreas de investigación de las tesis seleccionadas por los tesistas y su relación con otros atributos que definen al mismo. Para la construcción del modelo, se trabajó con la tarea descriptiva, reglas de asociación [8] dado que se buscó el hallazgo de reglas de asociación entre los atributos (carrera, edad, grupo familiar, área de trabajo) y el atributo objetivo (área de tesis). Se experimentó con los algoritmos de árboles de decisión: J4.8 de WEKA [9], ID3 en RapidMiner [10] y CART en Knime [11].

2.3.1 Análisis de los algoritmos de árboles de decisión

Los árboles de decisión son una de las formas más sencillas de representación del conocimiento adquirido. Dentro de los sistemas basados en árboles de decisión, habitualmente denominados TDIDT (Top Down Induction of Decision Trees) [12], se pueden destacar dos familias o grupos: la familia ID3, el más representativo es el propio algoritmo ID3 propuesto por Quinlan [12] y la familia de árboles de regresión,

cuyo exponente más significativo es Cart, desarrollado por Breiman et al. [13]. Los TDIDT se caracterizan por utilizar una estrategia de divide y vencerás descendente, es decir, partiendo de los descriptores hacia los ejemplos, dividen el conjunto de datos en subconjuntos siguiendo un determinado criterio de división. A medida que el algoritmo avanza, el árbol crece y los subconjuntos de ejemplos son menos numerosos. [12].

Los algoritmos generan reglas de decisión que son presentadas como un árbol, donde la población total (nodo raíz) es sucesivamente dividida (ramas-nodos intermedios) hasta obtener segmentos de similar comportamiento (nodos hojas) en relación con la variable objetivo. Las hojas contienen la predicción. En cada división se selecciona al predictor que mejor separa a la población con respecto a la variable objetivo. Si la variable objetivo es categórica, se llaman árboles de clasificación, en cambio, si es continua se llaman árboles de regresión [14].

Los diversos algoritmos se diferencian por razones como: naturaleza de los datos a clasificar, número de ramas que pueden dividir, criterios utilizados para la división, administración de los valores faltantes y métodos de poda (simplificación del árbol).

CART genera solo árboles binarios, es decir de cada nodo se desprende exactamente dos ramas. Mientras que J4.8 e ID3 pueden generar más de dos ramas. Si el predictor es de tipo nominal, CART agrupa las categorías en dos y genera solo dos ramas. J4.8 e ID3 por defecto generan una rama por cada categoría, pero además presentan la opción de agrupar categorías para generar menos ramas. Si el predictor es de tipo continuo, CART, J4.8 e ID3 buscan un valor de división y generan solo dos ramas. Si el predictor es ordinal, se pueden ingresar a los algoritmos CART, J4.8 e ID3 como si fuera continuo.

Para definir los criterios de división, se utilizan medidas de pureza para seleccionar los atributos que mejor dividen a las instancias. J4.8 e ID3 utilizan la medida de Razón de Ganancia por defecto (basado en entropía). En WEKA solo se dispone de esta medida, mientras que RapidMiner provee de Information-gain, Gain-ratio, Gini Index, Accuracy. CART utiliza el Índice de Gini por defecto.

Los Métodos de poda utilizados en los algoritmos son los siguientes; J4.8 e ID3 cuenta por defecto con el método de Error Pesimista y dos etapas de poda: local y global. WEKA adicionalmente ofrece el método de Error Reducido. En WEKA se cuenta con la opción de ejecutar ambas etapas, mientras que en RapidMiner solo presenta la etapa local. CART tiene por defecto el método de Costo-Complejidad.

En la tabla 2, se presenta una síntesis de la comparativa de los algoritmos empleados para la construcción del modelo.

Tabla 2. Comparativa de características de los algoritmos de árboles de decisión utilizados

Algoritmo	Variables predictorias	Tipo de división	Criterio de división	Método de poda	Implementación
J4.8	Continuas/Discretas	Binaria/ n-aria	Gain ratio (Entropía)	Pre- /Post-	Libre (Weka)
ID3	Discretas	n-aria	Ganancia de información (Entropía)	No	Comercial

CART	Continuas/Discretas	Binaria	Impureza (índice de Gini)	Post-	Libre Comercial

2.3.2 Construcción del modelo preliminar.

Para la construcción del modelo se utilizaron los algoritmos de árboles de decisión, J4.8 (de WEKA), ID3 (en RapidMiner) y CART (en Knime). Se utilizaron los atributos descriptos en la Tabla 1. Se seleccionó como atributo meta o clase, el área de tesis y se trabajó sin poda.

De la construcción del primer modelo en las tres herramientas, se observaron los siguientes tamaños de árboles: CART (59), J4.8 (45), ID3 (54), además de la compleja visualización e interpretación para la solución a nuestro problema planteado. Los tres algoritmos utilizan como nodo raíz, el atributo Área de trabajo. El atributo carrera en J48 e ID3 es el segundo nodo elegido, mientras que en CART es utilizado en las terceras o cuartas líneas de división.

Luego de estas comprobaciones y en correspondencia con la decisión ya tomada de investigar el problema a través de métodos de clasificación, sumado al interés de las cátedras de poder diferenciar entre ambas carreras, si existen diferencias entre las elecciones de los tesistas se dividió el conjunto de datos en dos. El primero para la carrera Licenciatura en Sistemas y el segundo para la carrera Ingeniería en Informática. La Figura 2, presenta el modelo logrado para la carrera Licenciatura, con el algoritmo J4.8 en WEKA. La Figura 3, presenta el modelo logrado para la carrera Licenciatura en Sistemas, con el algoritmo ID3 utilizando la herramienta RapidMiner La Figura 4, presenta el modelo logrado para la carrera Licenciatura en Sistemas, con el algoritmo CART utilizando la herramienta Knime (los tres modelos sin poda).

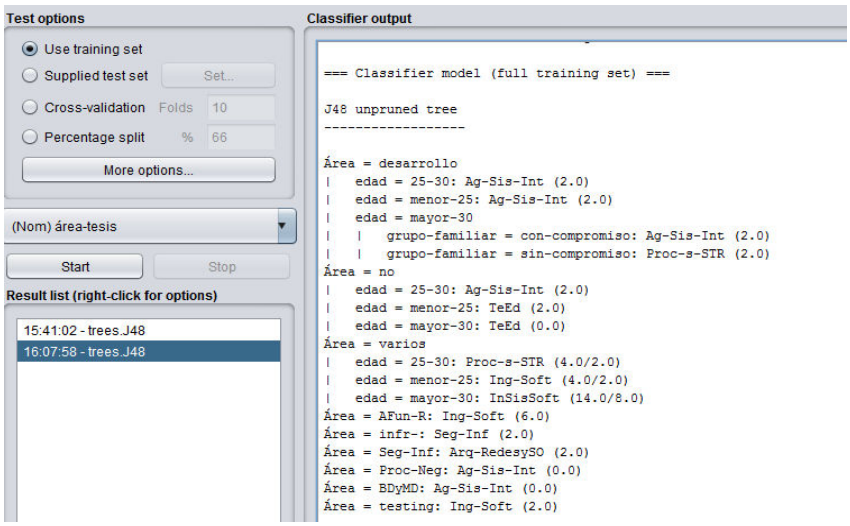


Fig. 2. Modelo para la carrera Licenciatura en Sistemas utilizando el algoritmo J4.8 en la herramienta WEKA.

```
Área = Anal-func-Requ: Ing-Soft {Ing-Soft=6, Ag-Sis-Int=0, In-Sis-Soft=0, Seg-Inf=0, Tec-Inf-Aeducacion=0, Arq-
RedesySO=0, Proc-Señales-STR=0}
Área = Seg-Inf: Arq-RedesySO {Ing-Soft=0, Ag-Sis-Int=0, In-Sis-Soft=0, Seg-Inf=0, Tec-Inf-Aeducacion=0, Arq-
RedesySO=2, Proc-Señales-STR=0}
Área = desarrollo
|   edad = 25-30: Ag-Sis-Int {Ing-Soft=0, Ag-Sis-Int=2, In-Sis-Soft=0, Seg-Inf=0, Tec-Inf-Aeducacion=0, Arq-
RedesySO=0, Proc-Señales-STR=0}
|   edad = mayor30
|   |   Grupo-familiar = con-compromiso: Ag-Sis-Int {Ing-Soft=0, Ag-Sis-Int=2, In-Sis-Soft=0, Seg-Inf=0, Tec-Inf-
Aeducacion=0, Arq-RedesySO=0, Proc-Señales-STR=0}
|   |   Grupo-familiar = sin-compromiso: Proc-Señales-STR {Ing-Soft=0, Ag-Sis-Int=0, In-Sis-Soft=0, Seg-Inf=0, Tec-
Inf-Aeducacion=0, Arq-RedesySO=0, Proc-Señales-STR=2}
|   edad = menor-25: Ag-Sis-Int {Ing-Soft=0, Ag-Sis-Int=2, In-Sis-Soft=0, Seg-Inf=0, Tec-Inf-Aeducacion=0, Arq-
RedesySO=0, Proc-Señales-STR=0}
Área = infraestructura: Seg-Inf {Ing-Soft=0, Ag-Sis-Int=0, In-Sis-Soft=0, Seg-Inf=2, Tec-Inf-Aeducacion=0, Arq-
RedesySO=0, Proc-Señales-STR=0}
Área = no
|   edad = 25-30: Ag-Sis-Int {Ing-Soft=0, Ag-Sis-Int=2, In-Sis-Soft=0, Seg-Inf=0, Tec-Inf-Aeducacion=0, Arq-
RedesySO=0, Proc-Señales-STR=0}
|   edad = menor-25: Tec-Inf-Aeducacion {Ing-Soft=0, Ag-Sis-Int=0, In-Sis-Soft=0, Seg-Inf=0, Tec-Inf-Aeducacion=2,
Arq-RedesySO=0, Proc-Señales-STR=0}
Área = testing: Ing-Soft {Ing-Soft=2, Ag-Sis-Int=0, In-Sis-Soft=0, Seg-Inf=0, Tec-Inf-Aeducacion=0, Arq-RedesySO=0,
Proc-Señales-STR=0}
Área = varios
|   edad = 25-30
|   |   Grupo-familiar = con-compromiso: Proc-Señales-STR {Ing-Soft=0, Ag-Sis-Int=2, In-Sis-Soft=0, Seg-Inf=0, Tec-
Inf-Aeducacion=0, Arq-RedesySO=0, Proc-Señales-STR=2}
|   edad = mayor30
|   |   Grupo-familiar = con-compromiso: In-Sis-Soft {Ing-Soft=2, Ag-Sis-Int=2, In-Sis-Soft=4, Seg-Inf=2, Tec-Inf-
Aeducacion=0, Arq-RedesySO=0, Proc-Señales-STR=0}
|   |   Grupo-familiar = sin-compromiso: In-Sis-Soft {Ing-Soft=0, Ag-Sis-Int=2, In-Sis-Soft=2, Seg-Inf=0, Tec-Inf-
Aeducacion=0, Arq-RedesySO=0, Proc-Señales-STR=0}
|   edad = menor-25
|   |   Grupo-familiar = sin-compromiso: Ing-Soft {Ing-Soft=2, Ag-Sis-Int=2, In-Sis-Soft=0, Seg-Inf=0, Tec-Inf-
Aeducacion=0, Arq-RedesySO=0, Proc-Señales-STR=0}
```

Fig. 3. Modelo para la carrera Licenciatura en Sistemas utilizando el algoritmo ID3 utilizando RapidMiner

```
CART Decision Tree

Area=(varios)|(no)|(infraestructura)|(desarrollo)|(Seg-Inf)
| Area=(varios)|(no)|(desarrollo)|(Anal-func-Requ)|(testing)
| | edad=(mayor30)
| | | Area=(varios)|(Anal-func-Requ)|(Seg-Inf)|(infraestructura)|(no)|(testing)
| | | | Grupo-familiar=(sin-compromiso): Ag-Sis-Int(2.0/2.0)
| | | | Grupo-familiar!=(sin-compromiso): In-Sis-Soft(4.0/6.0)
| | | Area!=(varios)|(Anal-func-Requ)|(Seg-Inf)|(infraestructura)|(no)|(testing)
| | | | Grupo-familiar=(sin-compromiso): Proc-Seales-STR(2.0/0.0)
| | | | Grupo-familiar!=(sin-compromiso): Ag-Sis-Int(2.0/0.0)
| | edad!=(mayor30)
| | | edad=(menor-25)|(mayor30)
| | | | Area=(no): Tec-Inf-Aeducacion(2.0/0.0)
| | | | Area!=(no)
| | | | | Area=(varios)|(Anal-func-Requ)|(Seg-Inf)|(infraestructura)|(no)|(testing): Ag-Sis-Int(2.0/2.0)
| | | | | Area!=(varios)|(Anal-func-Requ)|(Seg-Inf)|(infraestructura)|(no)|(testing): Ag-Sis-Int(2.0/0.0)
| | | edad!=(menor-25)|(mayor30)
| | | | Grupo-familiar=(sin-compromiso): Ag-Sis-Int(4.0/0.0)
| | | | Grupo-familiar!=(sin-compromiso): Ag-Sis-Int(2.0/2.0)
| Area!=(varios)|(no)|(desarrollo)|(Anal-func-Requ)|(testing)
| | Grupo-familiar=(sin-compromiso): Seg-Inf(2.0/0.0)
| | Grupo-familiar!=(sin-compromiso): Arq-RedesySO(2.0/0.0)
Area!=(varios)|(no)|(infraestructura)|(desarrollo)|(Seg-Inf): Ing-Soft(8.0/0.0)

Number of Leaf Nodes: 12

Size of the Tree: 23
```

Fig. 4. Modelo para la carrera Licenciatura en Sistemas utilizando el algoritmo CART utilizando Knime

Si bien el algoritmo J4.8 en la herramienta WEKA es más claro de interpretar, no clasifica correctamente todas las instancias. Por ejemplo: Área=Varios→ edad=25-30→: Proc-s-Str (4.0/2.0) clasifica 4 instancias de las cuales 2 son incorrectas.

ID3 en RapidMiner en Área= Varios → edad= 25-30 → de las 4 instancias distribuye 2 instancias en Procesamiento de Señales y las otras 2, en Agentes y Sistemas Inteligentes.

En CART la regla es más compleja ya que realiza divisiones binarias y para el caso de atributos con varios valores dificulta su lectura. Sin embargo, entre la edad 25-30 en los que trabajan en varias áreas, solo clasifica las instancias de Agentes y Sistemas Inteligentes, dejando sin clasificar las instancias de Procesamiento de Señales. Se concluye que este algoritmo no es recomendable para atributos con varios valores.

2.4 Fase de evaluación e interpretación

En la Tabla 3, se presenta un análisis comparativo del porcentaje que representan las áreas de trabajo de mayor incidencia en la selección del área de tesis.

Tabla 3. Comparativa de la incidencia del atributo área de trabajo en la selección del área de tesis (ambas carreras).

	Total, alumnos	**Ingeniería**	**Licenciatura**
Área Tesis	114	68	46
Innovación en Sistemas de Software	16%	18%	13%
Tecnología y Educación	14%	20%	
Agentes y Sistemas Inteligentes	40,50%	44%	35%
Ingeniería de Software	12%		26%
Total de las 3 áreas	83%	82%	74%
Área de Trabajo			
Desarrollo	17,50%	12%	17%
Varios	47%	47%	48%
Total de las 2 área de trabajo	64,50%	59%	65%

De la tabla 3, se visualiza que el área de tesis más seleccionada es Agentes y Sistemas Inteligentes para ambas carreras. Los tesistas se desempeñaban en desarrollo o en distintas áreas en el momento de inicio del trabajo de tesis.

Con los árboles de decisión obtenidos, se confirma que el área de Sistemas Inteligentes es la más seleccionada con una distribución porcentual no significativa para ambas carreras.

Las tendencias generales encontradas al evaluar los árboles de decisión son las siguientes:

- De la carrera Licenciatura en Sistemas, los árboles de decisión obtenidos muestran que las personas que trabajaban en el rubro Varios, no se inclinan por el área de tesis: Agentes y Sistemas Inteligentes. Tendencia mayoritaria entre todos los tesistas. En cambio, sí la seleccionan los tesistas pertenecientes al área de Desarrollo. Los tesistas que trabajaban como Analistas funcionales, todos se inclinan por el área Ingeniería de Software. Los tesistas que eligen Innovación en Sistemas de Software trabajaban todos en el rubro Varios.

- De la carrera Ingeniería en Informática, se observa que el área de tesis Agentes y Sistemas Inteligentes es elegida tanto por los que trabajaban en Desarrollo, como los que lo hacían en el rubro Varios. Los tesistas que trabajaban en Desarrollo se inclinan también por Innovación en Sistemas de Software. Los que seleccionan Tecnología y Educación trabajaban en las áreas: Analista Funcional, Infraestructura, y Seguridad en Informática.

Los modelos resultantes para cada una de las carreras (Licenciatura en Sistemas e Ingeniería en Informática) formalizan en gran medida la realidad que perciben los docentes de las cátedras de tesis. No obstante, para lograr una mayor cobertura de la realidad respecto a la elección del área de investigación, se debería considerar ampliar la cantidad de atributos que describen al tesista y experimentar con ellos.

3. Conclusiones y futuros trabajos

Los logros alcanzados son solo la etapa preliminar de la investigación que puede realizarse utilizando técnicas de DM. La continuación de este trabajo permitirá avanzar en la utilización de otras técnicas sobre el mismo conjunto de datos y datos futuros que surjan del incremento de la muestra.

Se aplicó el proceso metodológico sugerido en KDD, se resolvieron los problemas encontrados respecto a los datos y a la construcción de los modelos, esto nos permitió comprobar la naturaleza iterativa del proceso.

Se consiguió describir los perfiles de los tesistas aportando información útil, en relación con la incidencia del campo laboral en que se desempeña al inicio de su trabajo de tesis y el área de investigación seleccionada para la tesis.

Como trabajos futuros se identifican: a) el refinamiento de los modelos logrados mediante la incorporación de nuevos atributos para el posterior análisis de su incidencia en la solución del problema; b) la obtención de nuevos modelos, utilizando técnicas de DM aún no abordadas que permitan avanzar en la determinación de las variables incidentes en la selección del área de tesis.

Agradecimientos

La investigación que se en este artículo ha sido financiada por el Proyecto de Investigación titulado: ¨Aplicación de tecnologías inteligentes de explotación de información para el análisis de perfiles de tesistas de grado de carreras informáticas de la UM¨ (Código 17/01-MP-001) de la Secretaria de Ciencia y Tecnología de la Universidad de Morón.

Referencias

1. Plan de Estudios de la Carrera Ingeniería en Informática. Universidad de Morón. Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales. Página web: www.unimoron.edu.ar/area/informatica/stream/af40023e0-ingenieria-en

2. Plan de Estudios de la Carrera Licenciatura en Sistemas. Universidad de Morón. Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales. Página web: www.unimoron.edu.ar/area/informatica/stream/af40023e1-licenciatura-en
3. Marisa Panizzi, Iris Sattolo, Oscar Bravo, Javier Lafont and Nicolas Armilla. Aplicación de tecnologías inteligentes de explotación de información para el análisis de perfiles de tesistas de las carreras de grado de Informática de la Universidad de Morón. Actas de las XXIV Jornadas sobre la Enseñanza Universitaria de la Informática (JENUI 2018), Universitat Oberta de Catalunya, Barcelona.4 al 6 de julio 2018. ISSN: 2531-0607 ISSN: 2531-0607.
4. Iris Sattolo, Gastón Alvarez, Nicolás Armilla, Oscar Bravo, Matias García, Javier Lafont, Gabriel Mariuz, Lucila Mira, Marisa Panizzi. Hacia la caracterización de perfiles de tesistas de Carreras de Informática de la Universidad de Morón. XIII Congreso Nacional de Tecnología en Educación y Educación en Tecnología (TE&ET 2018). Universidad Nacional de Misiones. Posadas, Misiones. Argentina. 14 y 15 de junio 2018. ISBN en trámite.
5. Fayyad, U. M.; Piatetsky-Shapiro, G.; Smyth, P. "From Data Mining to Knowledge Discovery: An Overview" Advances in Knowledge Discovery and Data Mining, pp:1-34, AAAI/MIT. Press.1996.
6. Hernández Orallo José, Ramírez Quintana Maria José, Ferri Ramírez César. Introducción a la Minería de Datos. Ed. Pearson Educacion S.A. Madrid. (2004).
7. Red de Universidades con Carreras en Informática (RedUNCI). Página Web: http://redunci.info.unlp.edu.ar. Disponible online en junio 2018.
8. Waldo Hasperué. Tesis Doctoral en Ciencias Informáticas: ¨Extracción de Conocimiento en Grandes Bases de Datos Utilizando Estrategias Adaptativas¨. Universidad Nacional de La Plata. Facultad de Informática. Marzo 2012. http://sedici.unlp.edu.ar/handle/10915/4215
9. Weka. University of Waikato. Machine Learning Group. Página web: www.cs.waikato.ac.nz/ml/weka/downloading.html. Disponible online en junio 2018.
10. RapidMiner Management Team (S/A). RapidMinerStudio. Página Web: https://rapidminer.com/products/studio/. Disponible online en junio de 2018.
11. Kmine Analytics Platform versión 3.5.3. Página web: www.kmine.com. Disponible online en junio de 2018.
12. J. Quinlan. Induction of decision trees. Machine Learning, 1:81_106, 198. Página web: http://hunch.net/~coms-4771/quinlan.pdf. Disponible online en mayo 2018.
13. L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classi_cation and regresión trees. Wadsworth Int. Group, Belmont, CA, 1984.
14. Álex Moreno-Salazar, M. Puricación Vicente-Galindo & M. Puricación Galindo-Villardón. Aprendizaje basado en árboles de decisión: un estudio crítico desde Weka, RapidMiner y SPSS ModelerXXVI Simposio Internacional de Estadística 2016. Sincelejo, Sucre, Colombia, 8 al 12 de agosto de 2016.

Avances en la Personalización y Adaptación de Pruebas en un Ambiente Virtual de Aprendizaje

Marcela Gonzalez[1,2], Delia Esther Benchoff[1,2], Constanza Huapaya[1,2] , Cristian Remon[1], Guillermo Lazurri[1], Leonel Guccione[1], Franscisco Lizarralde[1,2]

[1]Grupo de Investigación en Inteligencia Artificial Aplicada a Ingeniería, Departamento de Matematica, Facultad de Ingeniería, Universidad Nacional de Mar del Plata, J.B. Justo 4302, Mar del Plata, Argentina

[2]Instituto de Psicología Básica, Aplicada y Tecnología (IPSIBAT),Facultad de Psicología, Universidad Nacional de Mar del Plata y CONICET, Funes 3280 - Cuerpo 5 Nivel 3, Mar del Plata, Argentina

mpgonza@mdp.edu.ar {ebenchoff.sead, constanza.huapaya, remoncristian, guillesky, leonel.guccione, francisco.lizarralde}@gmail.com

Abstract. Este trabajo plantea el análisis de una experiencia de personalización y adaptación para el aprendizaje de estudiantes universitarios. La experiencia presenta pruebas de evaluación formativas, personalizadas y adaptadas en un Ambiente Virtual de Aprendizaje, como complemento de las clases presenciales. El objetivo de esta comunicación es exponer, como contribución, nuestros métodos adaptativos para el aprendizaje. La personalización de las pruebas se basó en la identificación de los estilos de aprendizaje y los niveles de conocimiento previo de los estudiantes. La experiencia se inició en el año 2014 y continúa en la actualidad. Los resultados del análisis muestran que nuestras técnicas de personalización y adaptación, refinadas en el tiempo, mejoran el aprendizaje de los estudiantes, y orientan a los docentes en la toma de decisiones que impactan en la optimización de la propuesta pedagógica.

Keywords: pruebas personalizadas, adaptación, ambiente virtual de aprendizaje.

1 Introducción

Desde el paradigma educativo centrado en el estudiante, la investigación sobre personalización y adaptación del aprendizaje orienta en la actualidad numerosos desarrollos, particularmente en Ambientes Virtuales de Aprendizaje (AVAs). En este contexto, ambos conceptos pueden parecer similares, sin embargo presentan claras diferencias. La personalización en la educación considera la adecuación de la pedagogía, currículo y ambientes de aprendizaje, a las necesidades y estilos de aprendizaje de los estudiantes como individualidades [1]. El aprendizaje personalizado comprende cualquier acción que tanto un profesor como un software ejecuten para enseñar un tema a un estudiante entendido singularmente. Se identifican, por ejemplo, las características personales que diferencian a un individuo del resto del grupo: habilidades cognitivas, dispares niveles de conocimiento,

preferencias y estilos de aprendizaje, entre otras. La personalización puede entonces ser aplicada a ejercicios, explicaciones, pruebas acorde con el perfil individual.

Por otro lado, la adaptación del aprendizaje está relacionada con las funcionalidades adaptativas disponibles en los sistemas de gestión de aprendizaje (LMS). Estos sistemas brindan al estudiante acciones predefinidas, para que seleccione aquellas que orienten su proceso al aprender [2], [3]. Desde el punto de vista del docente, los LMS brindan la posibilidad de planificar la configuración de contenidos, la interface, la evaluación y/o retroalimentación y, de esta manera por un lado, mejorar la performance general del ambiente de aprendizaje centrado en el estudiante. [4], y, por otro, aconsejar y proveer retroalimentación efectiva [5].

En esta comunicación, presentamos los avances de una experiencia que muestra el diseño personalizado de pruebas de evaluación formativa para estudiantes universitarios. La guía principal de nuestro trabajo es la adaptación integrada en la modelización del usuario. Las esferas de análisis indagadas fueron los aspectos asociados a la interacción hombre/máquina en el proceso de enseñanza y de aprendizaje. Nos centramos en las tecnologías empleadas, los materiales utilizados y fundamentalmente, en uno de los aspectos menos tratados en la literatura, el diseño y uso de las pruebas de autoevaluación personalizadas.

2 Técnicas de Personalización y Adaptación

Las técnicas desarrolladas para el proceso de personalización y adaptación se basaron en atributos inherentes al estudiante (para la personalización), en atributos dinámicos de la interacción estudiante/AVA (para la adaptación) y en tópicos del dominio.

En particular, para la personalización se examinaron las preferencias individuales a través de los estilos de aprendizaje y el conocimiento previo sobre el dominio, usando encuestas. Para la adaptación, se analizó la interacción dinámica con los estudiantes, en un ambiente virtual de aprendizaje, se estudiaron los estilos de aprendizaje predominantes en el grupo de estudiantes y los contenidos de mayor dificultad detectados en el aprendizaje. En base a los atributos descriptos, las técnicas de personalización y adaptación fueron construidas centralmente con Cuestionarios Moodle trabajando el acceso, la temporalización y retroalimentación, así como el ofrecimiento de materiales y ayudas personalizadas. A continuación se presenta una breve descripción de los atributos de los estudiantes y de las pruebas formativas.

2.1 Estilos de Aprendizaje

Como individualidades, los estudiantes presentan diferentes estilos de aprendizaje, esto es, manifiestan modos particulares de seleccionar, procesar, absorber y retener información nueva, conformando patrones individuales de aprendizaje, flexiblemente estables [1]. Su consideración para plantear la personalización y la adaptación evidencia una influencia positiva como lo destacan numerosos estudios [6]. En este proyecto se aplicó el instrumento denominado Index of Learning Styles (ILS) [7] propuesto por Felder y Soloman, y dirigido a estudiantes de ingeniería. Ha sido traducido al español y validado en otros ámbitos de conocimiento. [8], [9]

El ILS es un cuestionario de 44 preguntas de opción múltiple que evalúa las preferencias de estilos de aprendizaje a través de cuatro dimensiones, cada una de ellas organizadas en dos categorías. Las dimensiones y sus respectivas categorías son: Procesamiento de información: *estudiantes Activos o Reflexivos*; Percepción de información: *Sensitivos o Intuitivos*; Recepción de información: *estudiantes preferentemente Verbales o Visuales* y Comprensión de la información, cuyas categorías representan *estudiantes Secuenciales o Globales*.

2.2 Conocimiento Previo

El conocimiento previo (CP), puede entenderse como el bagaje de ideas y conceptos que las personas poseen previamente en su memoria a largo plazo, organizado en un formato de redes semánticas. Es considerado como uno de los factores influyentes en el aprendizaje, dado que la puesta en relación entre la novedad y el CP, resulta esencial en el desarrollo del proceso enseñanza –aprendizaje.

De acuerdo con Carretero [10], en todos los niveles educativos, es necesario tener en cuenta los conocimientos previos de los estudiantes, puesto que los nuevos aprendizajes, se asentarán sobre los anteriores. Así, la posibilidad de aprender se encontrará vinculada con la cantidad y calidad de los aprendizajes previos realizados, y con las conexiones que se establecen entre ellos.

2.3 Pruebas Formativas con Cuestionarios

Las pruebas formativas son actividades que tienen por objetivo proporcionar la información necesaria para mejorar el proceso educativo, orientando a estudiantes y profesores. El concepto de retroalimentación efectiva es central tanto en el enfoque de Evaluación orientada al aprendizaje [11] como en las pruebas formativas. Carless afirma que la evaluación debería contribuir eficazmente en la mejora del aprendizaje. Los principios de este enfoque sugieren que: a) la adecuación entre objetivos, contenidos y tareas de evaluación facilita la experiencia de aprendizaje profundo; b) la participación de los estudiantes en la evaluación: autoevaluación, evaluación por pares, retroalimentación de los compañeros, promueve una mejor comprensión de los objetivos de aprendizaje; y c) la retroalimentación adecuada de la evaluación, propicia el compromiso del estudiante con su proceso y permite delimitar dificultades en el aprendizaje de tópicos del dominio y aplicar correcciones apropiadas.

Nuestras pruebas fueron diseñadas como complemento de las clases presenciales de la asignatura, blended-learning. Utilizamos el LMS Moodle, porque es un sistema de gestión del aprendizaje en continuo desarrollo, libre, abierto, escalable y gratuito, con una gran variedad de actividades y recursos que el docente puede incorporar en su propuesta pedagógica. Decidimos construir las pruebas usando la actividad Cuestionarios, por ser la más apropiada para diseñar pruebas formativas de autoevaluación. Esta herramienta permite al docente diseñar y plantear preguntas de diferentes formatos: emparejamiento, opción múltiple, verdadero/falso, respuesta corta y respuesta numérica, de acceso ordenado o aleatorio, con tiempo límite de resolución o no, con intentos ilimitados de respuesta, o restringidos a un número de

veces específico. Focalizamos la atención en la configuración de dos cuestiones relevantes para el logro de nuestro objetivo: la primera vinculada a la definición de las restricciones de acceso, y la segunda a la selección de las opciones de retroalimentación, según la respuesta dada por el estudiante.

3 La Experiencia

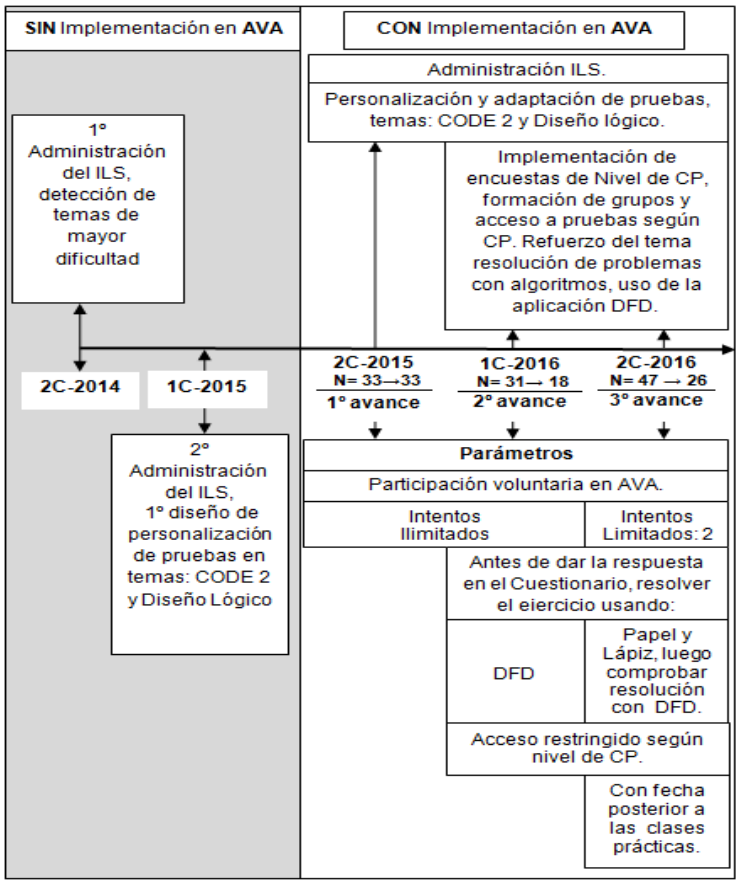
En la Universidad Nacional de Mar del Plata, la carrera Ingeniería en Informática comenzó a dictarse en el año 2014. La asignatura sobre la que se ha realizado este trabajo es Fundamentos de la Informática, ubicada en el 2º cuatrimestre del primer año. Se dicta en ambos cuatrimestres, es decir: cursada y contra cursada. Actualmente estamos cerrando el ciclo correspondiente a la 8º cohorte. Fundamentos de la informática es la primera asignatura de la carrera con contenidos específicos sobre informática, y brinda a los estudiantes una introducción sobre todos los contenidos que aprenderá a lo largo de la carrera.

El total de alumnos matriculados en los períodos analizados fue de 111, la primera implementación en el AVA, se hizo para la cohorte de estudiantes del 2º cuatrimestre del año 2015 (N=33), momento que identificamos como 1º avance, luego continuamos con ambos cuatrimestres del año 2016, (N=31 y N=47 respectivamente).

Para asegurar el mejor resultado posible de la experiencia, se decidió trabajar bajo un enfoque basado en procesos aplicando el Ciclo Deming PDCA: Plan, Do, Check, Act (Planificar, Hacer, Verificar, Actuar), para la mejora continua. [12].

Durante la primera cohorte, el equipo docente decidió incorporar servicios y recursos digitales para facilitar el acceso a los materiales de estudio (se subieron materiales en Blogger y en Google Drive y se mantuvo una comunicación fluida vía mail). Si bien iniciamos la tarea de personalización durante el 2º cuatrimestre del año 2014 al relevar el perfil predominante de estilos de aprendizaje; hubo una demora significativa en la implementación del LMS Moodle, ocasionada por la falta de disponibilidad de un servidor propio, en el cual pudiéramos instalar/configurar, y dar acceso a un AVA mediante una IP pública. Recién pudimos realizar la adaptación de las pruebas personalizadas en el AVA, para los estudiantes de la cohorte del 2º cuatrimestre del 2015. Tomando como insumos los perfiles de estilos de aprendizaje de las primeras cohortes, y los temas de mayor complejidad junto a los errores frecuentes detectados, se inició la personalización del diseño didáctico atendiendo a dichos temas: Lenguaje ensamblador y Diseño lógico. [13] Posteriormente, se consideró en particular el tratamiento del ítem de dominio Algoritmia, fundamental para la formación y abordaje de las siguientes asignaturas vinculadas a la programación, y por su implicancia en la formación específica del ingeniero en informática. Sin descuidar la totalidad de los temas incluidos en la asignatura, se invirtió mayor esfuerzo en el desarrollo de ejercicios de autoevaluación para el tema mencionado.

La Figura 1 resume el recorrido y los avances que realizó el equipo, en relación a la personalización y adaptación en el AVA. Si bien la experiencia continúa, en esta comunicación presentaremos los tres primeros avances mostrando los ajustes del diseño original y los resultados alcanzados.



DFD: software del grupo Smart (Universidad del Magdalena, Santa Marta, Colombia). Apoya la enseñanza de la lógica de programación utilizando diagramas de flujo. [14]

Fig. 1. Recorrido y avances de la experiencia de personalización y adaptación en el AVA

3.1 Perfil Predominante de Estilos de Aprendizaje

El inicio de la personalización se realizó durante el 2º cuatrimestre del año 2014, período en que comenzamos a identificar los estilos de aprendizaje. En el contexto de la experiencia, detallaremos el perfil de estilos de aprendizaje hallado en las cohortes en estudio. Se procesaron 88 protocolos completos: 25 correspondientes al 2º cuatrimestre de 2015, 24 al 1º cuatrimestre de 2016 y 39 estudiantes del 2º cuatrimestre de 2016. Los resultados indicaron balance en la dimensión activo/reflexivo, con una leve tendencia hacia el extremo activo; una mayor tendencia hacia los extremos secuencial y visual; y claro equilibrio en la dimensión sensorial/intuitivo. La Figura 2 ilustra la tendencia recopilada.

Estos resultados anticiparon preferencia por una modalidad de aprendizaje dinámica con el material didáctico y el trabajo en grupo (extremo activo) sin descuidar el acceso a materiales de estudio para el abordaje reflexivo; con interés por

los detalles y siguiendo un orden lógico, paso a paso, en la solución de problemas (extremo secuencial), y a favor del empleo de representaciones visuales, diagramas de flujo, videos (extremo visual).

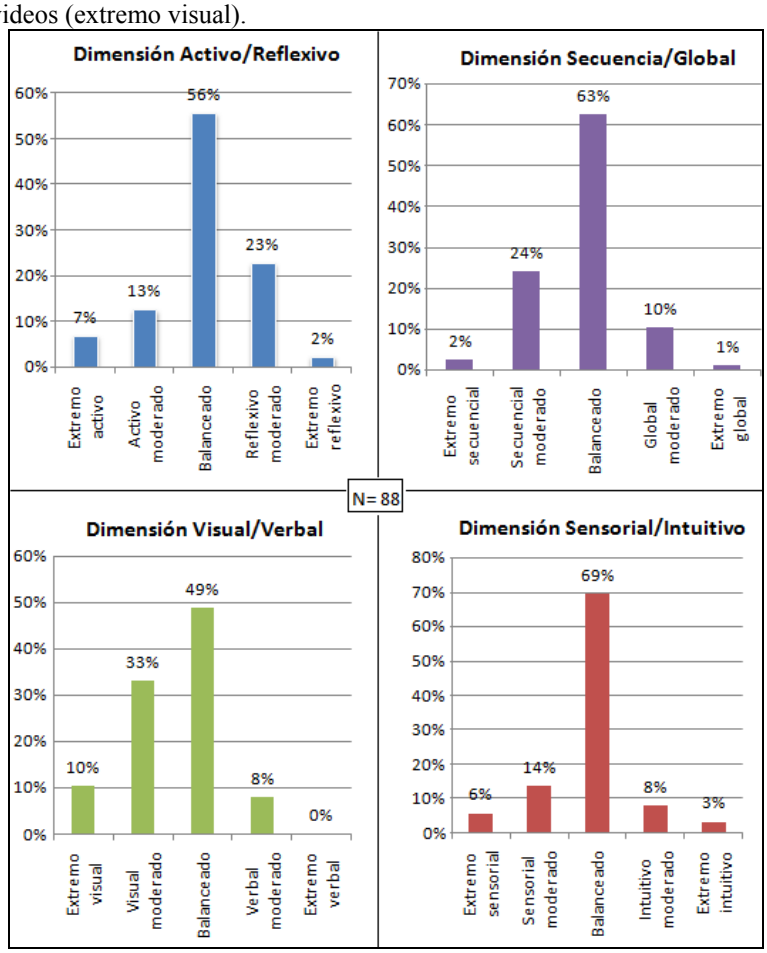


Fig. 2. Perfil de Estilos de aprendizaje, Cohortes: 1C-2015, 1 y 2C-2016

3.2 La Cohorte 2015 - 2º cuatrimestre

Participaron 33 estudiantes. Los temas fueron Diseño lógico y Lenguaje ensamblador, elegidos en consideración de las dificultades en el aprendizaje encontradas en las cohortes anteriores. La participación en el AVA fue voluntaria, aunque se hizo hincapié en la realización de las pruebas como instancia de ayuda y preparatoria a la primera evaluación parcial.

Las pruebas de autoevaluación, se diseñaron en base a los perfiles activo, secuencial y visual. Para cada tema, se confeccionaron dos Cuestionarios, uno de menor complejidad, y otro más complejo. En el tema Diseño Lógico se emplearon preguntas de emparejamiento, y de opción múltiple en Lenguaje ensamblador. Los

estudiantes podían realizar intentos de resolución de manera ilimitada. A fin de minimizar la cantidad de aciertos azarosos se eligió la Retroalimentación Inmediata con Puntuación Basada en Certeza (CBM, Certainty-based marking) en el comportamiento de las preguntas/respuestas. Además, se utilizó la retroalimentación específica según la respuesta del estudiante, ofreciendo ayudas visuales y textos complementarios en correspondencia con el extremo visual del perfil de aprendizaje señalado.

Los resultados de esta primera adaptación, mostraron que aunque todos los estudiantes de la cursada iniciaron los cuestionarios, no todos los finalizaron, ni se mantuvo una sostenida participación a lo largo de la experiencia. En los cuestionarios de baja complejidad de ambos temas, los porcentajes de finalización de los mismos fue mayor: 94% de los estudiantes en el primer tema y 76% en el segundo tema evaluado, en todos los casos con buenas puntuaciones, que incluyeron bonificaciones con CBM. En las pruebas de mayor dificultad, 54% de la muestra finalizó el cuestionario del primer tema y en el segundo tema, se redujo la cantidad de participantes a 24 alumnos, de los cuales el 63% finalizó el cuestionario. En estas pruebas de mayor complejidad, se incrementó la cantidad de intentos de resolución, y las penalizaciones con CBM, dato que permite inferir dudas en las respuestas o menor conocimiento adquirido.

Finalmente, el logro de la práctica complementaria como evaluación formativa, pudo observarse además en los resultados de la primera evaluación parcial de la asignatura, que incluyó los temas referidos. Del total de 33 estudiantes, 28 rindieron el primer parcial. Aprobó el 71% (20 alumnos) y desaprobó el 29 % (8 alumnos). En el conjunto de alumnos aprobados, el 70 % (14 alumnos) finalizaron las pruebas adaptadas de autoevaluación, incluyendo al menos uno de los ejercicios de alta complejidad. El rendimiento mostró que 9 alumnos obtuvieron nota mayor o igual a 7 y 5 alumnos alcanzaron notas entre 4 y 7 puntos.

3.3 La Cohorte 2016 - 1º Cuatrimestre

El total de inscriptos en esta cursada fue 31, y finalizaron el Cuestionario 18 estudiantes, sobre cuyo desempeño se realiza el análisis. El tema evaluado ha sido Algoritmia, destacado por ser transversal a las materias troncales de la carrera. Como insumos de la personalización de las pruebas, además de los perfiles de estilos de aprendizaje, se aplicó una encuesta que indagó sobre el conocimiento previo del tema. A partir de sus resultados, se establecieron 3 niveles de Conocimiento Previo (CP): Nulo (4 estudiantes), Bajo (11 estudiantes), y Medio/Alto (3 estudiantes).

La adaptación de la prueba de autoevaluación se diseñó en el Cuestionario, considerando los estilos de aprendizaje y cada nivel de conocimiento previo, de la siguiente manera: el Cuestionario contó con un total de 10 ejercicios, en un orden creciente de complejidad e integración. Mediante la configuración de una restricción de acceso, los estudiantes de CP Nulo, debían responder el cuestionario completo, cuyo puntaje esperado total era de 17 puntos. Para los estudiantes de CP Bajo, el cuestionario presentó siete ejercicios, donde el primer ejercicio se correspondió con el cuarto del cuestionario completo, es decir, iniciaban en un nivel de complejidad media; el puntaje en este caso fue de 14 puntos. Finalmente, los estudiantes de CP

Medio/Alto, resolvieron los tres ejercicios de mayor complejidad del cuestionario original, y el puntaje esperado fue de 6 puntos.

Atendiendo al estilo activo de aprendizaje, las actividades se centraron en la interacción dinámica con el material. La presentación gradual de los ejercicios con restricciones de acceso y temporalización, se configuró teniendo en cuenta la preferencia de secuencialidad; y para la preferencia visual se ofrecieron diagramas, videos, imágenes y textos complementarios. Para los CP bajo y medio/alto se presentaron videos de revisión de los contenidos necesarios para la solución de los ejercicios.

El desempeño en las pruebas de autoevaluación, no mostró diferencia en los resultados entre los niveles de CP nulo y bajo. Del total de 15 estudiantes ubicados en estos niveles, el 67% alcanzó el puntaje esperado de resolución (10 alumnos), y el 33% se ubicó por debajo de dicho puntaje (5 alumnos). Para el nivel medio/alto, sólo había 3 estudiantes: 2 lograron el máximo puntaje y 1 obtuvo una calificación por debajo de lo esperado.

El seguimiento de los estudiantes en los dos exámenes parciales, se realizó específicamente para el ítem de dominio evaluado. Siempre teniendo en cuenta el nivel de CP, se comparó el rendimiento entre grupos de estudiantes que participaron en el AVA y aquellos que no lo hicieron. Los resultados generales de este seguimiento en ambos parciales, indican que el conjunto de alumnos de CP nulo y bajo en el AVA obtuvo mejores calificaciones en el ítem: la mitad de estos estudiantes alcanzaron puntuaciones superiores al 50% del puntaje requerido en el ítem, mientras que en el grupo sin participación en el AVA, sólo un 25% obtuvo calificaciones equivalentes. Tuvieron además, un notorio ausentismo, debiendo recuperar los exámenes.

El nivel de CP medio/alto, en el AVA sólo tuvieron mejor rendimiento en el ítem correspondiente del 1º parcial, mientras que en el 2º parcial, los alumnos que no participaron del AVA, obtuvieron un desempeño levemente mejor.

3.4 La Cohorte 2016 - 2º Cuatrimestre

En este cuatrimestre, hubo 47 inscriptos, de los cuales 26 participaron en las actividades del AVA, que fueron recomendadas, aunque no obligatorias. Hubo un total de 20 estudiantes ubicados en los niveles de CP nulo/bajo, y 6 en el medio/alto. Se trabajó sobre el tema Algoritmia, al igual que en el cuatrimestre anterior.

A partir de la mejora continua del diseño pedagógico, se solicitó resolver cada ejercicio de la guía de trabajos prácticos usando papel y lápiz, y comprobar la solución con la aplicación DFD durante las clases presenciales. Esta práctica se completó con la presentación de las soluciones (a veces erróneas) de estudiantes en el pizarrón, y la orientación del equipo docente. En el AVA se agregó un parámetro de temporalización, que permitió el acceso al Cuestionario posterior a la fecha de las clases prácticas. Este cambio mejoró la comprensión de los ejercicios. También se modificó el parámetro de *Intentos ilimitados*, a sólo dos intentos, con el objetivo de minimizar las respuestas por ensayo y error, y reforzar la práctica previa. Los demás parámetros se mantuvieron iguales a los del primer cuatrimestre.

Los resultados del Cuestionario, indicaron que un 85% de los alumnos de CP nulo/bajo alcanzaron el puntaje esperado, y en el grupo de CP medio/alto, el 100%.

El seguimiento y desempeño en los parciales, al comparar a los alumnos que realizaron las pruebas de autoevaluación en el AVA, con los que no las hicieron, muestra lo siguiente: el 85% de alumnos de CP nulo/bajo, que realizaron los cuestionarios, (17 estudiantes), obtuvieron altas calificaciones en el ítem, en ambos parciales (superando el rendimiento del cuatrimestre anterior), y todos los estudiantes de CP medio/alto lograron el máximo del puntaje, también en los dos parciales (6 estudiantes). En el grupo que no realizó los Cuestionarios, el 35% de los alumnos de CP nulo/bajo (6 estudiantes), resolvieron el ítem en ambos parciales, y nuevamente se constató gran ausentismo. En el nivel de CP medio/alto, el 25% alcanzó los puntajes para aprobar el ítem (1 estudiante). [15]

4 Discusión y Conclusiones

En primera instancia, los resultados de las pruebas de evaluación formativa, muestran un mejor desempeño en los exámenes parciales de los estudiantes participantes. En tal sentido, la evaluación orientada al aprendizaje como principio rector de la experiencia presentada, [11] permite inferir que efectivamente está contribuyendo al aprendizaje significativo, mediante la adecuación de objetivos, contenidos y actividades. Por otra parte, el diseño de la adaptación apoyado en los estilos de aprendizaje predominantes, permitió la creación de actividades consonantes con las preferencias de los estudiantes, en concordancia con un enfoque pedagógico centrado en el alumno.

Las mejoras progresivas en la adaptación de los Cuestionarios, surgieron con base en la revisión del proceso, y realización de los ajustes necesarios. Como ejemplo, respecto de los contenidos, fue importante focalizar el tema Algoritmia, al relevar su centralidad y las dificultades en su comprensión para estudiantes de 1º año. Desde la perspectiva del diseño de los Cuestionarios, la revisión permitió introducir la variable de nivel de CP, adecuando las pruebas a cada nivel; y aplicando restricciones de acceso, mediante el uso de condicionales en Moodle por grupo y condicionales por fecha de acceso. La complementariedad con las clases presenciales prácticas, incrementó la ejercitación previa al acceso a los Cuestionarios y minimizó las respuestas azarosas. Sumado a ello, la efectiva retroalimentación específica que ofrece la herramienta Cuestionario, especialmente en respuestas incorrectas, completó una secuencia didáctica que favoreció el logro de un mejor desempeño.

Destacamos especialmente el proceso de los estudiantes con CP Nulo/Bajo, que mostraron un incremento en el rendimiento de la resolución de los Cuestionarios en el AVA, y en los puntajes asignados al ítem de dominio, en las evaluaciones parciales. De esta manera, podemos deducir que los estudiantes que iniciaron su proceso de aprendizaje con escasa o ninguna información previa sobre el tema, pudieron construir nuevas redes de conocimientos que resultaron significativas al poder ser aplicadas con éxito en la resolución de los ejercicios propuestos.

En tal sentido, se infiere una relación positiva entre el logro alcanzado por los estudiantes, y los procedimientos de personalización y adaptación aplicados a las pruebas de autoevaluación formativa. Si bien entendemos que las variables que actúan en el aprendizaje no se agotan en las que hemos tomado en esta experiencia (sería

necesario completar los perfiles en relación a motivación y características de personalidad de los estudiantes, por ejemplo), los resultados obtenidos en la experiencia de adaptación hasta el momento parecen indicar que el camino es promisorio.

Referencias

1. Klašnja - Milićević, A., Vesin, B., Ivanovic, M., Budimac, Z. y Jain, L.: E-Learning Systems Intelligent Techniques for Personalization. Springer. (2017). (eBook)
2. Gu Q., Sumner T.: Support Personalization in Distributed E-Learning Systems through Learner Modeling. In: 2nd Information and Communication Technologies, ICTTA, vol. 1, pp. 610–615, (2006)
3. Tian, F., Zheng Q., Gong, Z., Du, J., & Li, R.: Personalized learning strategies in an intelligent elearning environment. In Proceedings of the 11th International Conference on Computer Supported Cooperative Work in Design, pp. 973–978. (2007)
4. González, M.; Benchoff, D., Huapaya, C., Remon, C.: Aprendizaje Adaptativo: Un Caso de Evaluación Personalizada. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología, no. 19, pp. 65-72 (2017).
5. Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid Information Services for Distributed Resource Sharing. In: 10th IEEE International Symposium on High Performance Distributed Computing, pp. 181--184. IEEE Press, New York (2001).
6. Watkins, C.: Meta-learning in classrooms. Scott & Hargreaves (eds) Sage Handbook of Learning, pp. 321. (2015).
7. Soloman, B. A. y Felder, R.: Index of learning styles questionnaire. NC State University. (2005). Disponible en: https://www.webtools.ncsu.edu/learningstyles/
8. Felder, R. y Spurlin, J.: Applications, reliability, and validity of the Index of Learning Styles. International Journal of Engineering Education, Vol. 21(1), pp. 103 – 112. (2005).
9. Rodríguez, J.: Educación médica. Aprendizaje basado en problemas. Médica Panamericana, México. Capítulo 3 pp. 25-45 (2002).
10. Carretero, M.: Introducción a la Psicología Cognitiva. Buenos Aires: Ed. Aique. (1997)
11. Carless, D.: Learning-oriented assessment: conceptual bases and practical implications. Innovations in Education and Teaching International Vol. 44, No. 1, pp. 57–66 (2007)
12. Remón, C., Benchoff, D., González, M., Huapaya, C.: Aplicación de la mejora continua de la calidad para analizar el rendimiento de un grupo de estudiantes de ingeniería. V Encuentro Regional SAMECO (Mar del Plata, 2017)
13. Huapaya C., Gonzalez, M., Benchoff, E., Guccione, L., Lizarralde. F.: Estimación del Diagnóstico Cognitivo del Estudiante de Ingeniería y su mejora con pruebas adaptativas. TEyET 2015 (X Congreso de Tecnología en Educación y Educación en Tecnología). ISBN: 9789506561543. Pp 480-489. (2015)
14. Manual básico de instrucciones y commandos Smart DFD 1.0 (2014). Disponible en: https://informaticaic3.webs.com/MANUAL DFD 2.pdf
15. Benchoff D., González M., Huapaya C.: Personalization of Tests for Formative Self-Assessment. IEEE Revista Iberoamericana de Tecnologías del Aprendizaje. Volume: 13, Issue: 2. Pp.70-74. Print ISSN: 1932-8540. Online ISSN: 1932-8540. (2018)

Normalización de la evaluación de calidad de producto software para el ámbito educativo

Luciano Straccia[1], Paula Zanetti[1], Ma Florencia Pollo-Cattaneo[1]

[1]Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS). Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Argentina.
lstraccia@frba.utn.edu.ar, paula.zanetti@gmail.com, flo.pollo@gmail.com

Resumen. La evaluación de calidad de producto es parte inherente de los procesos de desarrollo de software. Existen diversos modelos y herramientas para realizar una evaluación que asegure maximizar la calidad, pero ningún estándar específicamente para el área disciplinar de la tecnología educativa. En este trabajo, se presenta una propuesta de normalización para la evaluación de calidad de producto para su uso en el ámbito educativo, basado en los objetivos y formatos de la Norma ISO/IEC 25010 y para ser utilizada en forma complementaria con la misma.

Palabras Claves: evaluación de calidad, software, educación, calidad, normas ISO.

1 Introducción

Dentro del ámbito de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional (UTN-FRBA) se ha conformado el Grupo GEMIS (Grupo de Estudios de Metodología para Ingeniería de Software y Sistemas de Información), un equipo de personas con interés en la investigación vinculada a Ingeniería en Sistemas de Información y Tecnología Aplicada a la Educación, que lleva adelante el Proyecto de Investigación y Desarrollo (PID) denominado "Intervenciones tecnológicas en dispositivos didácticos con herramientas de tecnología informática" cuyo objetivo es describir y analizar el uso de la tecnología informática en las intervenciones didácticas de los profesores de las asignaturas de la carrera de Ingeniería en Sistemas de Información (ISI) de la UTN-FRBA y desarrollar nuevos artefactos tecnológicos que favorezcan la mejora en las intervenciones didácticas.

El proyecto ha dado como resultado diversos productos software cuya evaluación de calidad resulta incompleta respecto de las necesidades disciplinares educativas. Por ello, en el presente trabajo se presenta una propuesta para la normalización de la evaluación de calidad de producto software en el ámbito educativo, tomando como referencia modelos preexistentes en la industria del software y se proponen atributos y subatributos de calidad que, utilizados conjuntamente con la ISO/IEC 25000:2005, permiten evaluar la calidad de un software educativo.

Este trabajo presenta conceptos vinculados a la temática del mismo (sección 2), el análisis de antecedentes en la evaluación de calidad de producto para el software

educativo (sección 3), la presentación de la propuesta de normalización para las evaluaciones de calidad de software de dicho ámbito (sección 4), la validación e implementación de la normalización propuesta, incluyendo el impacto de su aplicación (sección 5). Finalmente, en la sección 6, se presentan las conclusiones y futuras líneas de trabajo.

2 Calidad de software y normalización

La calidad de software se refiere al "cumplimiento de los requisitos de funcionalidad y desempeño explícitamente establecidos, de los estándares de desarrollo explícitamente documentados, y de las características implícitas que se espera de todo software desarrollado profesionalmente" [1]. Se entiende que la gestión de la calidad de software implica un proceso eficaz de desarrollo de software de forma tal que produzca un entregable útil que genera un valor, el cual es posible medir tanto para los que lo construyen como para sus destinatarios. Garvin [2] establece 5 diferentes visiones pragmáticas para la calidad: la visión trascendental (que supone que si bien la calidad es algo que se reconoce rápidamente, no es fácil de definir; resulta una suerte de cualidad innata que se percibe pero no necesariamente se describe), la visión del usuario (que es subjetiva y define la calidad en términos del fin particular del destinatario; de la atención de sus necesidades y conveniencias. En otras palabras, si el destinatario está satisfecho, esta visión asume que el producto tiene calidad), la visión del fabricante (si el producto cumple con sus especificaciones originales entonces tiene calidad), la visión del producto (la calidad está directamente relacionada con las funciones y características del producto, siendo medible y precisa a través de atributos del mismo) y, la visión del valor (que indica que la calidad es función del nivel de conformidad del producto, a un costo razonable).

En el ámbito de la calidad de software, existen diversos trabajos vinculados a la normalización o estandarización como " la actividad que tiene por objeto establecer, ante problemas reales o potenciales, disposiciones destinadas a usos comunes repetidos, con el fin de obtener un nivel de ordenamiento óptimo, en un contexto dado, que puede ser tecnológico, político o económico" [3].

Existen estándares internacionales de evaluación de la calidad del producto software como ISO/IEC 14598:1999 (que presenta pautas que ayudan al proceso de evaluación del producto software) e ISO/IEC 25000:2005 (que proporciona una guía para el uso de las nuevas series de estándares internacionales). Desde este último, se desprende el estándar 25010 que permite evaluar la calidad de producto software y establece las características de la calidad, reemplazando y actualizando el anterior estándar ISO/IEC 9126:2001. Estos estándares y normas de calidad no consideran las características propias del ámbito educativo ni de ninguna disciplina en particular, por lo cual cada una de ellas debe determinar su propio modelo complementario.

Dicho estándar se compone de ocho características con sus respectivas subcaracterísticas, que se refieren a propiedades estáticas de software y propiedades dinámicas del sistema informático que proporcionan coherencia terminológica para especificar, medir y evaluar la calidad del producto software y sistemas informáticos. Las características que componen el estándar son funcionalidad, rendimiento,

usabilidad, fiabilidad, seguridad, mantenibilidad, portabilidad y compatibilidad, cada una de ellas con sus correspondientes subcaracterísticas.

3 Antecedentes en la evaluación de calidad de software educativo

Existen metodologías de evaluación de software educativo. Según un trabajo de Alba García-Barrera [4] "desde la inclusión de las TIC en las aulas, han sido numerosos los autores que han recomendado el análisis de ciertas dimensiones de los recursos digitales por medio de indicadores de credibilidad que permitan evaluar su calidad y utilidad educativa. En general, las guías de evaluación que analizan estos factores surgieron en el ámbito de la biblioteconomía y en un principio se enfocaron a los recursos de información disponibles en internet" dando baja prioridad a aspectos educativos y limitando el análisis a ciertos tipos de recursos. Se ha hallado también un trabajo [5], en el cual se analizan diversas metodologías y se realiza la propuesta de una nueva basada en MOSCA (Modelo Sistémico de Calidad de Software); sin embargo, este trabajo, al igual que otros, plantea una propuesta centrada en aspectos informáticos dejando de lado aspectos pedagógicos. Por otro lado, Flores y Rivero [6] plantean la necesidad de generar instrumentos unificados para la evaluación del software educativo, aunque no realizan propuestas concretas.

En [7] se realiza un relevamiento de diversas metodologías y herramientas para la evaluación de software educativo. Su lectura permite comprender las diferentes propuestas generadas en la segunda mitad del siglo pasado. En tanto, en [8] se presentan diversos modelos de evaluación para el área. Según el análisis realizado en [9], la mayoría de estos métodos se basan en las propuestas de Pere Marques Graells [10].

Pere Marques Graells [10] considera que en la evaluación deben considerarse tres categorías de aspectos: funcionales (ventajas que le da al profesor como material didáctico y cómo facilita los aprendizajes de los alumnos), técnicos (diseño de pantallas, interface de comunicación) y pedagógicos (contenido pedagógico de programaciones didácticas). Respecto de este último aspecto, propone la calificación, mediante la valoración según una escala de 4 valores para los siguientes ítems: 1) plan docente (los objetivos docentes se presentan claros y explícitos); 2) motivación (si resulta atractivo e interesante); 3) contenidos (coherencia con los objetivos, veracidad, profundidad, calidad, secuenciación, estructuración, claridad, actualización, corrección gramatical); 4) relevancia de los elementos multimedia (relevancia de la información que aportan para facilitar los aprendizajes); 5) guías didácticas y ayudas (información clara y útil, buena orientación al destinatario); 6) flexibilización del aprendizaje (si incluye diversos niveles y/o itinerarios); 7) orientación del usuario (a través del propio material, consultas o tutoría, sobre el plan docente, los posibles itinerarios a seguir y las opciones a su alcance en cada momento); 8) tutorización de los itinerarios (en función de las respuestas de los usuarios en las actividades de aprendizaje sugiere automáticamente determinados contenidos y/o actividades); 9) autonomía del estudiante (toma de decisiones en la elección de itinerarios, recursos para la autoevaluación y el autoaprendizaje); 10) recursos didácticos (potencialidad y multiplicidad de los recursos didácticos que se utilizan.); 11) múltiples actividades (se

proponen múltiples y diversas actividades formativas); 12) enfoque crítico /aplicativo / creativo de las actividades dirigido a la construcción de conocimiento (no memorístico); 13) aprendizaje colaborativo (inclusión de actividades colaborativas y recursos para ello); 14) corrección de las actividades (la mayor parte de las actividades se corrigen adecuadamente de manera automática o por el consultor); 15) adecuación a los destinatarios de los contenidos y actividades; 16) evaluación de los aprendizajes (sistema de seguimiento y evaluación de los aprendizajes, pruebas de evaluación) y 17) sistema de apoyo docente y tutorial (servicio de consultas, aulas virtuales, tutoría virtual).

Por otro lado, en [9] se presenta un esquema general del Modelo de Aspectos Funcionales, Pedagógicos y Técnicos del Software Educativo, basado en el trabajo de Marques. Este esquema presenta tres etapas de evaluación del software educativo: la primera etapa, corresponde a aspectos funcionales y, la tercera etapa, a aspectos técnicos (ambas etapas vinculadas mayoritariamente a las normas ISO); en tanto la segunda etapa corresponde a los aspectos pedagógicos "que se refiere a la finalidad con la que el software será utilizado, en el cual se verificará el modo en que se han de entregar contenidos a los alumnos, pero también la participación de ellos en la búsqueda o recepción de la información" [9]. Este aspecto presenta los siguientes ítems: contenido temático; recursos didácticos; actividades de aprendizaje; refuerzos de motivación; evaluación clara y precisa; adecuación a los usuarios y facilidad de aprendizaje, aunque no especifica más detalles sobre técnicas de evaluación para los ítems. Finalmente, comparativamente con el modelo original de Marques, resulta de una mayor complejidad dada la falta de información que permita su implementación.

En [11], Reyes-Caballero propone la evaluación de software según tres aspectos fundamentales: pedagógicos, técnicos y tecnológicos. Respecto de los contenidos pedagógicos, está compuesto por seis características integradas por una serie de indicadores que son evaluados con una escala de 5 valores. Las características evaluadas y sus indicadores son: 1) cumplimiento de los objetivos; 2) forma de presentación de la información (imágenes, lecturas, videos, juegos, animaciones); 3) contenido (calidad del material, actualidad); 4) estructura del contenido (organización, presentación o secuencia; e información correcta, suficiente y adecuada); 5) interacción (capacidad de motivación, ambiente amigable, interactividad); 6) evaluación (evaluación adecuada; y preguntas y ejercicios de orientación). Al realizar una comparativa de este modelo con el de Marques, se puede observar una relación directa entre algunos ítems: el ítem 1 de Reyes-Caballero corresponde al ítem 1 de Marques, el ítem 5 de Reyes-Caballero puede ser vinculado al ítem 2 de Marques; el ítem 6, al 16 y los restantes al ítem 3 de Marques. Por lo cual el modelo de Reyes-Caballero resulta incompleto en comparación al de Marques.

En tanto, en [12] se realiza un análisis y comparación de diversos modelos de evaluación de software educativo tales como [13-18]. El trabajo presenta en sus conclusiones la categorización de variables en dos grupos: los aspectos técnicos y los aspectos pedagógicos. Sobre los aspectos pedagógicos contempla "las condiciones del software relativas a los destinatarios (edad, nivel educativo), la enseñanza (objetivos, motivación, retroalimentación, metodología usada), los contenidos (tipos y estrategias) y el aprendizaje que promueve la aplicación". Dicho trabajo propone esta categorización, pero no expresa cuáles elementos de cada uno de los modelos de

evaluación de software educativo corresponden a estas categorías, por lo cual requieren un mayor grado de análisis para su operatividad.

4 Atributos propuestos

La normalización de la calidad del software, tal como se presenta en la Norma ISO 25010, implica la definición de atributos y su desagregado en subatributos. En este trabajo además se consideran diversos ítems para cada subatributo que posibilitan su evaluación concreta. Para la definición de los atributos se utilizan las sugeridas en [12]: condiciones referidas a los destinatarios, enseñanza, contenido y aprendizaje, cuyo detalle se presenta en las siguientes secciones.

4.1. Condiciones referidas a los destinatarios

Para el atributo "condiciones referidas a los destinatarios" se han definido dos subatributos: individualización y adaptación socioculturalideológica. El primero se basa en la propuesta de [14] y evalúa si el software considera las diferencias individuales y si la interacción se adapta al progreso individual. En tanto el segundo, basado en [16], considera si se identifica a qué parte de la sociedad representa y cómo se establecen las otras sociedades y si se adecúa a lo previsto por el cliente y si el software se produce en y para medios socioculturales diferentes al del usuario.

4.2. Enseñanza

Para el atributo "enseñanza" se utilizan como subatributos los presentados en [12]: objetivos, motivación, retroalimentación, metodología.

El primer ítem del subatributo "objetivo" (los objetivos se definen en forma clara) fue obtenido de [18], aunque [13] y [15] presentan un ítem similar, mientras que el segundo (los objetivos son adecuados a los usuarios a quienes va dirigido el recurso informático) corresponde también a [18].

Los ítems correspondientes a "motivación" (para el cual Marques [15] propone ítems similares) y a "retroalimentación" (Clarke [14] presenta algunos ítems similares para esta subcaracterística) fueron obtenidos de [18]. Para "motivación" se consideran los siguientes ítems: motivación al conseguir una meta global interesante para el estudiante y si ofrece premios por respuesta correcta; mientras que para "retroalimentación", si la misma es positiva, amigable y convincente.

Los ítems de "metodologías" (que es redefinido como metodología de enseñanza, para diferenciarlo del subatributo de metodologías de aprendizaje), fueron seleccionados a partir de la conjunción de propuestas de [13; 15]:

- adaptación a la pedagogía (práctica de la educación en el aula);
- capacidad de adaptar el software para que refleje diversos estilos docentes;

- enfoque crítico/aplicativo/creativo de las actividades dirigido a la construcción de conocimiento (no memorístico);
- inclusión de actividades colaborativas.

4.3. Contenidos

Para la definición de subatributos dentro del atributo “contenidos” se utiliza en primer lugar la propuesta de Díaz [18], quien organiza diversos ítems en contenido, ejemplificación, ayudas y evaluación y registro de datos. La mayor parte de los ítems incluidos también son tomados de Díaz, aunque se incorporan otros propuestos por Marques [19].

Para el subatributo “contenidos” se definieron como ítems: si el contenido es claro, preciso, actualizado y de uso práctico y la tutorización de los itinerarios (si en función de las respuestas -acertadas o erróneas- de los usuarios en las actividades de aprendizaje sugiere automáticamente determinados contenidos y/o actividades). Para “ejemplificación”, si los ejemplos son claros y relevantes para ilustrar el contenido y suficientes para comprender al mismo. En el caso del subatributo “ayuda”, si contiene glosario de ayuda para el usuario y ofrece pistas para resolver las situaciones problemáticas. Finalmente, en “evaluación y registro de datos” los ítems seleccionados son:

- si la mayor parte de las actividades se corrigen adecuadamente de manera automática;
- el programa permite el registro del desempeño del estudiante automáticamente al usarlo; y
- se suministra un informe sobre el desempeño del aprendiz

Basados en la sugerencia de [19-20], se incorpora dentro del atributo "Contenidos" un subatributo denominado "valores", en el cual se incluyen diversos ítems tomados de las propuestas de los diferentes autores mencionados: si desarrolla actitudes favorables hacia los valores del ser humano y su ambiente; favorece el trabajo individual, de grupo cooperativo o competitivo; desarrolla contenidos de educación basados en conocimientos previos, edad, niveles educativos; y si rechaza la discriminación de sexo, clase social, raza, religión y creencias.

4.4. Aprendizaje

Los subatributos e ítems correspondientes al atributo de “aprendizaje” se seleccionaron a partir de las propuestas de [17-19], considerándose la evaluación de flexibilidad (del aprendizaje, incluyendo diversos niveles y/o itinerarios) y la relación con la metodología de aprendizaje (si el estudiante es un participante active en el proceso de aprendizaje, si permite aprender a partir de la experiencia y si están presentes o no las teorías del aprendizaje).

5 Validación e implementación de la normalización propuesta

5.1. Software evaluado

Google Forms es un producto que permite crear formularios personalizados para encuestas y cuestionarios [21]. En [22] se llevó a cabo un relevamiento en la cátedra de Sistemas y Organizaciones, correspondiente al primer nivel de la carrera de Ingeniería en Sistemas de Información de UTN-FRBA; allí se verificó el uso de evaluaciones escritas breves, con diferentes denominaciones, como parcialitos o test de lectura, vinculadas a la evaluación diagnóstica. Este tipo de evaluación tiene como objetivo "analizar la situación de cada alumno con respecto a los saberes y conocimientos que posee antes de iniciar un nuevo proceso de aprendizaje y conocer los puntos de partida, para luego estudiar las posibilidades de adaptación de los procesos a las necesidades detectadas" [23]. En algunos cursos de dicha asignatura, a partir del año 2016, la evaluación diagnóstica es realizada con Google Forms, que permite al equipo docente observar los resultados a medida que los alumnos completan el formulario.

5.2. Aplicación de la propuesta de normalización

Se realizó la evaluación del software Google Forms respecto de su uso y objetivos previstos, descritos en el punto anterior. En base a la propuesta de normalización de evaluación de software educativo que es presentado en el presente (y cuya información se encuentra ampliada en [24]) y, con la utilización de una planilla [25] se realizó una nueva evaluación y se concluyó que cumple satisfactoriamente con 7 subcriterios (objetivos, motivación, contenidos, ayuda, flexibilidad, retroalimentación, metodología de aprendizaje) y para 4 subcaracterísticas se ha considerado que no aplican a la evaluación (individualización, adaptación socio-cultural-ideológica, ejemplificación, desarrollo de valores). Respecto de las subcaracterísticas que se consideran que no aplican, se detallan sus fundamentaciones:

- individualización: no es personalizable individualmente ni adaptable, aunque se pueden definir trayectos diferenciales según respuestas; sin embargo, no es un objetivo para el uso planteado;
- adaptación socio-cultural-ideológica: la adaptación queda sujeta a la personalización de preguntas y respuestas desarrolladas por el autor de las evaluaciones;
- ejemplificación: por su uso como herramienta de evaluación, no requiere ejemplificación, aunque podría ser relevante en la instancia de retroalimentación;
- desarrollo de valores: el desarrollo de valores queda sujeto a la personalización de preguntas y respuestas desarrolladas por el autor de las evaluaciones.

Además, se han realizado observaciones en 2 subcaracterísticas que son cumplidas parcialmente: metodología de enseñanza (el informe es limitado y, entre otras cosas, dificultad una segmentación apropiada de los alumnos; por ejemplo, si el mismo formulario se utilizara en más de un curso, no se pueden obtener informes para cada curso) y evaluación y registro de datos (existe un registro adecuado y genera información sobre el desempeño del estudiante, sin embargo el criterio de evaluación de cada pregunta es limitado, ya que no permite respuestas parcialmente correctas ni la aceptación de respuestas alternativas).

5.3. Impacto de la aplicación de la normalización

Tras la identificación de subcriterios que no se cumplen, o se cumplen parcialmente, se han analizado alternativas que permitan resolver los inconvenientes que podrían producirse, buscando que el uso de Google Forms, conjuntamente con otras herramientas, puedan completar la totalidad de los atributos definidos en la propuesta de normalización.

Ante el impedimento de definir respuestas parcialmente correctas o aceptación de respuestas alternativas, se ha previsto el desarrollo de un módulo de revisión de resultados, que permite redefinir las calificaciones obtenidas y que se encuentra en etapa de diseño y codificación. Las especificaciones vinculadas a este módulo y sus enlaces de descarga pueden ser hallados en [26]. Respecto del informe, se ha definido el desarrollo de una herramienta que permite obtener la información de los resultados segmentada por cursos y permitiendo además realizar informes personalizados [27].

En la primera evaluación de la herramienta se consideró también el cumplimiento parcial del subatributo “retroalimentación” por considerar que la misma “es incompleta ya que, si bien informa los resultados, no otorga información adicional sobre los errores”. Sin embargo, a partir de esta identificación y en la búsqueda de posibles soluciones se halló que el producto contiene una funcionalidad que permite generar respuestas diferenciales al usuario en caso de resultado correcto o incorrecto para cada pregunta realizada.

6 Conclusiones

La aplicación del modelo de calidad y sus atributos definidos en la familia de Normas ISO 25000 permiten abordar la evaluación de calidad de un producto software. Sin embargo, cuando se espera que este software sea evaluado en función de características propias de la disciplina a la cual sirve, se requiere un modelo propio de evaluación de producto en el marco disciplinar correspondiente. En este trabajo se llevó a cabo el análisis de diversas propuestas de evaluación de software educativo, con el aporte sustancial de ser analizados en la búsqueda de construcción de un modelo que complemente a las normas de calidad de software generales. Se presenta una propuesta de normalización completa con sus atributos y subatributos y algunos ítems u observaciones que permiten comprenderlo de mejor manera. Tal como ocurre con las Normas ISO de evaluación de calidad de producto, es pertinente en la

continuidad de esta línea de trabajo generar herramientas que permitan evaluar de manera más clara y objetiva cada uno de los ítems previstos.

La aplicación de la normalización presentada en este trabajo ha permitido la identificación de requerimientos sobre el sistema de evaluación diagnóstica que, a través de alguna herramienta, complementen las funcionalidades que actualmente brinda el producto software Google Forms. Aunque el desarrollo y la implementación de estas nuevas herramientas no son objeto central de este trabajo, se presentan las especificaciones iniciales de las mismas que permiten comprender su vinculación con las necesidades surgidas como resultado de la aplicación de la normalización.

Tal como se expresó anteriormente, en próximos trabajos se analizarán herramientas para evaluar de manera más eficiente cada subatributo presentada. Además, se prevé la aplicación de la normalización en este trabajo presentado sobre algún software con menor grado de incidencia por parte del equipo docente sobre los contenidos, por ejemplo, WinQSB, herramienta vinculada al tratamiento de los problemas de gestión mediante herramientas cuantitativas aplicada en la asignatura Investigación Operativa, sobre la cual cuya evaluación el grupo se encuentra trabajando.

Referencias

1. Pressman, R. (2005). Ingenieria del Software. 6ª Ed. Mcgraw-Hill.
2. Garvin, D (1984). What Does Product Quality Really Mean?. Sloan Management Review.
3. IRAM (s.f.). ¿Qué es una norma?. Instituto Argentino de Normalización y Certificación.
4. García-Barrera, A. (2016). Evaluación de recursos tecnológicos didácticos mediante e-rúbricas. Revista de Educación a Distancia.
5. Díaz-Antón, G., Pérez, M., Grimán, A., Mendoza, L. (s/d). Instrumento de evaluación de software educativo bajo un enfoque sistémico
6. Flores, H.C. y Rivero, P. (2014). ¿Cómo selecciona el profesorado recursos digitales para enseñar el Historia? CLIO. History and History teaching, 40. Recuperado de http://clio.rediris.es/n40/articulos/FloresRivero2014.pdf
7. Cataldi, Z. (2000), Metodología de diseño, desarrollo y evaluación de software educativo. Disponible en http://laboratorios.fi.uba.ar/lsi/cataldi-tesisdemagistereninformatica.pdf
8. Ángel O., Y. y López S., M.A. (2014). Estudio comparativo de modelos para evaluar software educativo. Tesis de licenciatura, Universidad Juárez Autónoma de Tabasco, Cunduacán, Tabasco, México.
9. Rubén Jerónimo Yedra, Eric Ramos Méndez, Ferrer Sánchez Jesús Antonio (2015). Análisis para determinar cómo se valúa una herramienta de software de tipo educativo. En Revista CONAIC. Segundo Número Especial.
10. Marquès, P. (1995): Citado por Cataldi, Z. (2000): Metodología para la elaboración de software educativo en Software Educativo. Guía de uso y metodología de diseño. Barcelona: Editorial Estel.

11. Reyes-Caballero, F.; Fernandez-Morales, F.; Duarte, J. (2015). Herramienta para la selección de software educativo aplicable al área de tecnología en el nivel de educación básica. En Revista Entramado. Vol. 11 No. 1, 2015 (Enero - Junio)
12. Cova, A., Arrieta, X y Riveros V. (2008). Análisis y comparación de diversos modelos de evaluación de software educativo. Enl@ce: Revista Venezolana de Información, Tecnología y Conocimiento, 5(3), 45-67.
13. McDougal, A y Squires, D. (1997,2001). Cómo elegir y utilizar software educativo. Madrid: Ediciones Morata.
14. Clarke, P; Pete, M; Naidoo, A (1997). Evaluation of Software and its effect on Learning. Project: Kwazulu Concept Burger: HTML mind map / model of evaluation of educational software. SITE Conference, Orlando, US.
15. Marques, P (2005). Entorno formativo multimedia: elementos, plantilla de evaluación/criterios de calidad.
16. González, M. (2003). Evaluación de Software Educativo. (Orientaciones para su uso pedagógico). En Zea, C. (2003). Conexiones, Informática y Escuela: Un enfoque global. Universidad EAFIT, Medellín
17. OTA (1988). Power on! New tools for teaching and lear- ning. U.S.Congress, Office of Technology Asses- tment, SET-379, September, Government Prin- ting Office, Washington D.C.
18. Diaz, M; Perez, M.; Mendoza, L. y Griman, A. (2003). Calidad Sistémica del Software Educativo. Congreso Internacional Edutec 2003, Gestión de las Tecnologías de la Información y la Comunicación en los diferentes ámbitos educativos. UCV. Caracas.
19. Marqués, P. (2004). Plantilla para la Catalogación y Evaluación Multimedia.
20. Gómez del Castillo, M. (2002). Un instrumento para evaluar valores en el software educativo.
21. Google GSuite (s.f.). Formularios: encuestas y formularios fáciles de crear en todo el mundo. Disponible en https://gsuite.google.com/intl/es-419/products/forms/
22. Straccia, L.; Marino Aguirre, M.; Acosta, M.; Vegega, C.; Pytel, P.; Pollo-Cattáneo, M.F. (2015). El desarrollo de artefactos de tecnología informática como aporte a las intervenciones didácticas. Memorias de 3er. Congreso Nacional de Ingeniería Informática y Sistemas de Información Universidad Tecnológica Nacional, Facultad Regional Buenos Aires.
23. Dirección Nacional de Información y Evaluación de la Calidad Educativa, DiNIECE. Pautas pedagógicas para elaborar evaluaciones diagnósticas. Ministerio de Educación de la Nación, Argentina.
24. Straccia, L; Pollo-Cattaneo, M.F. (2017). Evaluación de Software Educativo. Documento de Trabajo Nro. 03. Grupo GEMIS. Disponible en https://drive.google.com/open?id=1gY1W0G68-V51o0qCFwMtBDc8wy7zjQhS
25. GEMIS (2018). Evaluación de Calidad de Software. Aspectos Educativos. Planilla de Trabajo Nro. 03. GEMIS. Disponible en https://goo.gl/jRJwzf
26. GEMIS (2018). Google Forms - Revisión de Resultados. Especificación de Componente. GEMIS. Disponible en https://goo.gl/eNtBXp
27. GEMIS (2018). Google Forms – Generación de Informes. Especificación de Componente. GEMIS. Disponible en https://drive.google.com/open?id=1YFtA-0Fl7iq-jfiHprgr8_Z7B0Kd-xEY

Análisis de Interacciones Sociales en Facebook

Aplicaciones en la enseñanza-aprendizaje

Paula Echenique[1], Regina Motz[2]
[1] Maestría en Informática PEDECIBA Informática, Montevideo, Uruguay
paula.echenique@cfe.edu.uy
[2] Instituto de Computación, Facultad de Ingeniería, Montevideo, Uruguay
rmotz@fing.edu.uy

Abstract. La aplicación de métricas de análisis de datos en grafos para dar explicaciones a las interacciones sociales que ocurren dentro de un grupo de Facebook aportan conocimiento útil para apoyar los procesos de enseñanza-aprendizaje. En este trabajo, se analizan e interpretan estas interacciones para los estudiantes y docentes que forman parte de un grupo de Facebook ilustrando la manera en que aportan a los docentes conocimiento clave para la toma de decisiones pedagógicas que beneficien al estudiante. Este análisis se puede aplicar a otros tipos de grupos de Facebook que no sean sólo de enseñanza y aprendizaje.

Keywords: redes sociales; Facebook; métricas; análisis de redes sociales; enseñanza-aprendizaje

1 Introducción

Los sitios de redes sociales (en más SRS) se han convertido en una parte importante de la vida cotidiana de las personas y por lo tanto, de los estudiantes. Los diversos SRS brindan a sus usuarios posibilidades de crear y participar en diversas comunidades a través de funciones como la comunicación, el intercambio, la colaboración, la publicación, la gestión y la interacción. Como se expresa en Vivian et al. (2014) [1] es común que los estudiantes de diversos niveles utilicen SRS para apoyar sus actividades de aprendizaje, formando parte también de su vida académica.

Los SRS y en particular, las redes sociales ofrecen a la enseñanza una amplia gama de posibilidades que facilitan la interacción social, creación de contenido, descubrimiento e intercambio de información, establecimiento de relaciones sociales y agrupación de información. Por otra parte la tecnología de estos servicios proporciona herramientas que posibilitan el acceso y la recolección de los datos generados tanto en forma de perfiles personales como en forma de interacciones.

La importancia de la interacción social en la construcción del conocimiento y el desarrollo de poderosas herramientas como los SRS motivan nuevos desafíos. Por un lado, el uso de redes sociales en contextos educativos se presenta como una oportunidad con múltiples ventajas. Por otro lado, la posibilidad de obtener en forma automática la información generada como resultado de la utilización de redes sociales en

contextos educativos, abre oportunidades en el plano del análisis de esta información con fines pedagógicos. La combinación del uso de redes sociales como apoyo al trabajo en el aula con herramientas de redes y de análisis de redes sociales puede proporcionar información relevante tanto para el docente como para los estudiantes.

Como caso particular se considera la red Facebook, mediante su herramienta de Grupo. Por sus características y posibilidades resulta ser un buen espacio de apoyo para los estudiantes de un curso presencial. En este trabajo se estudian las interacciones que se producen en un Grupo de Facebook al cual están asignados un docente y los estudiantes de un curso.

El desarrollo del trabajo consiste en la construcción de una herramienta de extracción de las interacciones ocurridas en un grupo de Facebook. Para el análisis de los datos se propone la aplicación de métricas de redes sociales. Se identifican los indicadores a los que aportan las interacciones sociales en contextos educativos. Se proponen interpretaciones semánticas para los datos obtenidos y se presentan los resultados de forma que sirvan como apoyo a los docentes en la toma de decisiones pedagógicas a partir de información obtenida de las interacciones de sus estudiantes.

El documento está organizado de la forma siguiente. En la sección 2 se presentan trabajos relacionados. En la sección 3 se presenta la herramienta creada para extraer las interacciones de un grupo de Facebook. En la sección 4 se definen las dimensiones y los indicadores para el análisis de las interacciones que se producen en un grupo de Facebook en un contexto educativo. En la sección 5 se introducen las métricas a aplicar a los datos para las mediciones, y se realiza una propuesta de interpretación.

2 Trabajos relacionados

En Said at al. (2010) [2] se argumenta acerca de algunas de las ventajas del uso de redes sociales en el aprendizaje. Las redes y las tecnologías asociadas apoyan el aprendizaje colaborativo, ya que posibilitan la creación de comunidades en línea que permiten a los estudiantes participar en actividades de aprendizaje social y en la construcción de conocimientos compartidos socialmente. En el aprendizaje colaborativo, se espera que los estudiantes trabajen juntos hacia el proceso completo de construcción del conocimiento y aprendan no sólo a apreciar su propio trabajo sino también las contribuciones de sus compañeros Un beneficio clave de participar en un trabajo grupal colaborativo es que el alumno tiene la oportunidad de participar, de aprender de otros compañeros, de aprender a asumir la responsabilidad de su aprendizaje y de esta forma desarrollar más autonomía en su aprendizaje. El rol del docente pasa de ser instructivo a ser de apoyo y actúa más como facilitador y coordinador para estructurar y guiar la dirección general del aprendizaje de los estudiantes.

En Said et al. (2013) [3] los autores realizan una clasificación de los tipos de interacciones asociadas con el aprendizaje que se producen Facebook, éstas son: estudiante - estudiante, estudiante - docente, estudiante - contenido y estudiante - interfaz. Según los autores las interacciones en Facebook pueden aportar datos sobre la dimensión psicosocial de los estudiantes de un curso, además de información sobre la adquisición de conocimientos y habilidades. Las interacciones pueden transmitir infor-

mación valiosa a lo largo del tiempo, además de proporcionar datos tanto informales como formales. Para analizar las interacciones utilizan un enfoque de cinco dimensiones: participativa, interactiva, social, cognitiva y de valores, tomando como referencias las definiciones de Henri (1992) [4].

Las dimensiones propuestas por Henri son:

- Dimensión participativa. Cantidad de mensajes o publicaciones de un estudiante.
- Dimensión interactiva. Utiliza una definición operacional de interacción en tres pasos: comunicación de información, una primera respuesta a esta información, una segunda respuesta relacionada con la primera.
- Dimensión social. Refiere a la dinámica social de los intercambios.
- Dimensión cognitiva. Refiere a la identificación los elementos dentro de los mensajes que dicen algo sobre las formas en que las personas aprenden.
- Dimensión metacognitiva. En un entorno de aprendizaje colaborativo mediado por computadora, el examen de los mensajes puede ser una valiosa fuente de información sobre actividades metacognitivas.

La propuesta de Henri es retomada por Pozzi et al. [5]. Estos autores establecen indicadores para investigar las dimensiones participativa, interactiva, social y cognitiva:

Para la dimensión participativa: cantidad de mensajes enviados, cantidad de documentos cargados, cantidad de mensajes leídos, cantidad de documentos descargados, distribución de la participación a lo largo del tiempo.

Para la dimensión interactiva definen como indicadores más significativos los que provienen del análisis de contenido cuantitativo y cualitativo de los mensajes y documentos intercambiados por los estudiantes durante el proceso de aprendizaje. Ejemplos: cantidad de mensajes leídos por un alumno antes de publicar los suyos propios, número de respuestas a los mensajes de otros estudiantes, el número de citas implícitas o explícitas de los mensajes de otros estudiantes.

Para la dimensión social proponen identificar señales que atestigüen el afecto y la cohesión dentro de los actos de comunicación. Se expresan a través de unidades temáticas, unidades de significado que transmiten un único elemento de información extraído de un segmento de contenido. Consideran los siguientes indicadores: unidades temáticas caracterizadas por el afecto o por la cohesión.

Para la dimensión cognitiva que la definen como la medida en que los alumnos pueden construir y confirmar el significado a través de la reflexión y el debate en una comunidad, proponen la medida de la presencia de unidades temáticas que se refieren a las siguientes fases: revelación, reconocer un problema; exploración, expresar acuerdo / desacuerdo; integración, conectar ideas, hacer síntesis; resolución, aplicaciones a la vida real, soluciones de prueba.

3 Extracción de datos

La extracción de los datos de interacciones de un grupo de Facebook se realiza mediante una aplicación construida en Python, que utiliza la interfaz que Facebook proporciona a los desarrolladores.
Facebook representa la red con un grafo, denominado grafo social de Facebook. A medida que Facebook crece, se construye su grafo. Cada nuevo usuario que se registra ocupa una posición en el grafo y a partir de él se genera nuevas relaciones. En el grafo de Facebook están integrados no sólo las personas, sino elementos que no son personas físicas como páginas, grupos, eventos. Todas estas entidades son tratadas en Facebook como objetos del grafo.

Cada objeto tiene un identificador único y todos los objetos se relacionan entre sí, igual que las personas. En el caso de los grupos, por ejemplo, cada uno se convierte en un objeto del grafo con su identificador único, que relaciona otros objetos del grafo mediante un tipo de relación, que no sería "es amigo de" como en el caso de las personas, sino "es miembro de".

Facebook provee una herramienta para desarrolladores que permite acceder al grafo para lectura y escritura. Esta se denomina API Graph. (API por las siglas en inglés de Interfaz de Programación de Aplicaciones y Graph por la representación de la red mediante un grafo). La herramienta para este trabajo se ha desarrollado utilizando la versión 2.9 de la API Graph, cuya descripción se obtuvo de [6]. Con respecto a esta herramienta de extracción para grupos de Facebook, es importante tener en cuenta que en abril de 2018 Facebook inició una serie de cambios estructurales en la API Graph, como consecuencia de los sucesos vinculados a Cambridge Analytica [7]. En el momento de finalizar este trabajo se ha actualizado la API Graph a su versión 3.0, donde se deshabilitan algunas de las funcionalidades utilizadas aquí.. El acceso a través de la API 2.9 tenía un conjunto de restricciones que han sido ampliadas en la versión 3.0, acorde a la nueva política de privacidad de Facebook.

La herramienta obtenía todas las interacciones ocurridas en un grupo de Facebook: publicaciones, comentarios y reacciones. Cada una de ellas contenía la identificación de su autor. La información de autor fue utilizada para el análisis individual y grupal de las interacciones que se describe en este trabajo. En la versión 3.0 de la API, la identificación del autor de cada interacción ya no está disponible.

4 Indicadores para el análisis de interacciones

El enfoque de Henri es la base para el análisis de interacciones abordado en este trabajo. Para poder estudiar las dimensiones propuestas se deben establecen indicadores que permitan adecuarse a las necesidades de presentar información al docente que resulte de interés y utilidad en la toma de decisiones. Para delimitar el alcance del trabajo al análisis de aspectos cuantitativos se seleccionan dos dimensiones: participativa e interactiva. Esta selección se realiza para establecer un conjunto de indicadores para el análisis cuantitativo de las interacciones. La propuesta se basa en el trabajo de Said et al. [8], que establecen una serie de indicadores para cada dimensión asociada a

intercambios en Facebook. A continuación se presentan los indicadores propuestos en [8] para las dimensiones seleccionadas y los que se consideran en este trabajo.

4.1 Dimensión participativa

La dimensión participativa se define por el concepto de "hacer algo" que muestre que el estudiante está presente. El interés de los indicadores definidos radica en que permiten observar no las acciones visibles de los estudiantes sino también su escasa o nula participación, y su discontinuidad. Estos pueden proporcionar información sobre situaciones de alerta al docente. En la Tabla 1 se presentan los indicadores..

Tabla 1. Indicadores de participación

Indicador	Propuestos en Said et al. [8]	Trabajo actual
Participación activa	Cantidad de publicaciones que los estudiantes hicieron en la discusión de Facebook.	Cantidad de publicaciones originales realizadas en el Grupo de Facebook más cantidad de comentarios realizados en publicaciones de otros participantes
Participación pasiva	Cantidad de veces que los estudiantes vieron publicaciones específicas dentro del grupo	Cantidad de reacciones de cada miembro del grupo a las publicaciones de otros
Continuidad	Distribución de la participación a lo largo del tiempo	Línea temporal de particupación activa y pasiva por separado

4.2 Dimensión interactiva

Las interacciones las definen como las relaciones que los estudiantes construyen durante el proceso de aprendizaje. Se mide por las influencias recíprocas que las intervenciones pueden tener y por la frecuencia de las intervenciones. En este trabajo se indicadores se definen por el contenido cuantitativo consideran desde el punto de vista cuantitativo del intercambio. En la Tabla 2 se presentan los indicadores de la dimensión interactiva.

Tabla 2. Indicadores de interacción

Indicador	Propuestos en Said et al. [8]	Trabajo actual
Cuantitativo	Número de respuestas a los mensajes de otros estudiantes	1. Cantidad de comentarios realizados a publicaciones de otros o a comentarios de otros.
		2. Cantidad de comentarios que otros realizan a las publicaciones o comentarios de un alumno.
		3. Frecuencia de comentarios a publicaciones o comentarios de otros considerando la línea temporal.

5 Aplicación de Métricas

El análisis de interacciones se realiza sobre indicadores medidos a partir de datos cuantitativos que arrojan información cualitativa sobre los grupos y sus estudiantes. Se propone analizar los indicadores descritos en la sección 4, mediante la aplicación de métricas de análisis de redes sociales. En esta sección se definen las métricas aplicadas y se presenta una posible interpretación de cada una de ellas en el contexto de las interacciones en un grupo de Facebook.

Para probar la aplicación de las métricas a datos extraídos de Facebook, se utilizaron datos de prueba. Se crean para ilustrar el conocimiento útil que el docente puede obtener y usar para tomar decisiones que impacten favorablemente sobre los procesos de aprendizaje de los estudiantes. Los datos generados no permiten aseverar nada sobre individuos específicos, pero sí permiten proponer la detección de posibles situaciones de riesgo o de uso positivo de la influencia de los diferentes miembros de la red social integrada por las personas que interactúan simultáneamente en un salón de clases y en un espacio virtual como los grupos de Facebook.

Se utiliza un modelo de grafo para representar los resultados obtenidos según los indicadores descritos. La ventaja del grafo es que proporciona una visualización organizada, que se puede presentar al docente de forma que la lectura de la información no requiera un esfuerzo adicional y facilite su labor. Se describe el modelo utilizado para algunos de los indicadores presentados anteriormente.

5.1 Indicador de participación activa:

Se construye un grafo dirigido donde los nodos representan a los alumnos y al docente. Se agrega un nodo Grupo que representa la página del grupo, de esta forma se modelan las publicaciones como aristas con origen en quien publica y destino el nodo Grupo. El peso que se le asigna a cada arista corresponde con la cantidad de publicaciones realizadas por la persona a quien representa el nodo. Para el análisis de la participación activa de esta dimensión se analiza la medida del grado de salida con peso, de forma que los nodos de mayor tamaño corresponden a las personas que han tenido mayor participación activa a lo largo del curso.

5.2 Indicadores 1 y 2 de la dimensión interactiva

Se construye un grafo en el que los nodos representan a los alumnos y al docente. Las aristas representan los comentarios realizados a publicaciones o comentarios de otros. Las aristas tienen origen en quien realiza el comentario y destino en el autor de la publicación a la que se realiza el comentario. El peso que se le asigna a cada arista corresponde con la cantidad de comentarios realizados por el origen en publicaciones realizadas por el destino. Para este grafo, se analizan por separado las métricas de centralidad.

5.3 Métricas

Para definir e interpretar las métricas se ha consultado [9] y los trabajos realizados en el marco del proyecto DIIA [10]. Las medidas de centralidad se han utilizado ampliamente para explotar las redes y sirven para detectar los nodos relevantes en un gráfico. En lo social las centralidades de análisis de red se utilizan para identificar, por ejemplo, personas influyentes en la red, comunidades, detectar comportamientos extraños en la red. Esta interpretación se aplica en este trabajo para una comunidad educativa en red.

La centralidad de grado se calcula como el valor promedio del número de aristas de entrada y de salida en un grafo dirigido. El grado de salida –indicador 1 de la dimensión interactiva - se pude interpretar como un indicador de la influencia. Un miembro con un alto grado de salida, tiene muchas conexiones directas con otros y se entiende que puede transferir información rápidamente a muchos otros.El grado de entrada – indicador 2 de la dimensión interactiva - se puede interpretar como un indicador de la importancia en la red a través del soporte que recibe de otros miembros de la red.

La centralidad de proximidad mide la distancia promedio desde un nodo a todos los otros nodos en la red. Indica qué tan cerca está un nodo de los demás. Un miembro con centralidad de proximidad alta tiene una comunicación rápida con todos los miembros de la red. La excentricidad mide la distancia desde un nodo hasta el nodo más lejano a él teniendo en cuenta las rutas ponderadas de la red. Permite encontrar los nodos que están lejos del nodo más central en la red. La centralidad de interconexión mide la frecuencia con la que un nodo aparece en el camino más corto entre nodos en la red. Es una medida que permite saber que tan importante es un nodo en la ruta más corta entre cada par de nodos. Un nodo con una centralidad de interconexión alta tiene una gran influencia en la red debido al flujo que pasa a través de él. La centralidad de vector propio mide la importancia del nodo en una red basándose en sus conexiones. Los nodos que poseen un valor alto de esta medida de centralidad están conectados a muchos nodos que a su vez están bien conectados, por lo que son buenos candidatos para difundir información.

Además de las métricas de centralidad se considera el coeficiente medio de clúster. Este indica cómo se incrustan los nodos en su vecindario, lo que puede considerarse como una indicación de lo que se denomina "mundo pequeño" o agrupamiento en la red. El hecho de tener vecinos más unidos influye en la ubicación estructural de los estudiantes en la red, lo que a su vez facilita la discusión y motiva a los estudiantes a permanecer en el curso. La falta de apoyo y de participación facilita la posibilidad de abandonar el curso.

En la Tabla 3 se presenta una propuesta de interpretación de las métricas de centralidad en el contexto de las interacciones en un grupo de Facebook de un grupo de aula. La propuesta pretende proporcionar información cualitativa a las mediciones. A cada nivel de la interpretación presentada en la Tabla 3 se le puede proveer de contenido semántico. Por ejemplo, para la entrada de la Tabla 3 correspondiente a Grado de Entrada Bajo en las interacciones estudiante – docente se podrían establecer diferentes conclusiones: Puede tratarse de un estudiante independiente y con muy buena escolaridad, o puede deberse a que es un estudiante desinteresado por el curso con riesgo de

abandonar. En la entrada Grado de Entrada Bajo en las interacciones estudiante – estudiante puede tratarse de un estudiante con muy buena escolaridad, pero sin interés en colaborar con sus pares o puede deberse a un caso de baja popularidad y aislamiento, o ambas.

Tabla 3. Interpretación de métricas de centralidad en interacciones de grupo

Métrica		**Tipo de interacción**	
		Estudiante-Docente	***Estudiante-Estudiante***
Grado de entrada	Alto	Estudiante o docente prestigioso, tiene mucha importancia dentro de la red.	Estudiante que tiene mucha importancia en la red. Cuenta con soporte de sus pares
	Bajo	Estudiante o docente que no tiene soporte en la red..	Estudiante que no tiene soporte en la red de pares.
Grado de salida	Alto	Estudiante o docente muy sociable, altamente comunicativo. Es influente en la red.	Estudiante con mucha influencia entre sus pares.
	Bajo	Estudiante o docente poco comunicativo o reservado.	Estudiante con baja influencia entre sus pares. Poco comunicativo y sociable.
Proximidad	Alto	Estudiante o docente que tiene mucho alcance en la difusión de mensajes.	Estudiante con capacidad de comunicar rápidamente a sus pares. Puede ser influenciado o influenciar a los demás
	Bajo	Estudiante o docente que tiene poco impacto (alcance) para difundir información en la red.	Estudiante que tiene poco impacto (alcance) para difundir información entre sus pares.
Excentricidad	Alto	Estudiante o docente con una amplia cobertura de la red. Es fácilmente accesible por todos los integrantes del grupo Idealmente el docente debería tener este valor.	Estudiante con un amplio rol dentro de la red. Es fácilmente accesible por todos los integrantes del grupo, puede ser influenciado o influenciar a los demás.
	Bajo	Estudiante o docente con una cobertura limitada dentro de la red. No accesible por todos los integrantes del grupo.	Estudiante con un rol limitado dentro de la red. No accesible por todos los integrantes del grupo.
Vector propio	Alto	Estudiante o docente influyente, importante para difundir información en la red.	Estudiante estratégico para difundir información entre sus pares.
	Bajo	Estudiante o docente con poca influencia para difundir información en la red.	Estudiatne con poca capacidad para difundir información
Interconexión	Alto	Estudiante o docente estratégico para comunicar entre subgrupos de la red.	Estudiante estratégico para comunicar entre subgrupos de la red de pares.
	Bajo	Estudiante o docente con poca influencia para comunicar en la red.	Estudiatne con poca influeencia y poca capacidad para intercomunicar a los miembros de su red.

6 Conclusiones y Trabajo futuro

Se ha presentado un análisis de interacciones en un entorno de aprendizaje en una red social. Para ello se ha construida una herramienta de extracción automática de los datos de interacciones de un grupo de Facebook. A los datos obtenidos se le han aplicado métricas de análisis de redes sociales. Se introducen grafos para la visualización de la información.

Los grafos son una forma de visualización flexible que facilita la representación de una forma natural de una gran diversidad de situaciones. Permiten enriquecer la información proporcionado vistas sobre las relaciones que ocurren entre los miembros de un grupo. Hacen posible la detección de diferentes situaciones individuales o grupales, en el contexto de aprendizaje pueden ser la presencia de estudiantes aislados o la presencia de estudiantes facilitadores de la comunicación, permiten detectar a tiempo situaciones de riesgo. Proporcionan información sobre el rol que desempeña el docente en la red del curso.

Es posible ampliar y enriquecer la información que se obtiene de los datos extraídos incorporando aspectos cualitativos el análisis que no han sido abordados en este trabajo. La incorporación de herramientas de análisis semántico se considera el próximo paso, para utilizar información cualitativa que aporte a la dimensión interactiva, también se incorporará el análisis de la dimensión social y de la dimensión cognitiva. El análisis de los datos extraídos muestra que es posible obtener información acerca de la comunicación como expresiones emocionales de afecto y de cohesión: emoticones, saludos, aliento, felicitaciones, agradecimiento. También es posible el procesamiento de las interacciones de forma que arroje información sobre el aprendizaje, si reproduce un enfoque, si repite lo que expresan otros, si expresa ideas nuevas, si realiza preguntas, si busca y proporciona material para ampliar un tema.

Es necesario incorporar análisis temporal de forma de proporcionar información sobre la continuidad tanto en la dimensión participativa como interactiva. Se trata de proporcionar datos sobre el estudiante en tiempo real, para facilitar el seguimiento y la intervención en el momento adecuado. El hecho de proporcionar al docente herramientas de análisis de información al comienzo, durante y al final de los procesos de aprendizaje facilita una mejor comprensión del proceso en sí.

7 Referencias

[1] R. Vivian, Barnes, Alan, Geer, Ruth y Wood, Denise, «The academic journey of university students on Facebook: an analysis of informal academic-related activity over a semester,» *Research in Learning Technology*, 2014.

[2] Said, Nihra, Forret, Mike y Eames, Chris, «An Online Group for Collaborative Learning in Malaysian Pre-Service Teachers,» de *3rd International Conference of Education, Research and Innovation*, Madrid, Spain, 2010.

[3] S. Ouf, M. Abd Ellatif, S. E. Salama y Y. Helmy, «A proposed paradigm for smart learning environment based on semantic web,» *Computers in Human Behavior*, vol. 72, pp. 796-818, 2017.

[4] F. Henri, «Computer Conferencing and Content Analysis,» de *Collaborative Learning Through Computer Conferencing*, 1992, pp. 117-136.

[5] F. Pozzi, S. Manca, D. Persico y L. Sarti, «A general framework for tracking and analysing learning processes in computer-supported collaborative learning environments,» *Innovations in Education and Teaching International*, 2007.

[6] Facebook, «Facebook, «Información general sobre la API Graph,» [En línea]. Available: https://developers.facebook.com/docs/graph-api/overview/. [Último acceso: 12 3 2018].

[7] M. Schroepfer, «An Update on Our Plans to Restrict Data Access on Facebook,» Facebook newsroom, 04 04 2018. [En línea]. Available: https://newsroom.fb.com/news/2018/04/restricting-data-access/. [Último acceso: 29 04 2018].

[8] M. N. Said y L. M. Tahir, «Towards Identification of Students' Holistic Learning Process through Facebook in Higher Education,» *Procedia - Social and Behavioral Sciences*, vol. 97, pp. 307-313, 2013.

[9] O. Cervantes, E. Gutierrez, F. Gutiérrez y J. A. Sánchez , «Social metrics applied to smart tourism,» de *1st International Conference on Smart Data and Smart Cities*, Split, Croatia, 2016.

[10] R. Motz y O. Cervantes, «No. FSED_2_2016_130712 , Proyecto ANII - DIIA: "Descubrimiento de Interacciones que Impactan el Aprendizaje",» Montevideo, Uruguay, 2016.

[11] L. Boykova, «Student interactions on Facebook: how Facebook student groups can enrich student's university experience,» *Raziskave in Razprave*, vol. 8, nº 2, pp. 4-39, 2015.

[12] B. Rieder, «Studying Facebook via data extraction,» de *5th Annual ACM Web Science Conference on - WebSci '13*, 2013.

[13] D. Yang, Sinha, Tanmay, Adamson, David y Rose, Carolyn Penstein, «urn on, Tune in, Drop out": Anticipating student dropouts in Massive Open Online Courses,» de *Conference: NIPS Workshop on Data Driven Education*, 2013.

Diseño, desarrollo, publicación y evaluación de Objetos de Aprendizaje. Un estudio de casos.

María Lucía Violini[1,2], Cecilia Verónica Sanz[1,3]

[1] Instituto de Investigación en Informática LIDI,
Facultad de Informática, Universidad Nacional de La Plata.
[2] Becaria TIPO A de la Universidad Nacional de La Plata.
[3] Investigadora Asociada de la Comisión de
Investigaciones Científicas de la Provincia de Buenos Aires.
La Plata, Buenos Aires, Argentina.
{lviolini, csanz}@lidi.info.unlp.edu.ar

Resumen. En este artículo se presenta un estudio de casos sobre creación de Objetos de Aprendizaje (OAs) y su uso en contextos educativos concretos. Primeramente, se realiza una revisión que revela las principales motivaciones de los docentes para crear OAs e incorporarlos en sus propuestas educativas. Luego, se realiza un análisis sobre un sub-conjunto seleccionado de casos de interés. Éste arroja resultados que permiten reflexionar sobre: metodologías de diseño de OAs (tecnológicas, pedagógicas, híbridas), tipos de escenarios de diseño (audiovisuales, interactivos, mixtos), herramientas de desarrollo, entornos de publicación y procesos de evaluación. El estudio realizado deja entrever un panorama positivo, y en progreso, en cuanto a la creación de OAs y su incorporación/uso en ámbitos educativos.

Palabras clave: Objetos de Aprendizaje, Materiales Educativos Digitales, Estudio de Casos.

1 Introducción

Con el avance de la web 2.0, los docentes comenzaron a involucrarse en la producción de sus propios *Materiales Educativos Digitales* (MEDs) [1]. Hace ya bastante tiempo que los MEDs se han convertido en un pilar de las propuestas educativas mediadas por *Tecnologías de la Información y la Comunicación* (TICs). En este escenario, se comienza a abordar el tema *Objetos de Aprendizaje* (OAs) [2].

Al momento, no existe una definición consensuada para el término *Objeto de Aprendizaje*. La falta de claridad conceptual es evidente en la cantidad de definiciones y usos que se hacen de los OAs [3]. Se adopta, para este trabajo, el siguiente concepto: "*Un OA es un tipo de material educativo digital, que se caracteriza, desde el punto de vista pedagógico, por orientarse a un objetivo específico de aprendizaje, y presentar mínimamente: una serie de contenidos con el fin de abordar la temática relacionada con el objetivo, actividades que permitan al alumno poner en práctica o problematizar el contenido presentado, y una autoevaluación que posibilite conocer*

al alumno, si ha podido comprender esos contenidos vinculados al objetivo. Desde el punto de vista tecnológico, se caracteriza por contener un conjunto de metadatos estandarizados para su búsqueda, y recuperación, y estar integrado, utilizando un modelo de empaquetamiento que respete estándares, y de esta manera, permita su diálogo con diferentes entornos tecnológicos." [4].

Los OAs son MEDs que pueden ser usados para facilitar un aprendizaje concreto, a partir del objetivo específico que proponen. Como así también, pueden ser integrados en lecciones, unidades o itinerarios de aprendizaje, bajo ciertas consideraciones y estableciendo entre ellos una relación determinada, lo que permitirá logros que cada uno por sí solo no sería posible de alcanzar [5]. Los OAs pueden ser incluidos tanto en propuestas de educación presencial, para su uso dentro del aula, como en propuestas de educación a distancia, a través de su publicación en *Repositorios de Objetos de Aprendizaje* (ROAs) o en *Entornos Virtuales de Enseñanza y Aprendizaje* (EVEAs) [6].

La decisión de integrar OAs, en propuestas educativas, lleva a los docentes a adentrarse en procesos de creación. Para esto, existen metodologías que tienen como finalidad guiar en las tareas involucradas en el diseño de OAs, el desarrollo y la publicación [7]. A su vez, existen numerosas herramientas que posibilitan la creación de estos MEDs, algunas con propósitos más específicos que otras [8].

El trabajo que aquí se presenta tiene como eje el estudio de un sub-conjunto seleccionado de casos, en los cuales se plantea la creación de OAs y su posterior uso en contextos educativos concretos. Se realiza un análisis que permite reflexionar sobre: las principales etapas por las que atraviesa un OA, las motivaciones que los docentes consideran para involucrarse en su creación y ciertos aspectos que revelan problemáticas en el estado del arte de la temática.

De aquí en adelante, este artículo se organiza de la siguiente manera. En la sección 2, se realiza una revisión de antecedentes en relación a las principales motivaciones de los docentes al momento de crear OAs e incorporarlos en sus propuestas educativas. En la sección 3, se presenta el análisis realizado sobre un sub-conjunto seleccionado de casos, basado en una serie de preguntas de investigación y criterios determinados. En la sección 4, se presenta una discusión sobre cuestiones clave de creación y uso de OAs, que se fundamenta en los resultados obtenidos a partir del análisis. Finalmente, en la sección 5, se detallan las conclusiones a las cuales se ha arribado y se proponen líneas de trabajo futuro.

2 Antecedentes en relación a las motivaciones para crear y usar Objetos de Aprendizaje

En esta sección, se describen algunos antecedentes indagados en relación a las motivaciones que involucran a los docentes en la creación y el uso de OAs. Cada docente tiene sus propios motivos por los cuales decide crear OAs e incorporarlos en sus propuestas educativas.

En [9-18] se hace referencia a que los docentes pretenden aprovechar las potencialidades de los lenguajes de los MEDs, bajo la consideración de que los OAs son MEDs. En éstos no sólo se presenta información textual sino que integran

recursos: visuales (imágenes fijas y en movimiento), auditivos, interactivos, entre otros.

En [10] [12-15] [17] se hace referencia a que los docentes consideran a los OAs como MEDs motivadores del aprendizaje. Se cree que el OA, desde su carácter innovador y con posibilidades de interactividad, despertará el interés de los alumnos para el abordaje de la temática en cuestión.

En [12-13] [16-18] se hace referencia a que los docentes buscan fomentar la participación activa del alumno en su propio proceso de aprendizaje. Un OA cuenta con un objetivo específico de aprendizaje y reúne: contenidos sobre el tema a tratar, actividades para el alumno en relación a esos contenidos y una autoevaluación donde el alumno pone a prueba sus conocimientos, así el alumno puede trabajar de manera autónoma con MED auto-contenido que promueve su compromiso por aprender.

En [10-12] [17] se hace referencia a que los docentes crean OAs con la intención de generar MEDs que queden disponibles para otros docentes y puedan ser reutilizados en otros contextos educativos. La reutilización es una característica clave y distintiva de los OAs respecto a otros tipos de MEDs [19]. Para alcanzar esta característica, los OAs cuentan con metadatos que son la ficha de catalogación digital e indican: qué contiene el OA, para quiénes fue elaborado, cuándo fue creado, sus autores, los requerimientos y recomendaciones técnicas y educativas de uso, los derechos de autor y si existen restricciones de uso o reproducción [20].

En la próxima sección se presenta el núcleo de este trabajo, vinculado al estudio de casos sobre creación de OAs y su uso en contextos educativos concretos.

3 Estudio de casos

En esta sección, se presenta el estudio de casos realizado para responder a una serie de preguntas de investigación vinculadas a creación y uso de OAs. En primer lugar, se formulan las preguntas de investigación. Luego, se describe el proceso de selección de casos llevado a cabo. Posteriormente, se definen criterios de análisis. Por último, se exponen los resultados del análisis de casos de interés.

3.1 Preguntas de investigación

A partir de un trabajo sostenido en la temática [2] [4] [7-8], se han propuesto una serie de preguntas de investigación para guiar el estudio sobre OAs y echar luz sobre diferentes aspectos en los cuales se requiere profundizar y trabajar. A continuación, se presentan las preguntas formuladas que guían el estudio de casos.

P1. ¿Qué motivaciones involucran, a los docentes, en la creación y el uso de OAs?
P2. ¿Qué metodologías utilizan, los docentes, para el diseño de OAs?
P3. ¿Qué tipos de escenarios de diseño utilizan los docentes?
P4. ¿Qué herramientas utilizan, los docentes, en las tareas de desarrollo de OAs?
P5. ¿Qué entornos tecnológicos utilizan, los docentes, para publicar OAs?
P6. ¿Qué procesos de evaluación aplican, los docentes, sobre los OAs?

3.2 Selección de casos

Inicialmente, se realizó un recorrido de la bibliografía en busca de casos que planteen la creación de OAs y su posterior uso en contextos educativos concretos. Como rango de búsqueda, se decidió considerar casos entre los años 2000 y 2018. Con respecto al idioma, se consideraron trabajos escritos en inglés o español. La búsqueda estuvo focalizada en las áreas *Tecnología* y *Educación*, y se llevó a cabo en: *ELSEVIER*[1], *IEEE Xplore Digital Library*[2], *ACM Digital Library*[3]; en las siguientes revistas científicas: *IEEE Journal of Latin-American Learning Technologies (IEEE-RITA)*, *Interdisciplinary Journal of Knowledge and Learning Objects (IJKLO)*; además se consideraron las siguientes conferencias internacionales: *Latin American Conference on Learning Technologies (LACLO)*, *International Conference on Engineering Education (ICEED)*, *International Conference on Advances in Mobile Computing & Multimedia (MoMM)*, *International Conference on Advanced Learning Technologies (ICALT)*, *International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM)*, *Frontiers in Education Conference (FIE)*. La cadena de búsqueda fue: *Learning Object*/*Objeto de Aprendizaje*. Luego de una primera revisión de los trabajos, a partir de sus títulos y resúmenes para determinar si se vinculaban con los fines de estudio, se seleccionaron 40 trabajos para lectura completa. Después de esto, se realizó una nueva selección en base a las siguientes prioridades.

En primer lugar, se decidió priorizar aquellos casos en los cuales el concepto de OA se acercara lo más posible a la definición de OA adoptada en este artículo. Así, se estableció, como primer filtro de selección, que el OA en cuestión debería: (1) ser un MED con un objetivo específico de aprendizaje; (2) presentar contenidos sobre la temática en cuestión; (3) proponer, al menos, una actividad/autoevaluación que permita al alumno poner en juego/autoevaluar los contenidos presentados.

En segundo lugar, se decidió contemplar aquellos casos que estuvieran explicados con mayor nivel de detalle, donde fuera posible analizar la mayoría o, en el mejor de los casos, todas las cuestiones implicadas en las preguntas de investigación que se formularon. Entonces, como segundo filtro de selección, se estableció que la descripción del caso debería referirse a: motivaciones de los docentes y diseño, desarrollo, publicación, evaluación del OA.

Finalmente, se obtuvo un sub-conjunto de 10 casos de interés para realizar el análisis. En la Tabla 1 se muestran los casos seleccionados, para cada uno se especifica: año de publicación del trabajo, autores y área disciplinar vinculada.

Tabla 1. Casos seleccionados: especificaciones.

CASOS SELECCIONADOS			
Nro. de caso	**Año de publicación**	**Autores**	**Área disciplinar**
1	2003	Chalk et al.	*Informática - Programación*
2	2005	Krauss y Ally	*Farmacología - Farmacocinética*
3	2006	Reis et al.	*Informática - Inteligencia Computacional*
4	2009	Ummi et al.	*Ingeniería - Electrónica*

[1] *ELSEVIER*: https://www.elsevier.com/.
[2] *IEEE Xplore Digital Library*: https://ieeexplore.ieee.org/.
[3] *ACM Digital Library*: https://dl.acm.org/.

5	2009	Matthews et al.	*Informática - Programación*
6	2012	Tuparov et al.	*Informática - Programación*
7	2014	Orozco et al.	*Matemática - Geometría*
8	2015	Tiosso et al.	*Arquitectura de Computadoras - Memoria*
9	2017	Violini y Sanz	*Informática - Programación*
10	2018	Redmond et al.	*Enfermería - Cuidado de heridas*

3.3 Criterios de análisis

Con intención de atender a las preguntas de investigación formuladas, se establecieron criterios de análisis. En la Tabla 2 se definen estos criterios, para cada uno de ellos se indica: con qué pregunta de investigación se relaciona, su descripción y los propósitos que persigue.

Tabla 2. Criterios de análisis: descripción y propósitos.

C1. Motivaciones | Relacionado a *P1*
Motivaciones que han llevado a los docentes a crear e incorporar el OA. Podrían ser: (a) fomentar aprendizaje activo y autónomo; (b) motivar el aprendizaje; (c) integrar información visual/interactiva; (d) generar material educativo reutilizable.
✓ Detectar las principales motivaciones para crear y usar OAs. ✓ Notar cuán motivadora resulta ser la característica de reutilización que poseen los OAs.
C2. Metodología de diseño | Relacionado a *P2*
Metodología que ha sido utilizada para diseñar el OA, si es que se siguió una metodología de diseño de OAs en particular. Se distinguen las metodologías de acuerdo a su enfoque: tecnológico (predominan aspectos tecnológicos, se vincula al desarrollo de *software*), pedagógico (predominan aspectos pedagógicos, se vincula al diseño instruccional) o híbrido (hay equilibrio entre aspectos tecnológicos y pedagógicos) [7].
✓ Detectar la utilización de metodologías que guíen el diseño de OAs. ✓ Notar qué tipo de enfoque es utilizado al momento de diseñar. ✓ Reconocer cómo repercuten estas cuestiones en la creación de OAs.
C3. Tipos de escenarios de diseño | Relacionado a *P3*
Tipos de escenarios que incluye el OA en las distintas partes que lo componen. Se consideran escenarios: audiovisuales (incluyen: audios, imágenes, videos), interactivos (permiten la interacción del alumno con los elementos que aparecen en la pantalla) y mixtos.
✓ Observar la incorporación de escenarios de cada tipo. ✓ Detectar si se apela a combinar ambos tipos de escenarios. ✓ Reconocer de qué manera se aprovechan las ventajas de cada uno.
C4. Herramientas de desarrollo | Relacionado a *P4*
Herramientas de autor utilizadas para implementar el OA. Se contemplan herramientas para: armar el OA, generar contenidos, generar actividades/autoevaluaciones, cargar metadatos [8].
✓ Observar la utilización de herramientas en las distintas tareas de desarrollo. ✓ Detectar la utilización de herramientas que integren metodologías de diseño. ✓ Reconocer el impacto de la utilización de herramientas en el quehacer docente.
C5. Entornos de publicación | Relacionado a *P5*
Entornos tecnológicos dónde ha sido, o será, publicado el OA. Podrían ser: ROAs o EVEAs.
✓ Observar la publicación de los OAs en entornos interoperables. ✓ Notar qué clase de entorno se utiliza para publicar OAs. ✓ Reconocer los beneficios que brinda la publicación de OAs en estos entornos.

C6. Procesos de evaluación | Relacionado a *P6*
Procesos de evaluación a los cuales ha sido sometido el OA, si es que ha sido evaluado de alguna manera, o a los procesos que se prevé aplicar. Esto incluye pruebas: con los destinatarios para los cuales fue creado el OA y con expertos (docentes/investigadores idóneos en la temática que aborda el OA).
✓ Detectar si los OAs son sometidos a procesos de evaluación. ✓ Observar la realización de pruebas en las que intervienen alumnos y docentes. ✓ Reconocer si los resultados de las pruebas influyen en mejoras de los OAs.

3.4 Análisis de casos de interés

De acuerdo a los criterios establecidos, se realizó un análisis de los casos de interés. En la Tabla 3 se detallan los resultados que se obtuvieron para cada criterio.

Tabla 3. Resultados del análisis de casos seleccionados.

CRITERIOS DE ANÁLISIS	**RESULTADOS OBTENIDOS A PARTIR DE LOS CASOS ANALIZADOS**
1 Motivaciones	En la Tabla 4 se especifican las motivaciones a las cuales se refiere cada caso.
2 Metodología de diseño	70% (casos 1-2, 4-5, 7, 9-10) → se refiere a metodologías, teorías y/o principios aplicados en el diseño de los OAs. Utilización de metodologías de diseño de OAs, como: la *Metodología CROA*[4] de enfoque híbrido (caso 9). Consideración de teorías del aprendizaje como marco para el diseño (casos 2, 4-5, 7, 9-10), como: las teorías constructivistas del aprendizaje (caso 10), la *Teoría del Aprendizaje Significativo* (caso 9), la *Teoría Cognitiva del Aprendizaje Multimedia* (casos 5, 10), la *Teoría del Diseño Instruccional* (caso 5). Aplicación de principios de diseño de OAs basados en *Ingeniería de Software* (caso 1).
	30% restante → no hace referencias a metodologías, teorías o principios aplicados en el diseño de los OAs.
3 Tipos de escenarios de diseño	100% (casos 1-10) → se refiere a la incorporación de escenarios audiovisuales, que incluyen: gráficos, imágenes, animaciones, videos. Éstos fueron utilizados, principalmente, para presentar contenidos dentro de los OAs.
	100% (casos 1-10) → se refiere a la incorporación de escenarios interactivos, que permiten: resolver ejercicios, recibir retroalimentaciones, controlar soluciones. Éstos fueron utilizados para las actividades/autoevaluaciones de los OAs.
4 Herramientas de desarrollo	60% (casos 2-4, 6-7, 9) → se refiere a la utilización de herramientas de desarrollo, como: *Ardora*[5] para generación de actividades (caso 9), *eXeLearning*[6] para armado de los OAs y carga de metadatos (casos 7, 9), *software* multimedia de autoría (caso 4).
	40% restante → no hace referencias a herramientas utilizadas para el desarrollo de los OAs.

[4] *Metodología CROA*: http://croa.info.unlp.edu.ar/.
[5] *Ardora*: http://webardora.net/.
[6] *eXeLearning*: http://exelearning.net/.

5 *Entornos de publicación*	80% (casos 1, 3-4, 6-10) → se refiere a la publicación de los OAs. Ya sea, en EVEAs que respetan el estándar de empaquetamiento de dichos OAs (casos 1, 6-7, 9), como: *Moodle*[7] (casos 6-7). O, en ROAs (casos 3-4, 8), como: *MERLOT*[8] (caso 8). Por otro lado, se hace referencia a la disponibilidad de los OAs en la *web* por un período de tiempo determinado (caso 10), pero no a su publicación en EVEAs/ROAs.
	20% restante → no hace referencias a la publicación de los OAs.
6 *Procesos de evaluación*	80% (casos 1-2, 5-10) → se refiere a pruebas de los OAs con alumnos, como: encuestas (casos 2, 6), cuestionarios previos y posteriores al uso de los OAs (casos 5, 10), evaluaciones en línea (casos 7-8, 10), pruebas de usabilidad (caso 2), sesiones de reflexión en voz alta (caso 2), entrevistas (casos 1, 5). La finalidad es obtener opiniones tanto a nivel técnico como didáctico y, además, detectar posibles problemas de uso, por ejemplo: de navegación (caso 2).
	20% restante → no hace referencias a pruebas de los OAs con participación de alumnos.
	50% (casos 2, 4, 7, 9-10) → se refiere a pruebas de los OAs con expertos/docentes, como: sesiones de muestra (caso 9), encuestas (casos 2, 9), revisiones (casos 2, 4, 10), evaluaciones con pares (caso 9). La finalidad es obtener valoraciones acerca de los contenidos y, también, sobre el diseño de los OAs, para contar con retroalimentaciones de mejora provenientes de idóneos.
	50% restante → no hace referencias a pruebas de los OAs con participación de expertos/docentes.

Tabla 4. Motivaciones de los casos seleccionados.

	MOTIVACIONES			
	a) Fomentar aprendizaje activo y autónomo.	**b) Motivar el aprendizaje.**	**c) Integrar información visual / interactiva.**	**d) Generar material educativo reutilizable.**
CASO 1			✓	
CASO 2		✓	✓	✓
CASO 3			✓	✓
CASO 4	✓	✓	✓	✓
CASO 5	✓	✓	✓	
CASO 6		✓	✓	
CASO 7		✓	✓	
CASO 8	✓		✓	
CASO 9	✓	✓	✓	✓
CASO 10	✓		✓	

[7] *Moodle*: https://moodle.org/.
[8] *MERLOT*: https://www.merlot.org/.

4 Discusión

Los resultados obtenidos, a partir del estudio de casos realizado, permiten reflexionar sobre cuestiones clave de la creación y el uso de OAs. A continuación, se presenta una discusión sobre cada pregunta de investigación, sin perder de vista los propósitos de los criterios de análisis relacionados.

- Sobre *P1* y *C1*: *Motivaciones*. Si bien las motivaciones por las cuales los docentes crean OAs y los incorporan en sus propuestas educativas son propias y particulares de cada caso, al revisarlas en conjunto se encuentran cuestiones en común que pueden sintetizarse de la siguiente manera: (a) se pretende fomentar la participación activa, y autónoma, del alumno en su proceso de aprendizaje; (b) se pretende motivar el aprendizaje del alumno, ofreciéndole un MED innovador y dinámico; (c) se pretende brindar, al alumno, un MED que apele a lo visual e interactivo para aprovechar los beneficios de los distintos lenguajes; (d) se pretende generar un MED reutilizable, por otros docentes, en otros contextos educativos. La reutilización, una característica fundamental de los OAs, no ha resultado ser la más motivadora para los docentes.
- Sobre *P2* y *C2*: *Metodología de diseño*. Se detecta una escasa utilización de metodologías de diseño de OAs propiamente dichas, no obstante, se manifiesta una amplia consideración de teorías del aprendizaje en el diseño de los OAs. Entre los distintos enfoques, el híbrido resultó ser utilizado al momento de diseñar. Las metodologías de diseño, en particular las híbridas, facilitan a los docentes la creación de OAs, sin dejar de lado aspectos tecnológicos ni pedagógicos.
- Sobre *P3* y *C3*: *Tipos de escenarios de diseño*. En los OAs, se incorporan escenarios tanto audiovisuales como interactivos. En general, se apela a combinar ambos tipos de escenarios. Se aprovecha la manera en que los escenarios audiovisuales permiten presentar contenidos, a través del audio, la imagen y el video. A su vez, se aprovechan las posibilidades que ofrecen los escenarios interactivos para proponer ejercicios que el alumno puede resolver, controlar y recibir retroalimentaciones.
- Sobre *P4* y *C4*: *Herramientas de desarrollo*. Se utilizan herramientas para las distintas tareas de desarrollo de OAs, pero no se detecta el uso de herramientas que integren metodologías de diseño (esto refuerza conclusiones previas de los autores de este trabajo [7] [8]). Disponer de herramientas, para implementar las distintas partes de un OA, es una gran oportunidad para los docentes interesados en crear OAs. Sin embargo, la utilización de cada herramienta implica para el docente saber cuestiones básicas, y a veces más complejas, sobre su manejo; si estas herramientas asumen un nivel avanzado de usuario, podría dificultarse el quehacer de docentes poco familiarizados con la tecnología.
- Sobre *P5* y *C5*: *Entornos de publicación*. La publicación de OAs se produce en entornos tecnológicos que respetan estándares de empaquetamiento. Los OAs se publican en EVEAs, principalmente, cuando van a ser usados en contextos educativos específicos. En cambio, cuando se pretende dejar los OAs disponibles para la comunidad educativa en general, éstos suelen ser

publicados en ROAs. La publicación de OAs en entornos interoperables posibilita su localización y acceso, y favorece la reutilización.

- Sobre *P6* y *C6*: *Procesos de evaluación.* En general, los OAs son sometidos a distintos procesos de evaluación. Se realizan pruebas con alumnos y, también, con docentes. Estas pruebas buscan obtener resultados/opiniones/valoraciones que permitan alcanzar mejoras de los OAs. Vale destacar que, en la mayoría de los casos analizados, tanto docentes como alumnos valoran positivamente el uso de OAs en los contextos educativos.

5 Conclusiones y Trabajo futuro

Luego de la investigación realizada, habiendo hecho una revisión acerca de las principales motivaciones de los docentes y un estudio de casos sobre creación y uso de OAs, se presentan a continuación conclusiones a las cuales se ha arribado.

Por un lado, se detectaron motivaciones comunes de los docentes al momento de crear OAs e incorporarlos en sus propuestas educativas: fomentar aprendizaje activo y autónomo, motivar el aprendizaje, integrar información visual e interactiva, generar material educativo reutilizable. Sin embargo, a pesar de que la reutilización es una de las características distintivas de los OAs, no ha sido ésta la motivación principal detectada en los casos revisados. Además, sólo un 30% de los casos hace referencia a la publicación de los OAs en ROAs. Esto es un indicador para seguir indagando y marca una posible causa de la falta de reutilización de OAs por parte de los docentes.

Por otro lado, con respecto a las distintas etapas de los OAs, se detectó: una escasa utilización de metodologías de diseño de OAs, una gran integración de escenarios audiovisuales e interactivos considerados en el diseño de los OAs, la utilización de herramientas (como: *eXeLearning* y *Ardora*) en las distintas tareas de desarrollo de OAs, la publicación de OAs en EVEAs (como: *Moodle*) y ROAs (como: *MERLOT*) que respetan estándares de empaquetamiento, la realización de evaluaciones de los OAs (previas, durante y posteriores a su uso).

Se vislumbra un panorama positivo, y en progreso, en cuanto a la creación de OAs y su incorporación/uso en ámbitos educativos.

Como trabajo futuro, se pretende profundizar esta investigación, considerando otras fuentes, para poder obtener un estado del arte más amplio. Asimismo, se pretende continuar con el desarrollo de *MarCOA*, un *framework* propuesto para la creación de OAs [7] que se basa en la *Metodología CROA* [4]. Se busca atender a la necesidad encontrada de abordar metodologías híbridas en el proceso de creación y considerar, como parte del diseño, la característica fundamental de reutilización.

Referencias

1. Del Moral, M., Villalustre, L.: Didáctica universitaria en la era 2.0: competencias docentes en campus virtuales. RUSC. Universities and Knowledge Society Journal, vol. 9, no. 1 (2012)
2. Sanz, C.: Los objetos de aprendizaje, un debate abierto y necesario. Bit & Byte (2015)

3. Polsani, P.: Use and abuse of reusable learning objects. Journal of Digital information, vol. 3, no. 4 (2006)
4. Sanz, C., Barranquero, F., Moralejo, L.: CROA: A learning object design and creation methodology to bridge the gap between educators and reusable educational material creation. EDULEARN16, pp. 4583--4592 (2016)
5. Prendes, M., Martínez, F., Gutiérrez, I.: Producción de material didáctico: los objetos de aprendizaje. Revista Iberoamericana de Educación a Distancia (RIED), vol. 11, no. 1 (2008)
6. Mora-Vicarioli, F.: Objetos de aprendizaje: Importancia de su uso en la educación virtual. Revista Electrónica Calidad en la Educación Superior, vol. 3, no 1, pp. 104--118 (2012)
7. Violini, L., Sanz, C., Pesado, P.: Propuesta de un Framework para la creación de Objetos de Aprendizaje. XXIII Congreso Argentino de Ciencias de la Computación (2017)
8. Violini, L., Sanz, C.: Herramientas de Autor para la creación de Objetos de Aprendizaje. ESTADO DEL ARTE. XXII Congreso Argentino de Ciencias de la Computación (2016)
9. Chalk, P., Bradley, C., Pickard, P.: Designing and evaluating learning objects for introductory programming education. ACM SIGCSE Bulletin, pp. 240--240 (2003)
10. Krauss, F., Ally, M.: A Study of the Design and Evaluation of a Learning Object and Implications for Content Development. Interdisciplinary Journal of Knowledge and Learning Objects (IJKLO), vol. 1 (2005)
11. Reis, F., Félix, Í, Stanzani, S., Campello, R., de Castro, L., Senger, H., Rosatelli, M.: A Learning Object on Computational Intelligence. Sixth International Conference on Advanced Learning Technologies (ICALT), IEEE, pp. 33-33 (2006)
12. Ummi, H., Aida, K., Azrita, A., Elia, H.: Development of learning object for engineering courses in UTeM. International Conference on Engineering Education (ICEED), IEEE, pp. 19--195 (2009)
13. Matthews, R., Hin, H., Choo, K.: Multimedia learning object to build cognitive understanding in learning introductory programming. 7th international Conference on Advances in Mobile Computing and Multimedia (MoMM), ACM, pp. 396--400 (2009)
14. Tuparov, G., Tuparova, D., Tsarnakova, A.: Using interactive simulation-based learning objects in introductory course of programming. Procedia-Social and Behavioral Sciences, vol. 46, pp. 2276--2280 (2012)
15. Orozco, C., Morales, E.: The eXeLearning and GeoGebra integration for teaching geometrics definitions and vectors representations through learning objects. Second International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM), ACM, pp. 639--645 (2014)
16. Tiosso, F., Bruschi, S., Souza, P., Barbosa, E., de Andrade, C.: Amnesia: A learning object for memory hierarchy teaching. Frontiers in Education Conference (FIE), IEEE, pp. 1--7 (2015)
17. Violini, L., Sanz, C.: Learning objects: How to insert an element into a vector?. Twelfth Latin American Conference on Learning Technologies (LACLO), IEEE, pp. 1--4 (2017)
18. Redmond, C., Daviesa, C., Cornallyb, D., Adama, E., Dalya, O., Feganc, M., O'Toolec, M.: Using reusable learning objects (RLOs) in wound care education: Undergraduate student nurse's evaluation of their learning gain. Nurse education today, vol. 60, pp. 3--10 (2018)
19. Sicilia, M.; García, E.: On the concepts of usability and reusability of learning objects. The International Review of Research in Open and Distributed Learning, vol. 4, no .2 (2003)
20. De La Torre, L.; Dominguez, J.: Las TIC en el proceso de enseñanza aprendizaje a través de los objetos de aprendizaje. Revista Cubana de Informática Médica, vol. 4, no. 1, pp. 83--92 (2012).

Análisis objetual de la interfaz gráfica de usuario de una plataforma de EaD. El caso MIeL (Materias Interactivas en Línea).

De Cicco, Juan[1], Morrone, Leandro[1]
Orthusteguy, Fernando[1], Rusticcini, Alejandro[1].

[1]Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigaciones Tecnológicas.
jdecicco@unlam.edu.ar, lmorrone@unlam.edu.ar, forthus@ unlam.edu.ar,
arusticcini@ unlam.edu.ar.

Resumen: El uso intensivo de TIC´s como mediadoras en el proceso educativo implica la utilización de lenguajes e interacciones que trascienden el tradicional docente-alumno, tornándose el medio en un actor más del proceso, por lo tanto un factor a diseñar a la hora de plantear un sistema educativo de estas características. Las interfaces gráficas de usuario, a partir de sus metáforas, se constituyen en el factor fundamental para lograr de forma eficaz este proceso comunicacional. El presente trabajo pretende indagar acerca de las posibilidades que surgen a partir de la incorporación de aquellos - y otros - elementos comunicacionales no verbales con el fin de mejorar la comprensión en la interfaz de EaD de la plataforma MIeL (Materias Interactivas en Línea).

Palabras clave: Interfaz Gráfica de Usuario, Interacción, Educación a Distancia (EaD), Metáfora Conversacional.

1 En busca de un paradigma de interacción.

Desde el comienzo de los sistemas informáticos, aún antes que se plasmara en una aplicación concreta, varios teóricos e investigadores trataron de establecer un paradigma que le permitiera a la gran mayoría de las personas (especialmente a aquellos que no tuvieran conocimientos técnicos) lograr una rápida y eficaz interacción con los sistemas informáticos.

Luego de varias teorías, diferentes contribuciones y ensayos a través de casi cuatro décadas; sería en el PARC (Palo Alto Research Center), creado por Xerox Corporation en 1970, el lugar donde naciera la primera computadora que incluyera una interfaz gráfica de usuario ("*Grafic User Interface*", GUI). Se trató de la Xerox Alto (1973), diseñada por un equipo de trabajo integrado por Ed McCreight, Chuck Thacker, Butler Lampson, Bob Sproull y Dave Boggs.

La Xerox Alto fue desarrollada con la intención de concebir una computadora lo suficientemente pequeña y transportable como para que pudiera ser ubicada en una oficina, además debería contar con un sistema operativo que se controlara desde una interfaz gráfica y que pudiera compartir información de manera sencilla.

Esta primera interfaz carecía de íconos pero tenía una especie de botones de texto, los cuales podían ser accionados a través del uso de un rudimentario mouse.

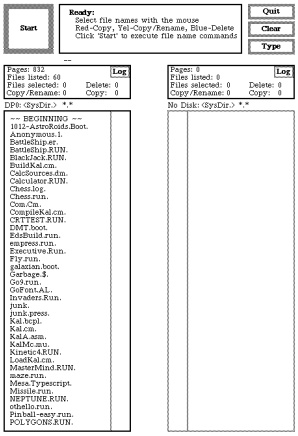


Fig. 1: Vista de la Interfaz de la Xerox Alto

La interfaz de la Xerox Alto constituye seguramente el primer acercamiento a la "metáfora del escritorio", paradigma de interacción conocido como WIMP (*Windows*, *Icons*, *Menus and Point devices*), que se mantiene vigente hasta nuestros días.

La misma se vería plasmada en la Xerox Star 8010 (1981), sucesora de la Xerox Alto, concebida en el mismo centro de investigación por un equipo de doscientos desarrolladores al mando de Don Massaro.

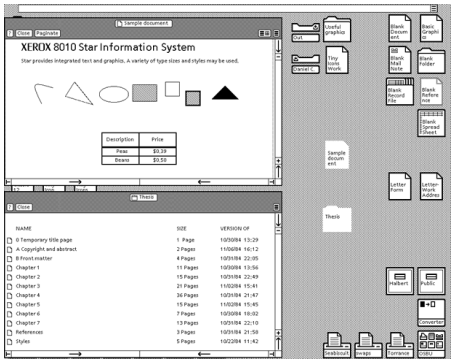


Fig. 2: Vista de la Interfaz de la Xerox Star 8010

Uno de los principales objetivos del proyecto consistía en establecer el concepto de oficina virtual, metáfora esta (conocida posteriormente como "metáfora del escritorio") que le permitiera al usuario interpretar rápida y eficazmente las tareas que podía desempeñar con la computadora al relacionarlas con las de una oficina real. Al respecto sostiene Ben Schneiderman (reconocido investigador especializado en la interacción hombre computadora) que "el modo en que la gente usa objetos del mundo real puede ser transferido a la manipulación de los objetos virtuales en los entornos digitales". (1998).

Es importante destacar que Xerox Star no fue un producto de utilización masiva ya que su costo de producción era muy elevado y por ende el acceso a ella era restringido; aunque su aporte fue decisivo para el desarrollo de posteriores interfaces gráficas de usuario, que si gozaron de suma popularidad en su momento, tales como las que utilizaron Apple Macintosh (1984) o Windows 95 (1995) por citar sólo algunas.

2 "Conversando" con la interfaz.

Desde un enfoque etimológico podemos aproximarnos al concepto que encierra el término interfaz, compuesto por dos vocablos INTER (del latín *inter*) que significa "en medio" y FAZ (del latín *facies*) que significa "superficie o lado de una cosa", entonces podemos deducir que la interfaz puede ser aquello que se ubica en medio de dos (o más) superficies.

Recurriendo a Wikipedia podemos ampliar bastante más las implicancias del término, y a la vez relacionarlas con el área de nuestra competencia: "... en informática se utiliza para nombrar a la conexión física y funcional entre dos sistemas o dispositivos de cualquier tipo dando una comunicación entre distintos niveles". Asimismo la enciclopedia describe su utilización en diferentes contextos: Interfaz como instrumento, Interfaz como Superficie e Interfaz como Espacio.

Con relación a lo anterior Carlos Scolari, en su libro "*Hacer clic. Hacia una socio semiótica de las interacciones digitales",* plantea además de las metáforas citadas,

una *metáfora conversacional*, en la cual "el proceso de interacción es visto como un proceso de comunicación donde el usuario y aplicativo de la computadora actúan ambos como emisores y receptores, y el aplicativo es visto como capaz de demostrar conductas comunicativas similares a las del *partner* humano". (Kammersgard en Scolari, 2006).

Llegado a este punto, y desde este lugar, en el cual la interfaz podría funcionar a la vez como instrumento, superficie, espacio o interlocutor; planteamos la necesidad de efectuar un análisis objetual de la misma ya que, "el análisis sintáctico de la interfaz nos aleja de la definición conceptual y nos acerca a la realidad objetual de la interfaz, como parte física del sistema informático. Desde esta perspectiva, la interfaz gráfica tiene peso, medidas, localización física, limitaciones tecnológicas y propiedades, que habría que analizar y describir." (Marrero Expósito, 2006).

En virtud de lo antes mencionado circunscribiremos nuestro análisis a la interfaz gráfica de usuario de MIeL (Materias Interactivas en Línea), la plataforma de EaD desarrollada por la Universidad Nacional de la Matanza.

3 A nuevas tecnologías, nuevas GUI´s y por tanto nuevos modos de interacción.

Una GUI hace el manejo de la computadora más intuitivo usando metáforas gráficas. Lo que en la interfaz de línea de comandos se hace escribiendo una orden, en la GUI se hace con una acción. Por ejemplo, si en una interfaz de línea de comandos escribimos el comando “CD” (*Change Directory*, cambiar de Directorio) para ver el contenido de una carpeta o directorio, en una GUI, tan sólo hacemos doble clic sobre el dibujo de la flecha que cambia de carpeta y nos proyectamos hacia otro carpeta, por medio de imágenes.

Si nos detenemos a realizar un análisis que pretenda determinar en qué sentido se mueven las GUI´s, el mismo debe inexorablemente contemplar la evolución de los dispositivos que las soportan y la expansión de sus posibilidades tecnológicas.

Dentro de los cambios producidos recientemente, resulta significativo comprender conceptual y funcionalmente a las “*Live Tiles”*, propuestas en la interfaz gráfica de Windows 8 (conocida inicialmente como *Metro* y luego denominada *Modern UI*) y heredadas de su hermano menor Windows Mobile, ya que establecen, a nuestro entender y desde diferentes perspectivas, un aporte más que interesante en la evolución de las GUI.



Fig. 3: Live Tiles en Windows 8

Inspiradas en la señalética del King County Metro Bus (de allí proviene la denominación primera), que recorre la ciudad de Seattle, tomaron de la misma distintos aspectos estéticos, funcionales y formales tales como la utilización de un sistema de colores contrastantes, combinados en áreas delimitadas por coordenadas cartesianas, tipografías de estilo "*humanist"* de rápida lectura y comprensión (especialmente la Segoe UI), en negro sobre fondo blanco o recortadas sobre color y combinando deliberadamente tamaños, variables tipográficas y posiciones así como la utilización de iconografías de líneas simples y fácil comprensión en general emplazadas en trazo blanco sobre fondo de color.

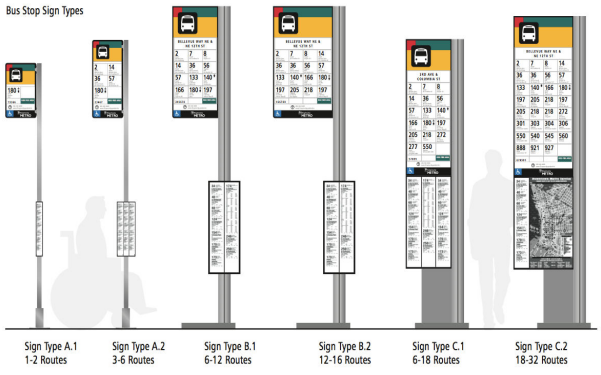


Fig. 4: Señalética del King Country Metro Bus

Este conjunto de elementos significantes, fue adaptado a las necesidades del dispositivo (la pantalla) y enriquecido con las posibilidades técnicas que pone a disposición el medio (la Web).

Entonces, es así que surgen las "*Live Tiles*" ("*tiles*" porque son cuadradas o rectangulares y se distribuyen como azulejos y "*live*" porque permiten actualización permanente mediante notificaciones desde medios remotos - "*push*" - e interacción

con el usuario), las que técnicamente funcionan como accesos directos hacia aplicaciones remotas o instaladas en el mismo dispositivo.

En las primeras GUI, los accesos a las aplicaciones estaban señalizados mediante íconos, los que nos permitían acceder a los contenidos una vez que accionabamos sobre ellos. La utilización de "*Live Tiles"*, les permite a los usuarios, comprender las interfaces más rápidamente y advertir el dinamismo de los contenidos a simple vista, puesto que poseen valor informativo agregado al objeto iconográfico implícito.

"El cerebro es una sofisticada herramienta de reconocimiento de patrones y por ello se desempeña en esta tarea con destreza y eficiencia. Cuanto más cerca de 'miro y entiendo' y más lejos de 'pienso y entiendo' está un contenido más fácil será su comprensión", afirma Daniel Mordecki desde su propia clasificación de las distintas modalidades de interacción por parte del usuario.

Las *Live Tails* se transforman en unidades discursivas independientes -y a la vez relacionadas entre sí-, que comienzan a ostentar una mayor capacidad enunciativa y complejidad de mensaje, generando un sistema de símbolos combinados, de muy sencilla comprensión. Características que a nuestro entender pueden funcionar como sustrato de una nueva sintaxis, que enriquezca sustancialmente a las interfaces de usuario y por consiguiente hagan más satisfactoria y a la vez más efectivas las experiencias de aquellos.

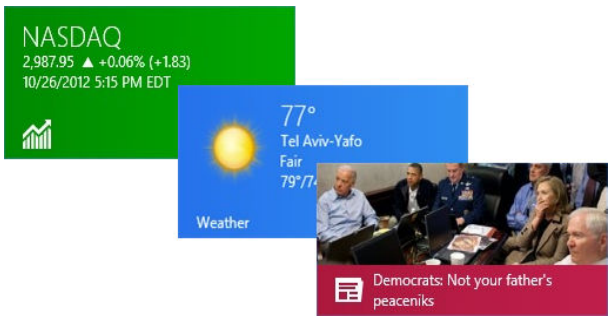


Fig. 5: Live Tiles en Windows 8

Existen actualmente variadas aplicaciones (sean o no desarrolladas por Microsoft), redes sociales y *Websites* que en sus versiones tanto *mobile* como de escritorio, adoptaron estructura, estética y funcionalidad semejantes a *Modern UI* para sus GUI con lo cual podemos inferir que progresivamente irá en aumento el número de usuarios que incorporen como propio este modelo de interacción.

Con la generalización de dispositivos de interfaz táctil (teléfonos celulares, tabletas, *notebooks,* etc.) el cambio mencionado de icono simple a superficie de interacción presenta aún más ventajas, por un lado, por el simple aumento de la dimensión del área de interacción y la consiguiente mejora en la usabilidad (la superficie de control de un dedo humano es mucho menos fina que el puntero del *mouse*), y por el otro, puesto que los usuarios empiezan a entender las interfaces como algo "maleable" y cercano, asociable a objetos reales, manipulables y no a algo que

requiere un dispositivo específico para operar (el *mouse*), se minimiza una posible barrera de acceso.

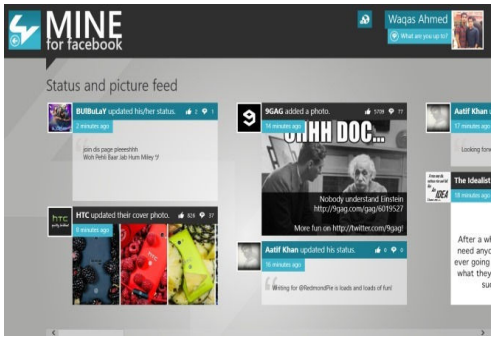


Fig. 6: Website de MINE for Facebook (www.mineforfacebook.com)

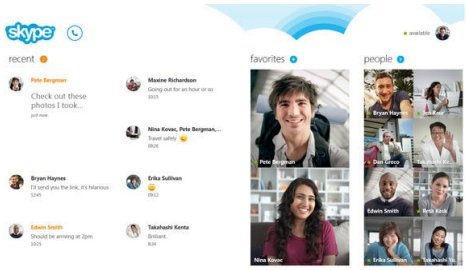


Fig. 7: Website de Skype (www.skype.com)

4 Impacto en las Plataformas de Educación a Distancia, la experiencia MIeL.

En una institución educativa, uno de los principales desafíos por venir, se halla en la integración de actores variados que además estén realizando múltiples intercambios. En un mundo donde la educación es buscada como fuente de valor agregado, por cada vez más individuos a los que, por edad, formación previa u otro tipo de condicionamientos, se los consideraba fuera del circuito formal de la educación. Desde esta lógica, las herramientas tecnológicas, deben funcionar como integradoras de estos actores en el ámbito educativo y no convertirse en verdaderas barreras de acceso.

La utilización de redes sociales, "*blogs*", "*wikis*", foros, entre otras, que funcionan actualmente como eficaz complemento de las plataformas de EaD y que además incorporan elementos innovadores similiares a los explicados en este articuloen sus interfaces, produce un acercamiento por parte de los usuarios (inclusive los menos

tecnológicamente formados) a estas modalidades de interacción y a la vez generan un ambiente propicio, propendiendo a incorporar, de manera eficaz, nuevos conceptos de interacción en las plataformas de EaD.

El desarrollo evolutivo de las interfaces, así como los cambios destinados a mejorar la usabilidad, entendida como la cualidad de ser comprensible e intuitiva, permite disminuir la brecha digital y de esa manera expandir las posibilidades de formación a una cantidad mayor de individuos.

En concordancia con lo expresado en los párrafos anteriores, se consideró desde el proyecto MIeL (plataforma de la UNLaM usada como mediadora tecnológica en los distintos modos de enseñanza), que el trabajo de investigación y desarrollo que pueda plasmarse en la incorporación de nuevas tecnologías en las plataformas de EaD, redundará en beneficio directo de la calidad de la interacción entre sus actores.

Lo dicho se manifiesta en la paulatina evolución del diseño de interacción de la plataforma MIeL, que, basado inicialmente en un concepto claramente iconográfico, incorpora actualmente superficies como espacios de interacción y de información dinámica.

Este permanente proceso de actualización, y adecuación a las prestaciones tecnológicas de los nuevos dispositivos, redunda en la posibilidad de incorporar nuevas herramientas que facilitan y a la vez amplían las capacidades operativas de la plataforma. Estas herramientas, nuevas o la refactorización de las existentes, se incorporan siempre teniendo en cuenta el desarrollo teórico expresado, como ejemplo, la incorporación de espacios dedicados al perfil del usuario, programación cronológica, detalles del curso y centro de alertas, que se observa en la parte superior izquierda de la interfaz (Fig. 8) muestran claramente la dirección seguida.

El cambio en la iconografía más simple se muestra en la limpieza grafica de su diseño y en la incorporación a cada uno de ellos de información adicional, alertas, etc. que le provean al usuario de conocimiento acerca de qué esperar una vez ingresado al espacio mediado por dicho ícono.

De esta manera se busca que los íconos, ya devenidos en espacios de interacción, se conviertan en actores de la interfaz y no simples medios de acceso, reduciendo la distancia discursiva entre usuario y sistema.

Con la finalidad de constatar nuestras apreciaciones preliminares establecimos una encuesta con un grupo de usuarios de la plataforma MIeL, basándonos principalmente en la variedad de orientaciones (variedad de carreras) y en que poseyeran experiencia previa sobre la versión anterior de la plataforma, a quienes se les propuso un cuestionario temático que abarca los aspectos considerados más importantes por el equipo de desarrollo de la plataforma

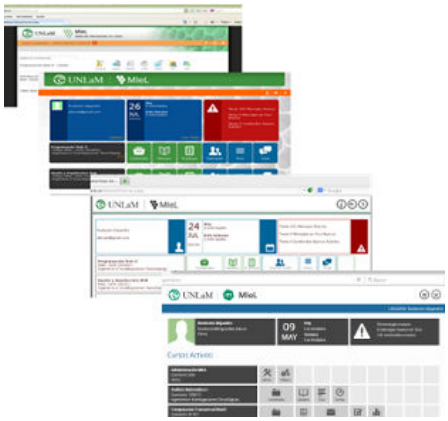


Fig. 8: Evolución de la interfaz de MIeL

La Universidad de La Matanza decidió impartir transversalmente, los conocimientos referidos a Ingles y computación, asumiéndolos como sustrato necesario para la formación universitaria. Específicamente Computación Transversal, es una asignatura en la que se viene utilizando la plataforma MIeL desde sus orígenes y el nivel 2 de esta materia, el que nos permite comparar diferentes versiones de la interfaz. La variedad de la composición del estudiantado, al ser transversal, se asemeja a la variedad observada en la población de la universidad permitiéndonos extrapolar los resultados.

Se construyó un instrumento para encuestar a los estudiantes de este nivel y se realizó sobre una muestra de 540 usuarios. En el mismo se indaga sobre algunas características del usuario como el nivel académico alcanzado y el tiempo y tipo de uso de la navegación en internet (en promedio navegan 2 horas diarias y usan la computadora 3 horas diarias). El resto de las preguntas están orientadas a la relación entre el usuario y el dispositivo: Encontrar el contenido buscado le resulta:(Imposible - 0%, Muy difícil – 5%, Difícil – 10%, Sencillo - 30%, Muy Sencillo – 55%). La navegación de la plataforma te resulta intuitiva?, como te resulta entenderlo (Es muy intuitiva – 35%, Es intuitiva – 40%, Es poco intuitiva – 15%, No es intuitiva – 10%). Adaptarse al uso de la plataforma le resultó (Imposible – 0%, Muy difícil – 5%, Difícil – 5%, Sencillo 55%, Muy Sencillo 35%). La iconografía le resultó (Muy comprensible – 70%, Comprensible – 25%, Poco Comprensible – 5%, Nada Comprensible – 0%). Cómo evalúa la facilidad de uso (Mucho mejor – 20%, Mejor – 70%, No tiene cambios significativos – 12%, Peor – 5%, Mucho Peor – 3%)

Por otra parte se tomaron medidas adicionales como: Tiempo medio de permanencia en la plataforma (en horas diarias), Como valoriza la estética (entre 1 y 10 puntos. Promedio 8 puntos), ¿Cuáles de las herramientas ofertadas por la plataforma le resultaron más sencillas de utilizar? (Calendario, Contenidos y Foros) y ¿Que herramientas que necesita no son provistas por la plataforma? (Comunicación con otras plataformas, email, sms, Facebook y facilidad de uso a través de dispositivos móviles).

5 Consideraciones finales

Desde la óptica del Diseño de Interacción, haciendo hincapié en la interfaces orientadas a la educación, es importante poseer el conocimiento de los aspectos claves de las interfaces y su utilización, para aprovecharlos en razón de posibilitar la inclusión de una mayor cantidad de usuarios y en la mejora de las interacciones hombre-máquina.

Las interfaces de usuario, como vínculo de inmersión del hombre en el entorno del trabajo tecnológico actual, exaltan su importancia en el desarrollo de nuevas aplicaciones, más eficaces, eficientes e interactivas, indispensables en la sociedad de la información.

La plataforma MIeL, como un actor involucrado en la mediación tecnológica del hecho educativo, debe propender a reducir cualquier barrera de acceso que dificulte o reduzca la posibilidad de los usuarios de acceder a los contenidos y herramientas mediados por ella. La utilización de conceptos discursivos accesibles al conjunto del universo objetivo y a su vez la consecución de una interfaz más sincrética que reduzca los tiempos de comprensión y acceso a los elementos mediada por ella, debe ser siempre un objetivo a seguir.

Los aspectos históricos y evolutivos deben ser afrontados de manera científica, para no dejar al azar, lineamientos, pautas y estructuras comunicacionales inherentes al diseño de las interfaces, con el propósito de obtener interacciones desarrolladas con la calidad necesaria para ser accesibles, tanto a los nativos como los inmigrantes digitales.

6 Bibliografía

1. Crumlish Ch. & Malone E. (2009). *"Designing Social Interfaces: Principles, Patterns and Practices for Improving the User Experience"*. Sebastopol; O´Reilly.
2. Marrero Expósito, C. (2006). "*Interfaz gráfica de usuario*". Aproximación semiótica y cognitiva. Tenerife: Universidad de la Laguna. Disponible en: http://www.chr5.com/investigacion/investiga_igu/igu_aproximacion_semio-cognitiva_by_chr5.pdf, recuperado el 10 de julio de 2014.
3. Mordecki, D. (2006). *"Miro. Leo. Luego Pienso."* Disponible en: http://www.mordecki.com/html/miroleopienso.php, recuperado el 15 de julio de 2014.
4. Scolari, C. (2004). "*Hacer clic. Hacia una socio-semiótica de las interacciones digitales*". Barcelona: Editorial Gedisa.
5. Scolari, C. (2009). *"Hipermediaciones. Elementos para una Teoría de la Comunicación Digital Interactiva"*. Barcelona: Editorial Gedisa.
6. Schneiderman, B. (2005). "*Diseño de interfaces de usuario*". Madrid: Pearson Education.
7. Szabo, K. (1995). *"Methapors and the user interface"*. Disponible en:
8. http://www.katalinszabo.com/metaphor.htm, recuperado el 10 de julio de 2014.

Desarrollo de Personajes Virtuales: estrategias y desafíos para la construcción del aula virtual

Alejandro Héctor González[1], Cristina Madoz[1],
Leandro Matías Romanut[2], Eugenia Olaizola[2]

[1]Instituto de Investigación en Informática III – LIDI – Facultad de Informática UNLP
[2]Dirección General de Educación a Distancia y Tecnologías UNLP
{agonzalez,cmadoz}@lidi.info.unlp.edu.ar
{leandro.romanut,eugenia.olaizola}@presi.unlp.edu.ar

Abstract. En el presente trabajo intentaremos presentar el desarrollo y los desafíos de un personaje virtual, herramienta que actualmente se está utilizando en el entorno virtual de enseñanza y aprendizaje AulasWeb basado en Moodle. A través del personaje virtual, se ofrece asistencia en el selector de actividades y recursos para orientar, promover, acompañar a los docentes en la inclusión de actividades colaborativas y, puntualmente, se expande la ayuda en el contexto del curso para colaborar con decisiones tecnológicas-pedagógicas-didácticas del aula virtual. Finalmente, se presenta la sistematización de los resultados logrados, a través de las experiencias en los cursos y encuestas, de uso de la herramienta en el marco de la Universidad Nacional de La Plata.

Keywords: personajes virtuales, entornos virtuales de enseñanza y aprendizaje, aprendizaje colaborativo, b-learning, aulas virtuales.

1 Introducción

Desde el 2004, la Dirección General de Educación a Distancia y Tecnologías (EADyT) dependiente de la Secretaría de Asuntos Académicos de la Universidad Nacional de La Plata (UNLP) viene generando un espacio de formación donde se realizan diferentes capacitaciones dirigidas a docentes que provienen de distintos niveles (pre-grado, grado, posgrado, extensión) y que están interesados en integrar las tecnologías digitales a sus prácticas de enseñanza. La experiencia en el dictado de diversas ediciones de los cursos de formación, permite visualizar que los docentes tienen dificultades al momento de plantear una propuesta mediada por tecnologías; específicamente al repensar las actividades, así como el diseño general de las aulas virtuales.
Al mismo tiempo, en el marco de esta problemática, en una tesina de grado[1] de la Facultad de Informática (UNLP) se analizaron diferentes entornos virtuales de

[1] Romanut, L. (2016) *Desarrollo de asistencia en el selector de actividades de Moodle para facilitar la incorporación de herramientas de trabajo colaborativo*. Lic. en Sistemas, Facultad de Informática - UNLP.

enseñanza y aprendizaje (EVEA): Moodle (AulasWeb[2] y Qoodle[3]), Sakai[4], IDEAS[5] utilizados en distintas universidades. En la tesina se analizó cómo dichos EVEA están organizados, en diversos aspectos del sistema, y en particular qué herramientas están disponibles para incorporar dentro de las aulas virtuales. Se profundizó en el estudio de las herramientas colaborativas provistas por estos entornos y que nivel de ayuda presentan cada una de ellas. Asimismo, se hizo énfasis en las actividades de trabajo colaborativo ya que a través de la sistematización de las capacitaciones y los datos extraídos de AulasWeb se había visualizado una escasa inclusión de este tipo de recursos.
AulasWeb es utilizado en la Dirección General de EADyT, y trabaja en forma conjunta con el soporte tecnológico de servidores y backups del III-LIDI instituto de investigación en informática de la Facultad de Informática de la UNLP.
A partir del relevamiento se propone, para AulasWeb desarrollar e implementar un personaje virtual que colabore como guía dentro de las aulas virtuales y, al mismo tiempo una serie de instructivos que incluyan posibles situaciones pedagógicas para las actividades de trabajo colaborativo. Estos instructivos se crearon basados en encuestas realizadas a profesionales en el uso y coordinación de aulas virtuales.

2 Marco teórico

Los procesos educativos mediados por tecnologías permiten repensar y resignificar las propuestas de enseñanza. La creación de los entornos virtuales de enseñanza y aprendizaje es una de las innovaciones más significativas en el campo de la educación a distancia; las funcionalidades de los EVEA han evolucionado, incrementado las posibilidades de realizar propuestas educativas colaborativas en línea.

2.1 Modalidades educativas

Las modalidades educativas están relacionadas con el ambiente en el que se desarrollan los procesos de enseñanza y aprendizaje. Definen la utilización de los medios, de los recursos educativos y establece las acciones de los sujetos del proceso -estudiantes, tutores, coordinadores-. [1]

Bates explica que la modalidad educativa puede cambiar de acuerdo a la cantidad de e-learning que interviene en el proceso. “El e-learning es un término conveniente para cubrir una amplia gama de usos de la tecnología en el aprendizaje y la enseñanza. Todas las actividades de la computadora y basadas en internet que apoyan en la enseñanza y el aprendizaje tanto “on-campus” como en la distancia. Esto incluye usos administrativos, así como académicos de las tecnologías de información y comunicación que apoyan el aprendizaje.” [2]

[2] https://aulasweb.ead.unlp.edu.ar/
[3] https://acceso.uvq.edu.ar/
[4] https://sakaiproject.org/
[5] https://ideas.info.unlp.edu.ar/login

El incremento de e-learning en los procesos de enseñanza dan lugar a dos términos: el b-learning y el aprendizaje distribuido. González define al b-learning (Aprendizaje híbrido o combinado) como la modalidad que combina la enseñanza presencial con la no presencial de tal manera que ambas experiencias de aprendizaje se vuelven imprescindibles para completar con éxito los objetivos de aprendizaje. [3][4]

Al involucrar de manera total el e-learning, se hace referencia a la educación a distancia, dado que involucra diferentes formas de comunicación. Algunos autores definen a la educación a distancia como:

"La enseñanza a distancia es un sistema tecnológico de comunicación bidireccional (multidireccional), que puede ser masivo, basado en la acción sistemática y conjunta de recursos didácticos y el apoyo de una organización y tutoría, que, separados físicamente de los estudiantes, propician en éstos un aprendizaje independiente (cooperativo)." [5]

"La educación a distancia es un aprendizaje planificado que ocurre normalmente en un lugar diferente al de la enseñanza, por lo tanto, requiere de técnicas especiales de diseño de cursos, de instrucción, de comunicación, ya sea por medios electrónicos u otro tipo de tecnología, así como de una organización especial." [6]

Por otro lado, Mena, Diez y Rodríguez señalan a la educación a distancia como una modalidad educativa que, mediatizando la mayor parte del tiempo la relación pedagógica entre quienes enseñan y quienes aprenden a través de distintos medios y estrategias, permite establecer una particular forma de presencia institucional más allá de su tradicional cobertura geográfica y poblacional, ayudando a superar problemas de tiempo y espacio. [7]

De las definiciones anteriores puede afirmarse que la educación a distancia en la actualidad es un proceso educativo que garantiza la relación pedagógica entre los docentes y los estudiantes a través de distintos recursos y estrategias, siendo este proceso mediado por tecnologías. Los entornos de aprendizaje y enseñanza apoyan la comunicación y el intercambio.

2.2 Las TIC en la enseñanza y el aprendizaje

La incorporación de las tecnologías de la información y comunicación (TIC) en la educación, produce una serie de cambios y transformaciones en las formas en que se representan y llevan a cabo los procesos educativos en general.

Desde una visión constructivista, explica Barriga, "el aula", lejos de referirse a un conjunto de recursos físicos, consiste en un sistema interactivo en el cual ocurren una serie de transacciones comunicativas. Este sistema genera un ambiente particular de trabajo para la construcción del aprendizaje, determinado por una serie de reglas de organización y participación, este ambiente se denomina "ambiente de aprendizaje". [8]

En la formación virtual, a diferencia de lo que ocurre en la enseñanza presencial, todas las actividades, intercambios y relaciones generadas ocurren mediados por tecnología. Los Ambientes Virtuales de Aprendizaje (AVA) son el contexto donde los elementos que conforman el aula virtual interactúan e intercambian información con el fin de construir conocimiento, basando su acción en una serie de reglas, todo ello, logrado por la mediación de herramientas informáticas.

Desde las corrientes actuales del aprendizaje se plantea la potencia de aprender con y de los otros; de buscar estrategias para que los estudiantes aprendan en grupo colaborando con sus pares para resolver problemas auténticos que supongan un uso situado del conocimiento. Es decir, se privilegian las perspectivas que permiten aprender a colaborar, es decir, a escuchar las opiniones de los otros, considerarlas, buscar acuerdos e instrumentar juntos las soluciones generadas por el grupo, expresan Edwards y Mercer, Wertsch, Coll y Mercer. [9] [10] [11] [12]

Gros define que, en un proceso colaborativo, las partes se comprometen a aprender algo juntos. Lo que debe ser aprendido sólo puede conseguirse si el trabajo del grupo es realizado en colaboración. Es el grupo el que decide cómo realizar la tarea, qué procedimientos adoptar, cómo dividir el trabajo, las tareas a realizar. La comunicación y la negociación son claves en este proceso. [13]

"El aprendizaje colaborativo es la adquisición de destrezas y actitudes que ocurren como resultado de la interacción en grupo." [14]

2.3 Entornos Virtuales de Enseñanza y Aprendizaje

Cañellas Mayor explica que un LMS (Learning Managment System - Sistema de Gestión de Aprendizaje) es un software online que permite administrar, distribuir, monitorear, evaluar y apoyar las diferentes actividades previamente diseñadas y programadas dentro de un proceso de formación completamente virtual (e-learning) o de formación semi-presencial (b-learning). [15]

3 Justificación y contexto de desarrollo

La tesina de investigación antes mencionada se enmarca dentro el proyecto de investigación: "*Tecnologías para sistemas de software distribuidos. Calidad en sistemas y procesos. Escenarios educativos mediados por TIC*", del Instituto de Investigación en Informática III-LIDI de la Universidad Nacional de La Plata. Durante la investigación se consideró la sistematización reflexiva de las experiencias en capacitación implementada por la Dirección General de Educación a Distancia y Tecnologías de la UNLP. Tanto en las capacitaciones como en la tesina se observó la falta y necesidad de incluir actividades orientadas al trabajo colaborativo en las propuestas de enseñanza de los docentes; tareas colaborativas que son posibles a través de las herramientas provistas por el entorno AulasWeb (basado en Moodle).

En este contexto y con el fin de profundizar sobre las herramientas de trabajo colaborativo en los EVEA se llevó a cabo una encuesta anónima y con preguntas abiertas; esta encuesta se realizó a los docentes y coordinadores de aulas virtuales de la UNLP con el fin de dar a conocer su trabajo en entornos desde una perspectiva no solo de usabilidad, sino de su experiencia en la enseñanza (interactuando con alumnos, materiales, comunicándose, interaccionando pedagógicamente). De las encuestas realizadas (54) se relevó que no es frecuente la integración de las herramientas de trabajo colaborativo y que generalmente las más utilizadas son los foros (en sus distintas opciones) y las tareas (espacio de entrega de actividades con retroalimentación). Concluido el análisis pudimos comprender que a los profesores se

le presentan similares problemas o complejidades al momento de utilizar las herramientas de trabajo colaborativo que se refieren a las siguientes cuestiones:

- **Configuración**. Las actividades de trabajo colaborativo poseen cierta complejidad en su configuración lo que provoca que los docentes se abstengan de usarlas.
- **Material bibliográfico**. Los docentes, por diferentes factores ajenos al presente análisis, no intentan dinamizar las actividades de sus cursos incluyendo actividades colaborativas dado que consideran que no es posible enseñar sus contenidos con tecnologías digitales. Generalmente trabajan con texto (recursos de tipo archivos en AulasWeb) convirtiendo el aula en un sitio "estático", donde los alumnos se descargan documentos. Esto sucede tanto en la modalidad de extensión del aula como en las propuestas a distancia.

En síntesis, si bien estamos convencidos de que los docentes están dispuestos a trabajar con las actividades colaborativas, todavía no rediseñan ni reformulan sus prácticas educativas mediadas. En este marco nos preguntamos: *¿cómo podemos favorecer trayectos a distancia con instancias de aprendizaje colaborativo?, ¿cómo acompañamos y orientamos a los docentes para la selección de las más adecuadas y pertinentes actividades para sus clases?, ¿qué necesitan conocer los docentes antes de implementar una actividad?*

4 Herramienta implementada

Con el fin de tener una herramienta flexible y que pueda ser incorporada fácilmente en cualquier entorno Moodle, se optó por realizar, en primera instancia, una modificación al código de AulasWeb para que permita incorporar un personaje virtual que colabore en la selección de actividades y recursos de Moodle. A partir de esta modificación, en la interfaz del selector de actividades, el docente puede ver el listado de todas las actividades y recursos que posee Moodle para agregar a las aulas. Cuando selecciona una de las actividades o recursos, se muestra sobre la misma interfaz una descripción de la herramienta. A partir de este selector, se extienden tres funcionalidades:

- Organizar el selector de actividades en tres secciones:
  - **Actividades de Trabajo Colaborativo** (Base de Datos, Chat, Foro, Glosario, Taller, Tarea, Wiki, y otras herramientas externas como VPL y BigBlueButton).
  - **Otras Actividades** (Consulta, Cuestionario, Herramienta Externa, Lección, Paquete SCORM).
  - **Recursos** (Archivo, Carpeta, Etiqueta, Libro, Página, URL).

- Ofrecer para cada actividad de trabajo colaborativo una ayuda pedagógica sobre en qué situaciones educativas podrían emplear dicho recurso.

- Ofrecer un asistente que permite reproducir la ayuda a través de un audio y ver un ejemplo de uso de la actividad seleccionada.

4.1 Personaje virtual

El personaje virtual se creó a través del sitio "Pocoyó" mediante su aplicación Pocoyize[6], la cual permite crear avatares de caricaturas y descargarlos de manera gratuita. La incorporación de un personaje intenta generar una estrategia que oriente a los docentes en la comprensión de las actividades de trabajo colaborativo.
En esta oportunidad se decidió trabajar con un personaje que cumple el rol de tutor tecnológico, que es aquel que realiza el acompañamiento tecnológico durante el aprendizaje de los alumnos y colabora en el entendimiento y uso de las tecnologías involucradas en el aula virtual.

Según Rib Davis, los ingredientes para el personaje son los que resultan de uno mismo y que hacen que cada uno sea diferente de los demás. Si bien en la obra de Rib se describe el armado de los personajes para teatro, cine y literatura, los elementos de creación resultan adaptables a la creación de personajes virtuales. [16] [17]
Para lograr un personaje se deben tener en cuenta tres aspectos:

- ¿Cómo es el personaje cuando nace (por su genética y su entorno)?
- ¿Cómo es el personaje por lo que va aprendiendo y llega a ser a través de la experiencia?
- ¿Cómo es el personaje ahora?



Fig. 1. Manu, el tutor tecnológico

5 Nuevas funcionalidades agregadas

En el año 2017, se realizó una actualización de la herramienta, se extendieron los instructivos pedagógicos a cada actividad (Consulta, Cuestionario, Herramienta Externa, Lección, Paquete SCORM) y recurso (Archivo, Carpeta, Etiqueta, Libro, Página, URL) del entorno AulasWeb, basado en Moodle. Así también, se incluyó el asistente virtual con sus funcionalidades ya desarrolladas: reproducir el instructivo y mostrar un ejemplo de la actividad/recurso seleccionada.

Actualmente, se están realizando nuevas mejoras:

1. La contabilización de la cantidad de clicks realizados sobre el personaje virtual con la finalidad de saber el uso y consulta del mismo.

[6]http://www.pocoyo.com/pocoyizador

2. Se ofrece espacio de preguntas frecuentes automatizadas que ofrece el asistente “Manu”, a nivel del curso para colaborar con los docentes en cuestiones de configuración y diseño del aula virtual.

3. Se presenta un listado de herramientas externas que se puedan incluir (embeber) dentro del entorno AulasWeb. Este listado constaría de una breve descripción, un tutorial de uso y el enlace a la página web de la aplicación. Por ejemplo: construir presentaciones online (Prezi), diseñar videos interactivos (Moovly), confeccionar collages (Fotor - BeFunky), organizar líneas de tiempo (Capzles), elaborar infografías (RAW), armar nubes de etiquetas (Tagxedo - Word it Out), desarrollar murales colaborativos (Padlet), construir imágenes interactivas (Thinglink), etc.

6 Experiencias

Se instaló y utilizó la modificación desarrollada al selector de actividades en el servidor donde reside el entorno AulasWeb (basado en Moodle) de la Dirección General de Educación a Distancia y Tecnologías de la UNLP.

Se realizó una primera encuesta, a través de un formulario online creado con la herramienta Google Drive, con el objetivo final de indagar el nivel de aceptación del prototipo en el entorno AulasWeb y extraer, algunas sugerencias y mejoras que puedan realizarse sobre el desarrollo realizado para este trabajo.

6.1 Instancias de formación con Manu, el tutor tecnológico

En el año 2015 se realizó una primera prueba piloto del asistente en el curso “Introducción al uso de los entornos virtuales de la UNLP soportados en Moodle (Aulas Web – Cavila – Cursos externos)” ofrecido por la Dirección General EaDyT – UNLP. El objetivo del curso es que los profesores, destinatarios del curso, se apropien de conceptos básicos para poder resignificar en sus propuestas y que se acerquen a la edición básica de las herramientas de las aulas virtuales del entorno Moodle.

En la actualidad las capacitaciones relacionadas con la enseñanza en línea y el uso del entorno AulasWeb son: “Entornos virtuales soportados en Moodle – Nivel 1” y “Entornos virtuales soportados en Moodle – Nivel 2”; ambas forman parte del trayecto obligatorio de un Ciclo de Formación para la Gestión de Proyectos de Educación a Distancia de la Dirección General de EADyT y que tiene como destinatarios docentes, principalmente, de la UNLP. En las instancias formativas se abordan diversas herramientas y aspectos del diseño de propuestas educativas implementadas en un aula virtual y en el contexto de distintas modalidades. Se hace hincapié en el uso del asistente para la resolución de las actividades del curso como herramienta de apoyo y consulta.

Al finalizar el curso se realiza una encuesta sobre el sentido y las posibilidades que ofrece el personaje virtual. A continuación presentamos algunas preguntas y resultados:

- **¿Utilizó los instructivos disponibles como ayuda para diseñar las prácticas educativas que involucren las actividades y/o recursos de AulasWeb que abordaron en el transcurso del curso?**

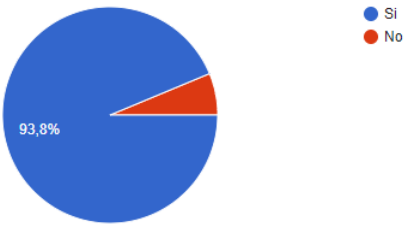


Fig. 2. Resultados de la primera pregunta

- **¿Qué mejoras propondría para la herramienta desarrollada (Personaje Virtual + Instructivos) en el Selector de Actividades?**

- *"Más que una mejora, es una ampliación. Propondría explorar algunos recursos más."*
- *"Cambiar la grabación a voz humana y no de computadora."*
- *"Tal vez tendría que haber una referencia que indique que hay que clickear sobre el personaje para acceder a ellos."*
- *"Quizá una primera vista con menos texto o esté más punteado puede hacer que uno lea más fácil de que trata la actividad o recurso y si uno desea saber más clikear en la información correspondiente."*

Estas respuestas permiten inferir que los propios docentes, destinatarios de la propuesta de intervención y desarrollo tecnológico, son quienes entienden la herramienta como una "ampliación" y proponen que a medida que se construya la demanda accedan a más información.

7 Conclusiones

En términos generales podemos afirmar que al incorporar el asistente en los LMS que se utilizan actualmente en la UNLP: AulasWeb, CAVILA y Cursos Externos, todos basados en Moodle, favorecimos las propuestas áulicas mediadas; es decir, con la implementación de "Manu" se realizó un aporte tecnológico, pedagógico, didáctico a la Dirección General de Educación a Distancia y Tecnologías de la UNLP y toda su comunidad. Entendemos que "Manu" como herramienta colabora en el armado, creación, integración, uso e inclusión de las actividades interactivas colaborativas en las propuestas educativas que los docentes llevan adelante en sus propios cursos.

Las sucesivas experiencias de formación denotan que el desarrollo permite colaborar en la comprensión de las actividades y recursos desde una perspectiva pedagógica como así también desde un punto de vista tecnológico.

Asimismo, la utilización del canal de audio es un primer intento de mejorar la accesibilidad a personas con capacidades diferentes atendiendo y también promoviendo inclusión digital, social y educativa.

8 Trabajo futuro

Frente a la preocupación de continuar construyendo espacios de enseñanza potentes, de promover aprendizajes significativos socialmente y atendiendo a nuevas propuestas, necesidades, demandas de la educación a distancia en la UNLP se proponen las siguientes acciones a futuro para favorecer y complejizar la herramienta desarrollada:

- Implementar nuevas encuestas como instrumento de evaluación de las capacitaciones de la Dirección para ver el impacto en la comunidad docente y obtener retroalimentación para mejorar las ayudas y sugerencias didácticas para cada herramienta.
- Implementar encuestas a todos los usuarios con rol docente de los entornos de la UNLP para conocer sus vivencias, experiencias y demandas.
- Fortalecer los espacios de formación docente que gestiona la Dirección General de Educación a Distancia y Tecnologías de la UNLP que intentan acercar a los docentes a la elaboración de materiales didácticos interactivos, digitales, y que reflexionan sobre los espacios de comunicación dentro del aula virtual, el rol de los tutores, las posibilidades pedagógicas de la virtualidad.
- Agregar nuevas funciones para personalizar el asistente en aspectos de imagen de acuerdo con como lo prefiera el docente del curso: cambiar la ropa, el cuerpo, la voz.

9 Bibliografía

1. Jiménez, D. Nuevos paradigmas educativos y modalidades educativas. Recuperado de: http://es.slideshare.net/auri_desi/modalidades-educativas-5551315 (2010)
2. Bates, A. T. Technology, e-learning and distance education. Routledge. (2005)
3. González, A. H. Educación a Distancia y Tecnologías Digitales en la Enseñanza Universitaria. Recuperado de https://prezi.com/v-vkyghu8zsc/2015-especializacion-en-docencia-universitaria-unlp/?utm_campaign=share&utm_medium=copy (2015)
4. González, C. Cuadro comparativo: Blended learning, Distributed Learning, Online Learning. Recuperado de http://es.calameo.com/read/001790350da562655eaed (2012)
5. García Aretio, L. La Educación a Distancia. De la teoría a la práctica. Ariel editorial S. A. (2001)
6. Moore, M. & Kearsley, G. Distance education: A systems view. Boston, MA: Wadsworth Publishing. (1996)

7. Mena, M., Diez, L. M., y Rodríguez, L. El Diseño de proyectos de Educación a distancia. Colección Itinerarios. Ediciones La Crujía y Editorial Stella. Buenos Aires. (2005)
8. Barriga, F. D. Educación y nuevas tecnologías de la información y la comunicación: ¿Hacia un paradigma educativo innovador? (2008)
9. Edwards, D. y Mercer, N. (1988). El conocimiento compartido: el desarrollo de la comprensión en el aula. Madrid: Paidós.
10. Wertsch, J. V. (1988). Vygotsky y la formación social de la mente. Buenos Aires: Aique.
11. Coll, C. (2001). Constructivismo y educación: la concepción constructivista de la enseñanza y del aprendizaje. En Coll, C., Palacios, J. y Marchesi, A. (Compiladores). Desarrollo psicológico y educación. Vol. II, (157-188). Psicología de la Educación Escolar. Madrid: Alianza.
12. Mercer, N. (2001). Palabras y mentes. Cómo usamos el lenguaje para pensar juntos. Barcelona: Paidós.
13. Gros, B. El Aprendizaje Colaborativo a través de la Red: Límites y Posibilidades. En Aula de innovación educativa, ISSN 1131-995X, Nº 162, 2007, págs. 44-50. (2000)
14. Salinas, J. El aprendizaje colaborativo con los nuevos canales de comunicación, 199 – 227. En Cabero, J. (Editor). Nuevas tecnologías aplicadas a la educación. Madrid: Síntesis. (2000)
15. Cañellas Mayor, A. LMS y LCMS: funcionalidades y beneficios. Recuperado de http://www.centrocp.com/lms-y-lcms-funcionalidades-y-beneficios/ (2014)
16. Rib, D. Escribir guiones: desarrollo de personajes. Editorial Paidós, manuales de escritura. Barcelona. España. (2004)
17. González, A. H. TICs en el proceso de articulación entre la Escuela Media y la Universidad. Tesis de Maestría, Facultad de Informática, UNLP. (2008)

Una Herramienta de Autor Web para Generar Aplicaciones Educativas Basadas en Posicionamiento con Interacción Física

Federico R. Mozzon Corporaal[1], Pedro A. Dal Bianco[1], Alejandra B. Lliteras[1,2], Julián Grigera[1,2], Silvia E. Gordillo[1,2]

[1] LIFIA, Facultad de Informática, UNLP
[2] CICPBA, Argentina
{fmozzon, pdalbianco, lliteras, jgrigera, gordillo}@lifia.info.unlp.edu.ar

Abstract. Este trabajo presenta una propuesta para que, mediante el uso de una herramienta de autor, docentes puedan generar sus propias soluciones para llevar adelante aprendizaje móvil basado en posicionamiento, considerando la interacción del alumno con elementos del mundo físico. Se describe también una herramienta de autor prototípica web que implementa los conceptos de la solución propuesta. La herramienta fue usada exitosamente para configurar una actividad educativa basada en posicionamiento en la temática de "Alimentación saludable".

Keywords: herramienta de autor, aprendizaje móvil basado en posicionamiento, interacción física.

1 Introducción

Cada vez es mayor la presencia de aplicaciones o herramientas orientadas a educación que permiten acercar al usuario al conocimiento, utilizando la tecnología móvil como intermediaria. Debido a esto, resulta natural que tanto los docentes como los estudiantes las encuentren cada vez más familiares y motivadoras. Si bien su utilización trae ventajas, es importante resaltar que, actualmente en la mayoría de los casos es el docente quien debe adaptar sus clases al formato propuesto por las aplicaciones existentes, y no son las aplicaciones las que se adaptan al contexto particular de cada docente.

Resulta relevante pensar en la manera en la que los docentes pueden adoptar la tecnología móvil y en particular, al aprendizaje móvil en su práctica docente [1]. De acuerdo a lo expresado en [2], un factor importante para su adopción es que el docente se sienta confiado y perciba que su uso es sencillo y significativo.

De acuerdo a [3], poder traducir conceptos de educación en herramientas informáticas constituye una habilidad muy valiosa para un docente. Sin embargo, de acuerdo a [4], la mayoría de los docentes no cuenta con habilidades de programación para desarrollar sus propias herramientas de aprendizaje móvil, siendo este el principal limitante a la hora de incorporar estas tecnologías en educación.

Para mitigar esta limitación resulta necesario brindarle al docente herramientas para que sea él quien diseñe la actividad educativa, y al mismo tiempo para que sea la aplicación la que se adapta a dicha actividad. Sería deseable que el docente tenga la posibilidad de llevar a cabo el diseño sin requerir de ningún conocimiento sobre programación ni de la asistencia de expertos en tecnología, preferentemente a través de una interfaz amigable y en un lenguaje familiar para él.

En este trabajo se propone una versión prototípica de una herramienta de autor web para la configuración de una aplicación educativa basada en posicionamiento. Con este aporte se espera contribuir con un marco de trabajo y discusión ante la necesidad de herramientas fáciles de usar y que resulten significativas para el docente. El fin de la propuesta es colaborar en la adopción del aprendizaje móvil en el contexto educativo.

Este trabajo se estructura de la siguiente manera: en la Sección 2 se presentan algunos trabajos relacionados. En la Sección 3 se describe el trabajo adoptado como base para la presente propuesta. En la Sección 4 se detalla la propuesta de solución. En la Sección 5 se muestra una posible implementación. Por último, en la Sección 6 se presentan conclusiones y trabajos futuros.

2 Trabajo Relacionado

Existen en la literatura algunos trabajos que permiten usar posicionamiento para reforzar las experiencias en aplicaciones educativas. También pueden encontrarse trabajos que permiten que usuarios finales (con mayor o menor nivel de experiencia en programación) puedan generar aplicaciones educativas con relativamente poco esfuerzo. En esta sección se presentan algunos de estos trabajos y su relación con la herramienta de autor propuesta.

En [5] se presenta una herramienta de autor in situ para diseñar experiencias educativas basadas en posicionamiento empleando GPS y códigos QR como mecanismo de sensado de la posición. La herramienta toma como base el trabajo presentado en [6] y exporta en archivos XML las capas de contenido y posiciones. La herramienta no permite la generación de códigos QR para identificar las posiciones en las que se brindará contenido y si bien plantea la posibilidad de trabajar con elementos del mundo físico (considerando lo propuesto en [7]), tampoco provee un mecanismo de generación de códigos QR para identificarlos ni genera experiencias ejecutables.

Por otro lado, el MIT provee un entorno de programación visual llamado App Inventor [8] que permite crear aplicaciones funcionales para dispositivos móviles Android. Posee una herramienta que soporta programación en bloques, permitiendo la creación de aplicaciones complejas, en un tiempo menor respecto a entornos de programación tradicionales. Uno de sus objetivos es el de "democratizar el desarrollo del software y empoderar a las personas, especialmente a jóvenes, a moverse desde el consumo de la tecnología a la creación de la misma" [8]. Una aplicación se conforma por un conjunto de componentes que pueden ser visuales (por ejemplo: botones y cajas de texto) y no visuales (por ejemplo: la cámara del dispositivo y sensores). La funcionalidad de estos componentes se indica mediante bloques de programa [9]. Este

entorno, de acuerdo a lo expresado en [4], ha sido usado, por un lado, para enseñar conceptos de programación y por otro, para que docentes puedan crear sus propias aplicaciones para aprendizaje móvil. De acuerdo a los citados autores, este entorno "permite a los docentes, sin un conocimiento avanzado de programación, desarrollar aplicaciones personalizadas para aprendizaje móvil". Esta herramienta es de uso general y, dado que emplea programación en bloques para la creación de aplicaciones educativas, requiere que los usuarios aprendan conceptos básicos de programación.

En [10] los autores presentan una herramienta de autor para desafíos sensibles al contexto. Su objetivo es permitir a los docentes crear sus propias experiencias educativas y lúdicas basadas en posicionamiento. Utilizan tres mecanismos de posicionamiento, GPS, Beacon y códigos QR. Es posible usar un conjunto de actividades predefinidas para crear nuevas actividades a partir de ellas y no requiere de conocimientos de programación para su uso. Las actividades son una colección de puntos de interés, donde cada punto tiene asociada una actividad, como por ejemplo un mini juego o un cuestionario contextualizado. Esta herramienta no considera elementos del mundo físico para interactuar como parte de la actividad.

3 Trabajo de Base

En un trabajo previo, se creó la aplicación prototípica "Aprendo Jugando", en la cual, y a modo de prueba de conceptos de un enfoque de modelado presentado en [7], se propuso una actividad educativa para recolectar elementos concretos de acuerdo a su uso, para luego depositarlos en base a la clasificación por su material principal (papel, metal o plástico). En [7], se presentaron dos casos de uso de la aplicación. Luego, en base al enfoque de modelado usado para desarrollar "Aprendo Jugando" y habiendo propuesto algunas variantes para incorporarle los conceptos de scaffolding (andamiaje) [11], "feedback formativo" [12] y recolección de datos para poder realizar analíticas del aprendizaje ("learning analitycs") [13]; propusimos la aplicación "Resuelvo Explorando", un prototipo de aplicación educativa basada en posicionamiento [14]. La aplicación "Resuelvo Explorando" es configurable, permitiendo crear actividades en diferentes espacios físicos (por ejemplo, aulas) y utilizando elementos concretos del mundo físico [14] y [15].

La aplicación "Resuelvo Explorando" [14], utiliza como metáfora un viaje en el cual se tienen que recorrer diferentes lugares y realizar tres tareas, en las cuales los alumnos recolectan elementos del mundo real y los colocan en depósitos acorde a una clasificación. Con esta metáfora se busca que el alumno pueda hacer una representación mental de la dinámica de uso, para facilitar la adopción de la aplicación. En el viaje metafórico, hay un origen y un destino preestablecidos, y existen tres lugares que deben visitarse en un orden dado para recolectar elementos del mundo real, mientras que el lugar donde se depositan los elementos recolectados puede visitarse en cualquier momento, ya sea luego de una tarea de recolección o al final del viaje.

La aplicación "Resuelvo Explorando" [14], es configurable mediante la carga de información relacionada al espacio físico y el contenido educativo. Actualmente, el contenido educativo, se ingresa mediante la configuración de un archivo JSON que

contiene los detalles de la actividad educativa, más los de sus correspondientes tareas de recolección, de depósito, y sus elementos del mundo real, incluyendo el código que permitirá que se los identifique luego a través de códigos QR. Adicionalmente, se ingresan imágenes con representaciones gráficas del espacio físico donde se destacan los lugares relevantes en los que se brindará cada tarea de la actividad. Luego, la aplicación configurable es capaz de recuperar esta información dinámicamente, en tiempo de ejecución, para que pueda ser usada durante una experiencia educativa (ver Fig. 1).

Una característica a mencionar de la aplicación configurable es el mecanismo de posicionamiento que usa. En particular se emplea posicionamiento indirecto, mediante códigos QR, lo cual ha resultado de gran efectividad por su simplicidad de aplicación e independencia de los dispositivos (sólo hace falta una cámara), sobre todo considerando su uso en ambientes internos y pequeños, como puede ser un aula dentro de una escuela.

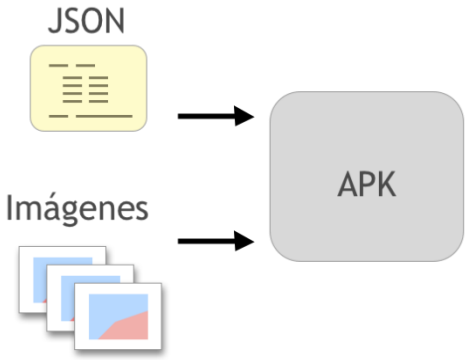


Fig. 1. Esquema de la arquitectura del trabajo de base.

Una limitante de dicho trabajo fue que el archivo de configuración JSON debió ser completado por expertos en tecnología. Es importante destacar también que la aplicación configurable no genera los códigos QR necesarios, y que tanto su creación como el mapping de códigos con ubicaciones se realiza manualmente. Esto motiva el desafío de brindarle al docente la posibilidad de ser él quien configure todos los aspectos sin necesidad de un experto como se menciona en [14] y [15] .

4 Propuesta

En esta sección se describe la propuesta de una herramienta de autor a través de la cual un usuario final (en el caso del dominio propuesto, un docente) pueda configurar mediante una interfaz web, una aplicación educativa basada en posicionamiento de manera integral. La herramienta es capaz de generar una aplicación configurada, lista para ser instalada en un dispositivo móvil Android.

Como se explicó en la Sección 3, la información necesaria para configurar la aplicación es:

- Contenido educativo, mediante un archivo JSON que contenga información de: i) la actividad, ii) las tareas de recolección, iii) los depósitos, iv) los elementos del mundo real que pertenecen a cada tarea de recolección y vi) la información adicional que se corresponde con cada tarea.
- Espacio físico de uso, mediante imágenes que representen dicho espacio y que indiquen hacia dónde debe dirigirse el alumno para realizar la siguiente tarea o ir a depositar.

Y para la interacción con el mundo físico son necesarios códigos QR que se correspondan con: i) las tareas, ii) los depósitos y iii) los elementos.

Tanto el contenido educativo para configurar la aplicación, como los códigos QR necesarios para la interacción con el mundo físico, serán generados por la herramienta en función de lo que ingrese el usuario, mientras que las imágenes del espacio físico, podrán ser cargadas por la herramienta desde una dirección indicada por el usuario. En base a esto, la herramienta generará tres paquetes de información: i) los archivos necesarios para copiar en el dispositivo móvil, ii) un paquete con los códigos QR necesarios para la interacción con el mundo físico y iii) un archivo explicando de qué manera se debe configurar el espacio físico y los elementos de cada tarea de recolección.

Un esquema de la solución propuesta en este trabajo puede ser visualizada en la Fig. 2.

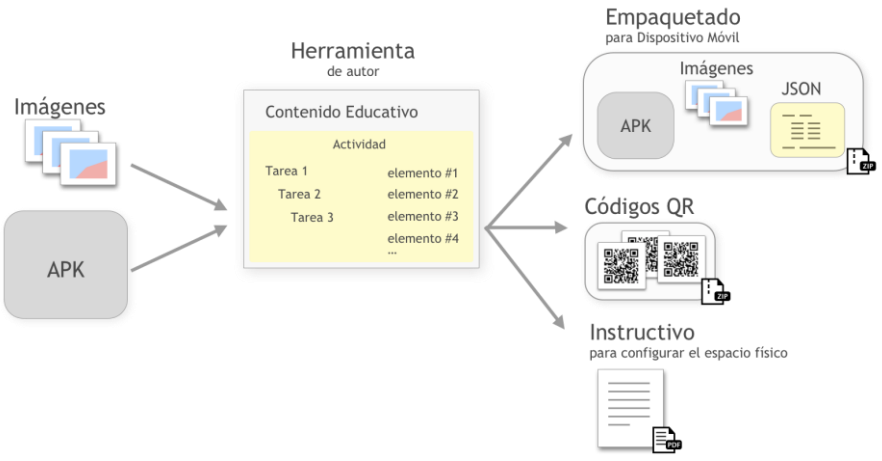


Fig. 2. Esquema de la solución propuesta.

5 Implementación

Para lograr el objetivo planteado en la Sección 4, en esta sección se propone una herramienta de autor, que sirve para configurar la aplicación y para generar los códigos QR necesarios para la interacción física. La herramienta está diseñada para guiar al docente a través de diferentes formularios y pantallas, en los cuales se ingresa la información sobre la actividad educativa, sus depósitos, tareas de recolección, y los elementos del mundo real asociados a cada tarea. Esta información es necesaria para definir el archivo JSON mencionado en la Sección 4.

En un principio se ingresa el nombre de la actividad, el idioma en el que estará dicha actividad (por el momento Español o Inglés) y el objetivo de la misma. Luego se configuran los dos depósitos con su criterio y descripción. También puede agregarse un tercer depósito en caso de que la actividad lo requiera. Posteriormente las tareas se ingresan de una forma similar. En este caso, es posible agregar para cada tarea y de manera opcional, material adicional, que puede ser, por ejemplo, un link a un video, una imagen explicativa o un texto. Además, a cada tarea de recolección se le asocia uno o más elementos del mundo físico. Para cada elemento se ingresa un nombre, una descripción y se le asocia el criterio de depósito al que corresponde. También se debe indicar si es un elemento correcto para ser usado como respuesta para esa tarea.

La Fig. 3 muestra el formulario en el cual se carga la información relacionada a una tarea de recolección.

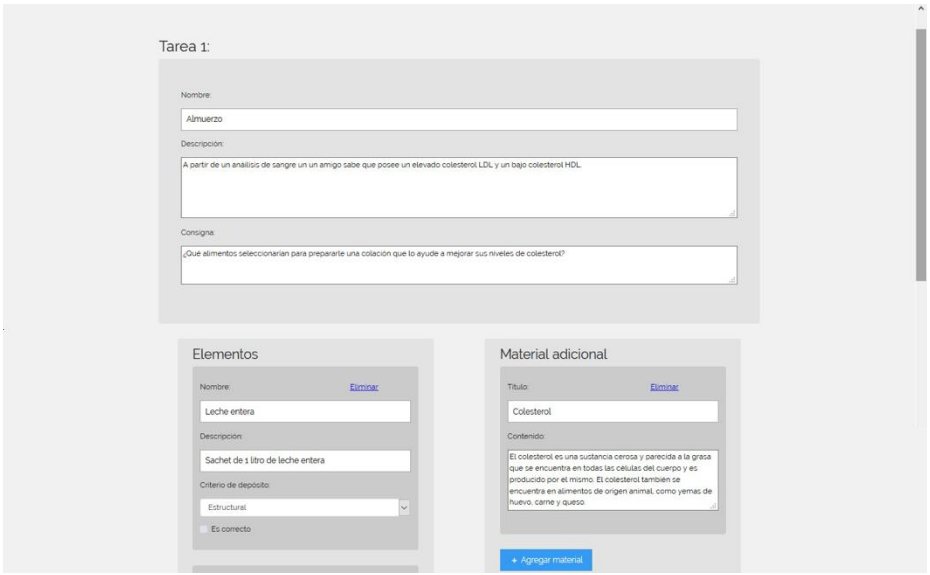


Fig. 3. Formulario de carga de información referida a una tarea de la actividad. En la parte superior se ingresan los detalles de la tarea, en la parte inferior izquierda pueden ingresarse elementos del mundo físico y a la derecha el material adicional.

Al cargar una tarea de recolección, además del nombre, descripción y consigna, se ingresan los elementos del mundo real a considerar en la tarea y de manera opcional material adicional (por ejemplo, un texto).

Usando la herramienta, es posible seleccionar y cargar, desde una galería, imágenes con los lugares relevantes identificados dentro del espacio físico. Con toda la información ingresada y cargada por el usuario, la herramienta se encarga, por un lado, de generar: i) un archivo JSON, ii) los códigos QR para ubicar e identificar a las tareas, depósitos y elementos del mundo real y iii) un instructivo para la configuración del espacio físico usando las imágenes cargadas, información de los códigos QR generados y de la actividad educativa en general. Por otro lado, se recupera del servidor el APK de la aplicación "Resuelvo Explorando".

La herramienta utiliza lo anteriormente descripto para dar como resultado tres paquetes de información. Un primer paquete con un archivo en formato ZIP que contiene el APK junto con el archivo JSON de configuración y las imágenes del espacio físico. Un segundo paquete, con otro archivo ZIP el cual contiene los códigos QR para la interacción física, y por último, un archivo PDF con un instructivo en el cual se le indica al docente cómo configurar el espacio físico y cada tarea de recolección con sus elementos.

Una vez cargados los formularios de la Herramienta, se visualiza una pantalla con el resumen de toda la información ingresada y un botón que permite confirmar la configuración para luego llevar al usuario a una pantalla para descargar los tres paquetes generados.

En la Fig. 4 se visualiza la pantalla desde la cual es posible descargar cada paquete previamente descripto.

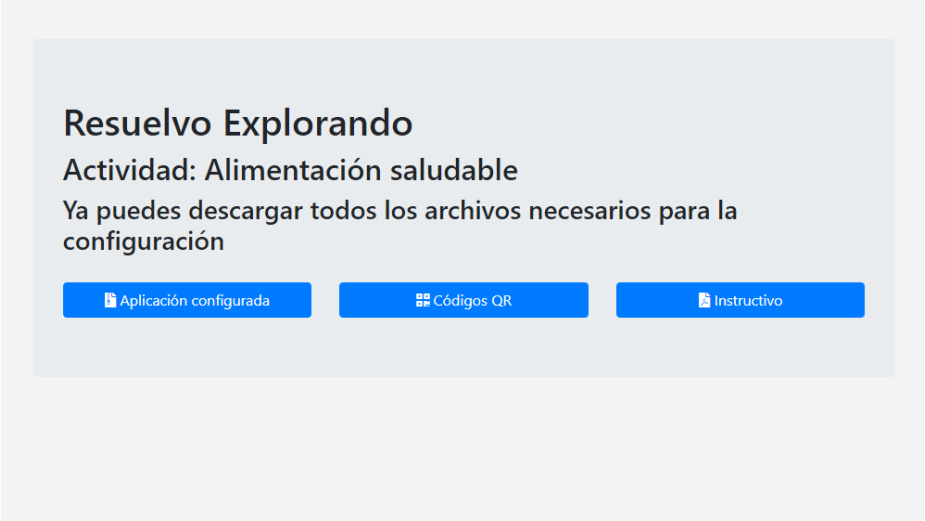


Fig. 4. Pantalla de descargas.

Desde los aspectos tecnológicos usados para la implementación de esta herramienta se utilizó Boostrap[1] como framework CSS. También se usó la librería JQuery[2] de Javascript.

Se utilizó PHP plano para el *back-end*, y sólo se utilizaron dos librerías: i) PHP QR Code[3], que permitió a partir de identificadores internos de los elementos, tareas de recolección y de depósito, generar sus respectivos códigos QR; ii) TCPDF[4], para crear el documento PDF con el instructivo para la configuración de los espacios de uso correspondientes para la actividad y los QRs correspondientes a cada elemento y tarea junto a su respectivo identificador.

Si bien la herramienta de autor se encuentra en una etapa temprana, se ha realizado una primera configuración tomando el diseño conceptual realizado por un equipo multidisciplinario que empleó el framework conceptual propuesto en [16] para el diseño de una experiencia educativa basada en posicionamiento. La configuración propuesta forma parte de un tema curricular para alumnos de 4to. año del secundario[5].

6 Conclusiones y Trabajos Futuros

En este trabajo se presentó una herramienta web de autor que le permite a docentes configurar una actividad educativa para que luego pueda ser puesta en práctica con alumnos mediante la aplicación "Resuelvo Explorando", descripta en la sección 3. La herramienta de autor, ha sido diseñada para ser intuitiva y para que el usuario de la misma, no requiera de ningún conocimiento acerca de programación. De este modo, el docente puede generar una configuración para una actividad educativa basada en posicionamiento, que pueda ser utilizada en cualquier dispositivo Android. El diseño de la actividad educativa, puede ser llevado a cabo empleando el framework conceptual presentado en [16] que incluye diseño in-situ de la actividad educativa. Se destaca que si bien el diseño puede ser in-situ, no es necesario que el uso de la herramienta propuesta lo sea.

La solución propuesta permitirá al docente incorporar el aprendizaje móvil en su práctica docente, con actividades educativas diseñadas por él y para ser llevadas a cabo mediante el uso de dispositivos móviles. Esto no sólo permitirá el empoderamiento del docente, el cual será capaz de evaluar la apropiación del conocimiento por parte de sus alumnos, sino que también los alumnos tendrán una experiencia diferente, mediada por tecnología móvil.

Como trabajo futuro, se planifica sumar un módulo de validación de usuarios a la herramienta de autor, a partir del cual, cada docente podrá mantener información de sus configuraciones, las escuelas y grados/años dónde se llevaron a cabo las puestas

[1] Bootstrap – https://getbootstrap.com
[2] JQuery – https://jquery.com
[3] PHP QR Code - http://phpqrcode.sourceforge.net
[4] TCPDF - https://tcpdf.org
[5] Diseño curricular para la educación secundaria. Biología. 4º Año. URL: http://servicios.abc.gov.ar/lainstitucion/organismos/consejogeneral/disenioscurriculares/secundaria/materias_comunes_a_todas_las_orientaciones_de_4anio/biologia_4.pdf.

en práctica de las aplicaciones educativas configuradas por él, los alumnos que la usaron y toda información que le sirva de retroalimentación a su quehacer docente.

Por otro lado, se planifica trabajar en una comunidad de uso, a fin de fomentar la reutilización de los aspectos involucrados: contenidos educativos, representación del espacio de uso, y la aplicación configurable.

Se planifica realizar pruebas con docentes para que puedan generar sus propias actividades mediante la configuración de la aplicación educativa basada en posicionamiento, y de este modo recibir feedback para analizar la usabilidad de la herramienta con el fin de mejorarla.

Se trabajará además en ampliar la herramienta de autor para dar soporte a la implementación de otros aspectos del enfoque de modelado presentado en [7].

Agradecimientos. Esta herramienta se presenta en el marco del "Proyecto de Aplicaciones e Innovación", convocatoria 2017, Facultad de informática, UNLP. Res. HCD. 04/18. Para el proyecto presentado: "Preservación de especies del mar de la Costa Atlántica Argentina: una experiencia mediada por tecnología móvil". Los autores agradecemos a las personas que participaron en el diseño de la experiencia educativa usada para probar la herramienta presentada: Flavia Eliana Bianchini, Lic. en Biología con orientación en Ecología. Fac. de Cs. Naturales. y Museo, UNLP, Argentina. Juan Pablo Diez, Veterinario. Fac. de Cs. Veterinarias, UNLP, Argentina. Ambos profesores de biología en escuelas secundarias de la ciudad de La Plata, Pcia. de Bs. As., Argentina. A Liliana Rodríguez, Médica. Fac. de Cs. Médicas, UNLP, Argentina. (MP. 114975, Pcia. de Bs. As., Argentina) y Especialista en Nutrición.

Referencias

1. Chiu T.K.F., Churchill D.: Adoption of mobile devices in teaching: changes in teacher beliefs, attitudes, anxiety. Interact. Learn. Environ., vol. 24, no. 2, pp. 317--327, doi: 10.1080/10494820.2015.1113709 (2016)
2. Christensen R., Knezek G.: Readiness for integrating mobile learning in the classroom: Challenges, preferences and possibilities. Comput. Human Behav., vol. 76, pp. 112--121, doi: 10.1016/j.chb.2017.07.014 (2017)
3. Mota, J.M., Ruiz-Rube, I., Dodero, J.M., Person, T., Arnedillo-Sánchez, I.: Learning Analytics in Mobile Applications Based on Multimodal Interaction. In Software Data Engineering for Network eLearning Environments, pp. 67--92. Springer, Cham (2018)
4. Ortega García, A., Ruiz-Martínez A., Valencia A.: Using App Inventor for creating apps to support m-learning experiences: A case study. Computer Applications in Engineering Education vol. 26, no. 3, pp. 431-448 (2018)
5. Zimbello, A.M., Alconada Verzini, F.M., Challiol, C., Lliteras, A.B., Gordillo, S.E.: Authoring tool for location-based learning experiences. In Proceedings of the 4th International Conference on Mobile Software Engineering and Systems, pp. 211-212. IEEE Press (2017)
6. Alconada Verzini, F.M., Tonelli, J.I., Challiol, C., Lliteras, A.B., Gordillo, S.E.: Authoring tool for location-aware experiences. In Proceedings of the 2015 Workshop on Narrative & Hypertext, pp. 21--25. ACM. (2015)
7. Lliteras, A.B.: Un enfoque de modelado de actividades educativas posicionadas que contemplan elementos concretos. Master Thesis, Faculty of Informatics, National University of La Plata, Argentina, http://hdl.handle.net/10915/50030 (2015)

8. AppInventor, http://appinventor.mit.edu
9. Xie, B., Abelson, H.: Skill progression in MIT app inventor. In Visual Languages and Human-Centric Computing (VL/HCC), 2016 IEEE Symposium on, pp. 213--217. IEEE (2016)
10. Hauge, J.B., Stefan, I.A., Stefan, A., Cazzaniga, M., Yanez, P., Skupinski, T., Mohier, F.: Exploring Context-Aware Activities to Enhance the Learning Experience. In International Conference on Games and Learning Alliance, pp. 238--247. Springer, Cham. In: Dias J., Santos P., Veltkamp R. (eds) Games and Learning Alliance. GALA 2017 (2017)
11. Janson A., Soellner M., Leimeister J.M.: The Impact of Procedural Scaffolding on Mobile Learning Outcomes. Acad. Manag. Proc., vol. 2017, no. 1, p. 10802, doi: 10.5465/ambpp.2017.10802abstract (2017)
12. Faber J.M., Visscher A.J.: The effects of a digital formative assessment tool on spelling achievement. Computers & education, vol. 122, pp. 1--8, doi: 10.1016/j.compedu.2018.03.008. (2018)
13. de Witt C., Gloerfeld C.: Mobile Learning and Higher Education. The Digital Turn in Higher Education, pp. 61--79, Springer VS, Wiesbaden, doi: 10.1007/978-3-658-19925-8_6 (2018)
14. Lliteras A.B., Gordillo S.E., Dal Bianco P.A, Mozzon Corporaal F. R.: A customizable location-based Mobile Learning Prototype: A case of study, XIII Latin American Conference on Learning Technologies (LACLO), in press, (2018)
15. Lliteras A.B., Grigera J., Dal Bianco P.A, Mozzon Corporaal F.R., Gordillo S.E. : Challenges in the design of a customized location-based mobile learning application, XIII Latin American Conference on Learning Technologies (LACLO), in press, (2018)
16. Lliteras A.B., Challiol C., Gordillo S.E.: Location-based mobile learning applications: a conceptual framework for co-design. Proc. Twelfth Latin American Conference on Learning Technologies (LACLO), IEEE Press, pp. 358--365, doi: 10.1109/LACLO.2017.8120946 (2017)

Tecnologías aplicadas en el aula universitaria – Caso de aplicación en la enseñanza de programación

Pablo Martín Vera, Rocío Andrea Rodríguez, Edgardo Javier Moreno

Universidad Nacional de La Matanza, Departamento de Ingeniería e Investigaciones Tecnológicas, Grupo de Investigación Desarrollo e Innovación de Software (GIDFIS).
Florencio Varela 1903, San Justo, Buenos Aires, Argentina
{pvera, rocio.rodriguez, emoreno}@unlam.edu.ar

Resumen. El uso de tecnología ya es parte de la vida cotidiana. Computadoras, tablets y en especial los teléfonos celulares son fuente de consulta permanente, no sólo como medio de comunicación sino como un complemento para las tareas diarias. Los estudiantes que son en su gran mayoría nativos digitales nacieron con estas herramientas y por lo tanto su uso en el aula debe promoverse ya que simplemente para esta generación y las venideras la tecnología es parte de su vida por lo que alejarlos de ella sólo causa desgano, desconcierto y falta de interés. El presente trabajo muestra un caso de aplicación donde, tanto dentro como fuera del aula, se utilizan distintas herramientas tecnológicas para facilitar el acceso a la información e incentivar el uso de la tecnología logrando mayor compromiso y aceptación por parte de los estudiantes.

Palabras Claves: Nativos Digitales, Gamificación, Materiales didácticos, dispositivos móviles, teléfonos celulares

1 Introducción

Vivimos en un mundo tecnológico, la tecnología ha llegado para quedarse y es necesario aprender a utilizarla y sacar provecho de ella en todos los aspectos. El uso de la tecnología en la educación no puede quedar relegado. Algunos estudios demuestran que los docentes aprovechan a las TIC (Tecnologías de la Información y Comunicación) al momento de planificar sus clases. Incluso en carreras que no son de informática, como el trabajo [1] vinculado con la carrera de medicina, se destaca que: "El 70.4% de los docentes utiliza TIC en sus clases… El PowerPoint es el material didáctico TIC más utilizado (93.8%), además utilizan… YouTube (39.5%), red social (43.7%). El 49.4% utiliza aula virtual. Los docentes refieren una falta de capacitación en TIC (38.3%) y solicitan capacitación en diseño de aulas virtuales (50.6%)" [1]. Pero aún no se contempla la tecnología que los estudiantes pueden traer consigo y como aprovecharla para que sea utilizada a favor de la clase. Los estudiantes, en su mayoría, traen consigo dispositivos con gran capacidad de procesamiento y medios de comunicación que pueden ser aprovechados y utilizados durante el dictado de las

materias. “La mayoría tiene un teléfono inteligente, aunque opinan que en general no utilizan las aplicaciones para actividades académicas, con excepción del Dropbox; por el contrario, un alto porcentaje las utiliza para jugar, chatear u otra actividad. No obstante, están conscientes de que el uso de los teléfonos móviles les afecta en su rendimiento académico” [2].

“El uso de la tecnología móvil en el aula es un tema de actual debate entre todo tipo de docentes y equipos directivos de centros educativos. Es importante entender que no hablamos de mobile learning, es decir, el conjunto de metodologías de enseñanza-aprendizaje exclusivamente mediante tecnología móvil para educar y aprender en cualquier momento y lugar; sino del uso del móvil como una herramienta más dentro del aula como los ordenadores, las tabletas, los libros o las pinturas” [3].

Podría considerarse que el teléfono celular es una fuente de distracción dentro de un aula, ya que quita el foco de atención en la explicación del docente pero ¿qué sucedería si se utiliza a favor del aprendizaje y se lo incorpora en las tareas áulicas? Si el dispositivo se utiliza dentro de una cursada, el mismo se incorpora y ya pasa a ser un objeto de utilidad y no uno de distracción. “Es entendible que haya ciertas dudas respecto de los dispositivos móviles en el aula. La mayor preocupación es que distraigan a los alumnos… pero en realidad, las distracciones siempre han existido”[4].

La UNESCO (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura) indica que “el incremento sin precedentes del número de teléfonos móviles, tanto en las naciones industrializadas como en los países en desarrollo, genera posibilidades inéditas para aumentar el acceso, la equidad y la calidad en materia de educación. El aprendizaje mediante dispositivos móviles, una modalidad de rápido crecimiento entre las aplicaciones de las TIC a la educación, tiene la capacidad de influir considerablemente en las prestaciones educativas” [5]. “En general se puede decir que la innovación docente en el uso de tecnologías móviles es un ámbito de investigación creciente, fruto de su progresiva implantación en las universidades [6]”.

Los estudiantes en su mayoría nativos digitales tienen otra forma de interactuar y de ver el mundo. No están acostumbrados a los apuntes en papel ni a leer textos demasiado largos. Los métodos tradicionales deben cambiar y adaptarse. El desafío está en ¿cómo presentar los contenidos a estos estudiantes, de qué forma se puede llegar a ellos sin perder su atención y a su vez motivándolos a investigar, a participar, a involucrarse con la materia?

Este trabajo presenta un caso de estudio para la materia Elementos de Programación, que es una materia de primer año para las carreras de Ingeniería de la Universidad Nacional de La Matanza (Civil, Electrónica, Industrial, Informática, Mecánica) donde se aplican distintas estrategias para incorporar el uso de tecnología dentro y fuera del aula. El artículo se organiza de la siguiente forma, en la sección 2 se presentan recursos que son utilizados en el aula, pudiendo ser accedidos desde cualquier dispositivo móvil (incluyendo por supuesto al teléfono celular), en la sección 3 se presentan recursos que son ofrecidos para ser utilizados fuera del aula, esto permite que el alumno tenga recursos adicionales que faciliten la práctica y el repaso de temas utilizando los dispositivos que tenga disponible. Finalmente se presentan las conclusiones alcanzadas y los trabajos futuros.

2 Recursos para el Aula Universitaria

Aprovechando la disponibilidad de los dispositivos móviles en el aula se muestran a continuación distintas herramientas para que los estudiantes puedan sacar provecho de sus dispositivos y los utilicen durante la clase.

2.1 Guías de ejercicios

Todos los materiales de la cátedra están cargados en la plataforma virtual de la Universidad Nacional de la Matanza, denominada MIEL (Materias Interactivas en Línea), por lo que cada vez es más común que los estudiantes descarguen las guías de ejercicios en formato pdf y las consulten en el aula desde sus dispositivos. También muchos docentes utilizan dropbox para compartir archivos adicionales a sus alumnos. Pero el material en formato pdf no siempre es lo más cómodo desde un dispositivo móvil, es por ello que como complemento se utiliza una plataforma especialmente diseñada para mostrar contenidos adaptados a dispositivos móviles llamada GECODIMO (GEstor de COntenidos para DIspositivos MOviles). GECODIMO como su nombre lo indica es un gestor de contenidos donde los docentes pueden cargar información para que los estudiantes puedan visualizarla en forma correcta usando sus dispositivos móviles. Permite organizar los contenidos con una jerarquía de categorías generando un árbol de navegación flexible (ver figura 1).

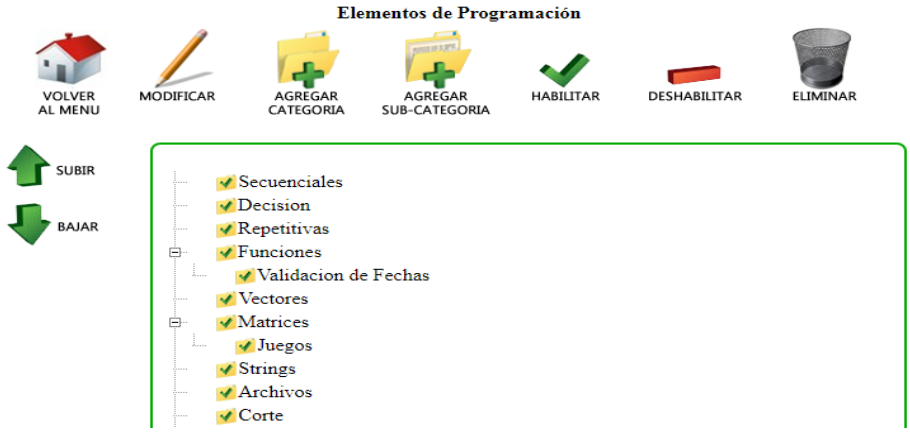


Fig. 1. Pantalla de administración de categorías

GECODIMO fue desarrollado por el grupo de investigación y se utiliza constantemente desde su construcción, con distintos fines [7]. Las páginas web que genera son livianas (ver figura 2) por lo que no requiere grandes conexiones ni gastos de datos para funcionar haciendo además el acceso muy rápido. Incluso se puede visualizar en cualquier dispositivo con acceso a internet ya que no está restringido a Smartphones. GECODIMO además de tener las guías de ejercicios, se utiliza para cargar en él material adicional propio de cada docente con el que quiere trabajar en

clase, ya que los estudiantes lo pueden consultar en forma directa y el docente puede controlar cuando la información estará visible para su consulta.



Fig. 2. Visualización de los contenidos desde un dispositivo móvil

2.2 Compilador

Los cursos de primer año en ingeniería en la Universidad Nacional de La Matanza son numerosos (aproximadamente 90 personas) y si bien la universidad dispone de laboratorios informáticos, es complicado poder llevar a todos los alumnos al mismo tiempo a realizar prácticas. Es por ello que durante las primeras unidades de la materia donde los códigos a desarrollar son cortos y además se requiere práctica para irse familiarizando con la sintaxis de programación, se les sugiere a los estudiantes descargar un compilador para sus Smartphones. Hay varias opciones gratuitas que incorporan entornos de desarrollo para el lenguaje C sobre teléfonos celulares. Un ejemplo es cppdroid [8] que brinda un entorno 100% funcional donde es posible codificar y probar los programas realizados. De esta forma, sumado a que muchos estudiantes traen también otros equipos de mayor tamaño como tablet, netbooks y notebooks, se logra sin necesidad de utilizar un laboratorio dedicado codificar los ejercicios planteados. En la figura 3 puede verse a la izquierda el editor donde se escribe el código fuente y los errores de compilación si los hay. A la derecha se muestra la ejecución de un programa de prueba donde se ingresaron datos por teclado y se calcula la suma de dos números.

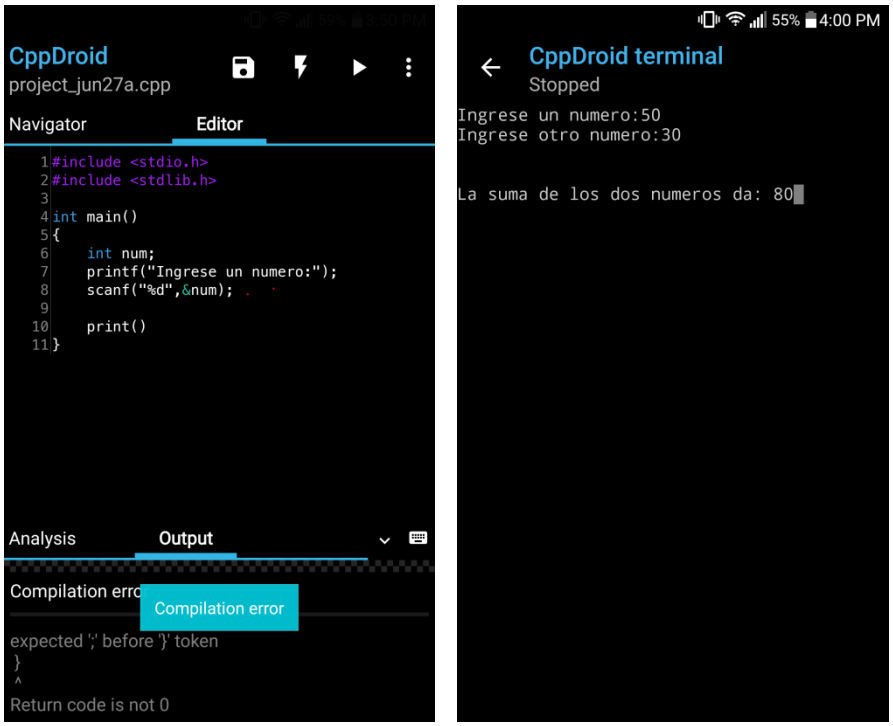


Fig. 3. Pantallas del IDE cppDroid para Android

2.3 Cuestionarios

Otra forma de fomentar el uso del teléfono celular en el aula es mediante el uso de cuestionarios. Muchas veces durante la clase se quiere conocer la opinión de los estudiantes sobre determinado tema o comprobar si los conceptos se fijaron correctamente. El docente oralmente puede consultar a los alumnos, haciendolos levantar la mano ante determinada pregunta pero muchas veces varios estudiantes no participan por vergüenza o directamente se acoplan a la respuesta en la que más gente levanta la mano. En el trabajo [9] se muestra un desarrollo previo del grupo de investigación de una aplicación basada en Raspberry PI [10] (un SBC -Single Board Computer- que tiene el tamaño de una tarjeta de crédito), que permite realizar cuestionarios a los alumnos sin necesidad de internet haciendo que puedan responderlo desde sus dispositivos de forma anónima fomentando la participación y haciendo la clase más interactiva. Además el docente puede visualizar en tiempo real el porcentaje de alumnos que opta por cada respuesta. Esto permite evidenciar las principales dificultades que se presentan en cada tema.

2.4 Servidor de Archivos (off-line)

Hay casos en que los estudiantes no tienen conexión a internet, es el caso en que la red wifi está saturada y no disponen de datos para acceder mediante línea telefónica. En este caso el docente tiene en su dispositivo los materiales en .pdf y se los pasa a los alumnos en forma directa mediante bluetooth. Este método es lento y tedioso ya que requiere enviar los archivos uno a uno y en forma individual a cada equipo. Además se encuentra limitado en ciertos equipos por ejemplo los Iphone que solo permiten compartir archivos entre sí, por lo tanto si el docente tiene un dispositivo con Android no tiene forma de pasarle en forma directa los archivos sino hay internet.

Otra forma de compartir archivos sin disponer de internet es utilizando un servidor de archivos portátil. El grupo de investigación ha diseñado un desarrollo sobre Raspberry Pi que permite configurarla a fin de compartir mediante una interfaz web los archivos que el docente traiga en un pendrive en una carpeta específica. Configurando la Raspberry como punto de acceso y servidor web (tal como lo indicado para el caso de los cuestionarios mostrado en [10]) se agregó un módulo que permite compartir archivos de forma directa de modo que todos los alumnos puedan conectarse a la red generada y descargar el material necesario directamente sin necesidad de internet.

3 Recursos fuera del espacio áulico

3.1. Plataforma Online Gamificada para Práctica

El grupo de investigación ha desarrollado una plataforma gamificada que permite resolver problemas en un ambiente online y obtener devolución inmediata mediante una corrección automática del ejercicio planteado. “Gamificar es aplicar estrategias (pensamientos y mecánicas) de juegos en contextos no jugables, ajenos a los juegos, con el fin de que las personas adopten ciertos comportamientos” [11]. “El objetivo general de la gamificación es influir y motivar a los usuarios para conseguir adquirir hábitos y alcanzar objetivos” [12].

La plataforma gamificada presenta los ejercicios de una forma distinta por medio de desafíos organizados por niveles. Un sistema de corrección automática permite asignar puntos según la efectividad en el desarrollo, detectando tanto errores de codificación como lógicos. Los puntos son el punto inicial para una serie de elementos de gamificación que se incorporan a la plataforma con el fin de captar la atención de los estudiantes brindándoles una herramienta de gran utilidad para avanzar en la materia. La plataforma se denomina “Programados” y sus detalles pueden verse en [13]. La figura 4 muestra algunas capturas de pantalla de la plataforma. Este es un recurso que se utiliza por fuera del tiempo de cursada, de forma no obligatoria. “Para mantener el interés del educando, se debe aprovechar la energía natural del alumno para sentirse capaz y orientado hacia el logro de metas. Es necesario que realice de manera voluntaria, lo que se espera que haga; y que desarrolle sus aptitudes para que alcance la meta deseada. Así, cuanto más capaz se sienta un alumno de desarrollar una actividad, más motivado estará para persistir en

ella, a su vez, le dará la sensación de éxito o de mejoría y le ayudará a mantenerse motivado" [14].

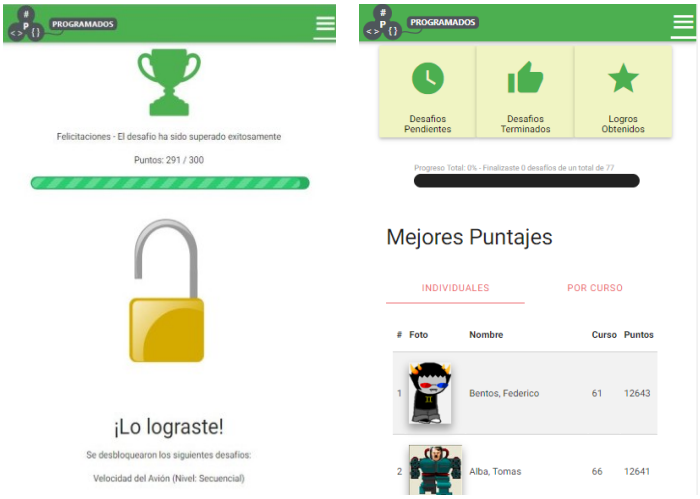


Fig. 4. Pantallas de la plataforma "Programados"

Actualmente el 71% de los cursos de la materia tiene alumnos que utilizan la plataforma, llegando al 46% el porcentaje de alumnos por curso que utilizan la plataforma en el curso con mayor cantidad de participación. Obtener estos porcentajes resulta más que significativo para una actividad que no es obligatoria. Se efectuó además una encuesta de uso para este recurso que ha sido el último incorporado. El 100% de los encuestados quisiera que en otras materias se utilizaran plataformas de este tipo. Incluyendo a aquellos que la utilizan eventualmente. Cabe destacar que la mayor parte de los alumnos que usan la plataforma lo hacen en forma semanal (ver figura 5). El 85% indica que le resulto Fácil o incluso Muy fácil, el uso de Programados (ver figura 6).

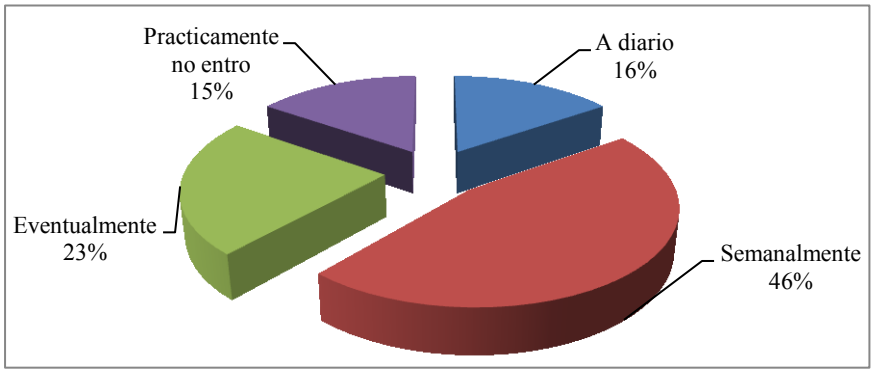


Fig. 5. Frecuencia de acceso a la plataforma

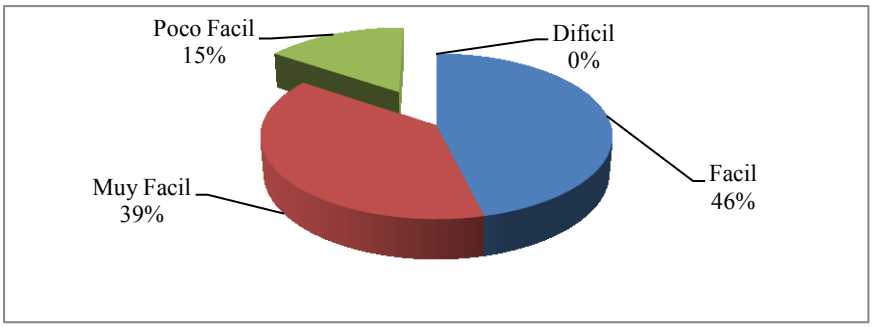


Fig. 6. Facilidad de Uso de la plataforma

3.2. Materiales Didácticos Adicionales

Los estudiantes actuales están perdiendo el hábito de la lectura. Están más acostumbrados a materiales multimedia, imágenes y videos son parte del día a día. Internet y los videos en youtube se han vuelto la principal fuente de consulta cuando un tema no es comprendido lo que trae varios problemas. En la web hay miles de recursos de fuentes no reconocidas que explican temas y cuestiones que no son correctas o simplemente no están debidamente fundamentadas. En una carrera universitaria y en especial en Ingeniería es fundamental conocer el porqué de las cosas, comprender profundamente su funcionamiento para así tener las herramientas necesarias para solucionar los problemas que se presenten con las herramientas disponibles. Los apuntes y libros de consulta son necesarios pero deben ser complementados con materiales didácticos de apoyo más cercanos a los estudiantes actuales para que puedan comprender determinados temas.

Para la materia se han desarrollado explicaciones de mini temas puntuales para que puedan ser consultados a modo de repaso. Para el desarrollo de los materiales se ha elegido powerpoint ya que con una sola fuente puede ofrecerse el material en distintos formatos adaptándose a las preferencias y circunstancias del alumno.

Los materiales se desarrollan en powerpoint con las siguientes premisas: (a) Debe contener el texto mínimo necesario para explicar los conceptos; (b) Deben agregarse explicaciones visuales del concepto a trasmitir; (c) No realizar superposiciones de contenido para que luego puedan ser impresas las diapositivas si se desea.

Una vez desarrollado el tema de esta forma, ya se puede generar una primer versión del material en formato PDF para que puedan descargarlo o generando una imagen por cada diapositiva y subiéndolas a una página web para consulta directa. En la figura 7 puede verse un ejemplo de una página de un material didáctico para el tema de búsqueda secuencial. Basado en esta primera versión se agregan animaciones para hacer más claras las explicaciones realizadas y por lo tanto el material puede ofrecerse directamente como una presentación powerpoint para que el estudiante pueda apreciar dichas animaciones. Luego a esta segunda versión se le agrega un audio pre-grabado explicando el tema y se genera un video. De esta forma con un mismo archivo fuente (el powerpoint) se ofrece el material en distintos formatos: (a) Video: material completo con animaciones y audio del docente; (b) Presentación: solo con las animaciones; (c) PDF: para descargar y llevarlo en el dispositivo móvil

teniendo una fuente de consulta al instante; (d) Secuencia de imágenes: subidas a la web generando la visualización del material directamente en línea pero sin necesidad de descarga ni de tener un software de visualización particular. Para ello se desarrolló una página web que permite visualizar las imágenes directamente desde el dispositivo móvil permitiendo desplazarse entre las imágenes mediante gestos en la pantalla táctil (swipe). Estas páginas se desarrollaron con el framework materializecss utilizando su componente carrousel [15]. En [16] puede verse un ejemplo del material cargado con este componente. De esta manera el estudiante puede seleccionar su preferencia de visualización del material. Probablemente utilice más de una forma y al principio vea el video explicativo y luego una vez adquiridos los conocimientos recurra al material visual en pdf o imágenes.

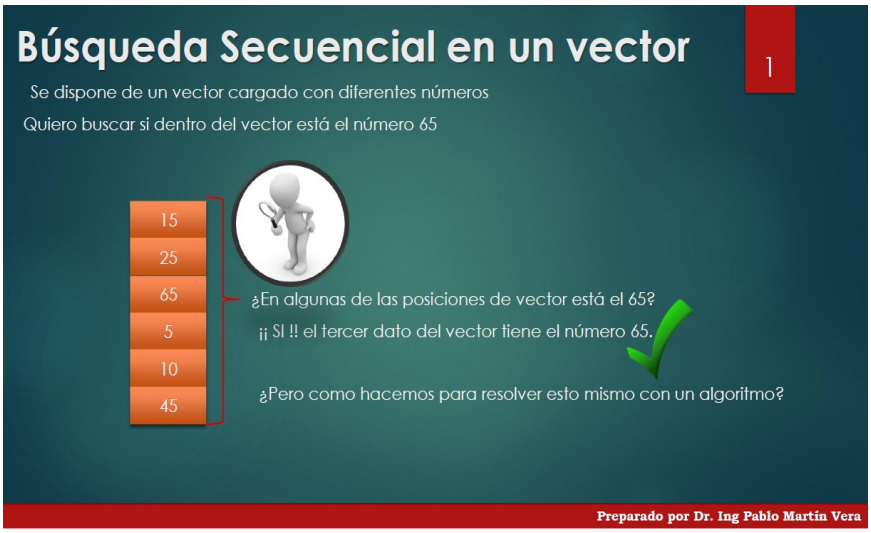


Fig. 7. Ejemplo de una página de un material didáctico desarrollado con powerpoint

4 Conclusiones

El uso de las distintas estrategias tecnológicas trae varias ventajas, que pueden desprenderse del presente trabajo, entre ellas: (a) Se aprovechan los recursos ya existentes en el aula como ser los smartphones o tablets. (b) Al utilizar el dispositivo móvil con un fin académico, la dispersión del alumno es menor, ya que el dispositivo ya se está usando por ejemplo para codificar un programa o leer un enunciado lo que hace menos probable que se utilice al mismo tiempo una aplicación no académica que pueda desviar el foco del estudiante. (c) Se usa tecnología que los estudiantes en su mayoría están habituados a utilizar porque lo que son aceptadas rápidamente. (d) El uso de compiladores móviles hace más rápida la inserción al mundo de la programación haciendo también que se puedan aprovechar tiempos de viaje o espera para codificar un programa. (e) Los materiales didácticos son una alternativa a los textos para que los estudiantes puedan repasar o terminar de fijar los conocimientos

sobre un tema dado. (f) La plataforma de práctica gamificada hace más entretenido el hecho de realizar ejercicios y motiva a los alumnos a esforzarse al competir de una forma sana con sus compañeros para lograr posicionarse en el ranking y además fomenta la auto superación incorporando metas individuales a alcanzar dentro de la misma plataforma. Creemos que esta iniciativa pude replicarse en otras instituciones comprendiendo las ventajas que aporta aprovechar el uso de los dispositivos con los que cuentan los alumnos en vez de restringirlo para evitar distracciones.

Referencias

1. Rojas, Hubert James Mendoza, and Maritza Dorila Placencia Medina. "Uso docente de las tecnologías de la información y comunicación como material didáctico en Medicina Humana." Investigación en Educación Médica (2017).
2. Sánchez, Beatriz Herrera. "El uso de los teléfonos móviles, las aplicaciones y su rendimiento académico en los alumnos de la DES DACI." Revista Iberoamericana para la Investigación y el Desarrollo Educativo ISSN: 2007-261912 (2015).
3. Proyecta, ¿El móvil en el aula? Ideas, ventajas, retos y posibilidades. http://www.plataformaproyecta.org/es/recursos-educativos/el-movil-en-el-aula-ideas-ventajas-retos-y-posibilidades
4. Robledo Jhoanna. Dispositivos móviles para el aprendizaje: Lo que usted necesita saber. (2012). https://backend.edutopia.org/sites/default/files/pdfs/guides/edutopia-guia-aprendizaje-dispositivos-mobiles-espanol.pdf
5. UNESCO. Aprender mediante el uso de dispositivos móviles (2013) http://www.unesco.org/new/es/education/themes/education-building-blocks/teacher-education/single-view/news/learning_with_mobile_technologies/
6. Figueras-Maz, Mònica, Joan Ferrés, and Julio César Mateus. "Percepción de los/as coordinadores/as de la innovación docente en las universidades españolas sobre el uso de dispositivos móviles en el aula." Revista Prisma Social 20 (2018): 160-179.
7. Giulianelli, Daniel Alberto, et al. "Desarrollo e implementación de GECODIMO (GEstor de COntenidos para DIspositivos MÓviles) para municipios." XIII Workshop de Investigadores en Ciencias de la Computación. 2011.
8. CppDroid – C/C++ IDE https://play.google.com/store/apps/details?id=name.antonsmirnov.android.cppdroid&hl=es_AR
9. Rodríguez, R. A., Vera, P. M., Giulianelli, D. A., & Cammarano, P. (2017). Implementación con Raspberry PI de un servidor portátil de contenidos. In XXIII Congreso Argentino de Ciencias de la Computación (La Plata, 2017).
10. Raspberry Pi 3 Model B, https://www.raspberrypi.org/products/raspberry-pi-3-model-b/
11. José Luis Ramírez C. Gamificación: mecánicas de juegos en tu vida personal y profesional. Alfaomega, 2014.
12. Morales Artero, Juan José. "La gamificación en la universidad para mejorar los resultados académicos de los alumnos." (2013): 1-15.
13. Vera, P. M., Moreno, E. J., Rodríguez, R. A., Vázquez, M. C., & Vallés, F. (2016, September). Aplicación de técnicas de gamificación para la enseñanza de programación a alumnos de primer año de ingeniería. In XI Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2016).
14. Hernández, Ana Polanco. La motivación en los estudiantes universitarios. Revista Actualidades Investigativas en Educación, 2011, vol. 5, no 2
15. Materialize, Componente Carousel. https://materializecss.com/carousel.html
16. Gidfis, Material didáctico de Busqueda Secuencial, http://gidfis.unlam.edu.ar/GIDFIS/programacion/BusquedaSecuencial.html

CadiDF una Herramienta de Soporte para la enseñanza-aprendizaje del Proceso de Normalización de Bases de Datos

Nadia Ramos[1], Rodolfo Bertone[2], Pablo Thomas[2], Germán Tejero[1],

[1] Universidad Nacional de Tierra del Fuego, Instituto de Desarrollo Económico e Innovación, Hipólito Irigoyen 880, 9410 Ushuaia, Tierra del Fuego, Argentina

[2] Universidad Nacional de La Plata, Facultad de Informática, Instituto de Investigación en Informática LIDI (III LIDI), 50 y 115 1er Piso, 1900 La Plata, Buenos Aires, Argentina

{npramos, ctejero}@untdf.edu.ar, {pbertone, pthomas}@lidi.info.unlp.edu.ar

Resumen. Las tecnologías informáticas y de la comunicación son utilizadas para la construcción del aprendizaje y el perfeccionamiento de habilidades, redefinen el rol del docente y de los estudiantes, mejorando todo el proceso educativo.
En las carreras de la disciplina Informática, no puede concebirse la enseñanza sin la transversalidad de estas tecnologías. En los últimos años, siguiendo una tendencia que se aplica en todos los ámbitos de la educación, los cursos vinculados a Bases de Datos buscan apoyarse en el uso de TICs a fin de favorecer los procesos de enseñanza y de aprendizaje.
En la asignatura introductoria a las Bases de Datos que se dicta en la Universidad Nacional de Tierra del Fuego (UNTdF) se desarrollan temas entre los que están presentes las formas normales y el proceso de normalización de relaciones basado en dependencias funcionales. Para ésto, no se contaba con una herramienta de libre acceso, con una interfaz amigable, y que permita asistir a los alumnos en la identificación y aplicación de los algoritmos vinculados a su estudio.
En este trabajo se presenta el desarrollo de un prototipo "ad-hoc" llamado CadiDF, que permite trabajar sobre conjuntos de dependencias funcionales en una Base de Datos, la forma normal que se halla y su proceso de normalización.

Palabras clave: Enseñanza y aprendizaje de Bases de Datos, enseñanza y aprendizaje del proceso de normalización. Herramienta educativa. Formas normales. Bases de Datos.

1 Introducción

Las herramientas o plataformas informáticas y los diversos medios de comunicación masivos, pueden utilizarse como soporte para la mejora de los procesos de enseñanza-aprendizaje. Estas herramientas modifican el rol del docente, posicionándolo no sólo

como un gestor del conocimiento sino en el rol de creador del mismo, y en su vinculación con los alumnos [5].

El rol que desempeña el docente, como así también el alumno, el cambio en las fuentes de conocimiento y la didáctica, presuponen un nuevo paradigma en la educación. Este paradigma está basado en la optimización del proceso [5][6].

En una carrera relacionada con los sistemas informáticos resulta difícil contemplar el rol del docente como un gestor del conocimiento sin el apoyo de una herramienta informática, sobre todo en una asignatura referente a Bases de Datos.

Comunicar el conocimiento hacia los alumnos no sólo se circunscribe al horario de clases. El alumno debe afianzarlo también fuera de este ámbito, por lo que muchas de las dudas se plantean fuera del mismo. Actualmente la mayoría de las personas tienen acceso tanto a una computadora como a Internet para poder obtener conocimiento. Pero la apropiación de ese conocimiento no es tarea fácil, ya sea por la dificultad en su interpretación o por la falta del marco adecuado.

La utilización de herramientas informáticas como asistentes en el proceso de enseñanza-aprendizaje brinda un apoyo de suma importancia en el proceso educativo, permitiendo al alumnado afianzar el conocimiento adquirido durante los horarios de clase.

Más específicamente, la herramienta presentada en este trabajo brinda un soporte a los alumnos en la práctica de los conceptos de normalización de relaciones en una Base de Datos (BD). El producto desarrollado brinda un soporte al rol docente fuera y dentro de los horarios de clases.

El resto de este artículo se organiza del siguiente modo; en el capítulo 2 se presentan las ventajas de la utilización de las herramientas informáticas en el proceso de normalización de BD en contraposición al método tradicional, en el capítulo 3 se analizan cuatro herramientas web disponibles en el mercado para la enseñanza del proceso de normalización de BD. Luego en los capítulos 4 y 5 se presenta un breve marco teórico y la realización del prototipo CadiDF, respectivamente. Finalmente se presentan las conclusiones y trabajo futuro.

2 Enseñanza de la normalización

El proceso de normalización de una Base de Datos es un tema que se encuentra en la currícula de cualquier curso referido a Bases de Datos Relacionales en las instituciones de educación superior.

Los métodos de enseñanza tradicionales se han utilizado para presentar el proceso de normalización de las Bases de Datos. Con la creciente incidencia de los medios informáticos y comunicacionales, se brinda a los alumnos recursos de aprendizaje alternativos y visualmente atractivos, que pueden mejorar el proceso de enseñanza-aprendizaje.

Para las nuevas generaciones de alumnos, el uso de tecnologías no sólo les proporciona ambientes más orientados a lo visual, sino que se responsabilizan de su

aprendizaje. Estos hechos, no hacen que el método de enseñanza a través de la utilización de las tecnologías reemplace a la metodología tradicional, sino que funciona como un complemento y como una extensión al rol docente.

Aquellos alumnos que todavía no han tenido un contacto práctico con BD les puede resultar complejo entender el concepto de anomalías. Puede ser comprensible la forma que estas anomalías impactan con el diseño de BD, pero no han tenido la posibilidad de experimentar los inconvenientes que se producen con modelos que las tengan. Es por ello que la enseñanza de la teoría de normalización de BD se hace bajo conceptos que son abstractos y difíciles de comprender.

En la enseñanza tradicional hay una interacción con el docente. Se utilizan recursos clásicos como proyector y pizarrón para la explicación del tema. Se presentan al alumno ejercicios que deberán resolver durante el horario de clases o en sus horarios de estudio. Luego en el resto de la práctica, los alumnos realizan ejercicios teniendo al docente disponible en caso de surgir inquietudes.

Cuando el alumno resuelve ejercicios fuera del horario de clase, las dudas deberán ser aplazadas hasta el siguiente encuentro presencial o consultadas mediante otra vía: internet o libros disponibles en biblioteca.

Una de las mayores ventajas del método tradicional descripto es la interacción entre el docente y el alumno. Pero la enseñanza asistida con herramientas informáticas permite el acceso del docente al ámbito extra académico del alumno extendiendo sus funciones más allá de los límites físicos del aula y los horarios de clase.

En el caso del proceso de normalización de BD, los métodos visuales impartidos por una herramienta de software permiten a los alumnos apropiarse de los conceptos vistos en clase. El alumno puede verificar la correctitud de sus ejercicios y mejorar el ritmo de aprendizaje.

Las herramientas utilizadas aportan flexibilidad, pero no debe ser un reemplazo del rol docente. El alumno puede perjudicarse al trabajar aislado de sus compañeros y del docente, es por ello que las herramientas deben ser complementarias.

Más allá del método de impartir conocimiento que se utilice, tradicional o asistido con herramientas informáticas, la enseñanza de temas vinculados a BD continúa siendo un pilar de las carreras informáticas, sobre todo el diseño adecuado de BD a través del proceso de normalización.

El concepto a formalizar a través del prototipo CadiDF presentado en este trabajo, no necesita la implantación de un “ecosistema complejo” que almacene y manipule los datos. Estos pueden ser almacenados temporalmente dentro del navegador y utilizados durante el tiempo que permanezca el alumno utilizando la aplicación.

Al descartar la necesidad de almacenamiento de datos en una BD propiamente dicha, ya no es imperioso desarrollar un software back-end que permita manipular esos datos y administrar los accesos. Por ello se utilizó como almacenamiento temporal el store proporcionado por el navegador, no sólo para la relación sino para los cálculos derivados de la interacción con esta herramienta.

Esta aplicación web es una Single Page Application (SPA) generada con la librería ReactJS, aplicando el patrón Flux, a través del uso de Redux para almacenar los datos ingresados por el usuario y los resultados de los algoritmos de normalización en el

store. Una arquitectura SPA está compuesta por código HTML, CSS y JavaScript que se carga en un navegador. Carga el código necesario en una única página, que se actualiza en forma dinámica, en la medida en que es requerida. [7]

Por las características propias de la aplicación, no necesita implementación del lado del servidor. Únicamente se utilizará un servidor HTTP para aceptar la petición de la página en que se encuentra la aplicación.

En CadiDF sólo es necesaria la interacción con el usuario, éste no busca persistir los ejercicios ingresados. Se busca resolver el ejercicio en el momento. Esto hace que no haya separación entre front-end y back-end. Simplemente se elimina el back-end y se trabaja la resolución del problema del usuario en el cliente, es decir en el navegador del usuario.

3 Herramientas para la enseñanza del proceso de normalización

Si bien existe una gran cantidad de herramientas para la normalización de Bases de Datos, este proyecto tiene como propósito la creación de un software propio, orientado a mejorar el proceso de enseñanza aprendizaje.

En el desarrollo realizado se describen los conceptos básicos del proceso de normalización a partir de un conjunto de dependencias funcionales. El objetivo consiste en analizar las dependencias entre los atributos y la partición de entidades. No se busca obtener diseños correctos a partir de un diagrama de entidad interrelación.

Para el presente trabajo se analizaron algunas herramientas que son gratuitas y orientadas a la enseñanza, pero sólo se limitará a una interfaz web.

Las herramientas analizadas, fueron:

- Griffith University – Normalization Tool
- Tool for Database Design
- Relational Database Tools
- LDBN – Learn Database Normalization

No se evaluaron criterios de desarrollo o de cálculo, sino la forma en que se distribuye la información y cómo se muestra al usuario. Asimismo una característica importante es que la interfaz de la herramienta se vea en una forma amigable en los dispositivos móviles, esto considerando la masividad de estos dispositivos y su portabilidad.

El resultado del análisis de las cuatro herramientas mencionadas, ha considerado características que algunas de ellas poseen y funcionalidades que no incluyen, las cuales han sido incorporadas al prototipo desarrollado. Las características analizadas se presentan en la *Tabla* 1, a través de un cuadro comparativo.

Característica	Herramientas			
	Griffith University – Normalization Tool	**Tool for Database Design**	**Relational Database Tools**	**LDBN – Learn Database Normalization**
Carga intuitiva de atributos y dependencias	SI	NO	NO	SI
Repositorio de ejercicios	SI	NO	NO	SI
Descarga de ejercicio	SI	NO	NO	NO
Explicación de proceso	SI	NO	SI	NO
Incluye hasta Forma Normal Boyce Codd	SI	NO	SI	SI
Permite saber en que Forma Normal esta la relación	SI	NO	SI	NO
Interface amigable	SI	NO	NO	NO
Idioma Español	NO	NO	NO	NO
Diseño adaptable a móvil y tablet	SI	NO	NO	NO
Código disponible	NO	NO	SI	SI

4 Marco teórico para el desarrollo de CadiDF

El modelo relacional de BD plantea un conjunto de relaciones que representan la información del mundo real. Su representación se realiza a través de las relaciones y permite la manipulación de los datos en forma de actualización y consulta.

Un esquema de relaciones correctamente diseñado, no presenta anomalías de actualización. Las anomalías de actualización pueden encontrarse como repetición de datos o inconsistencias entre datos que debieran ser iguales. Las operaciones de eliminación y modificación de datos consumen mayor tiempo por necesitar la búsqueda de cada repetición de los datos a modificar. Además el agrupamiento incorrecto de atributos puede generar pérdida de información.

Las formas normales, surgen para identificar la presencia de anomalías en las relaciones. Cada relación está en una determinada Forma Normal (FN) siempre que cumpla con determinados requisitos. El proceso de normalización está basado en el análisis del esquema de relaciones de la BD, las dependencias funcionales que existen entre sus atributos y la identificación de atributos especiales denominados claves.

Existen siete (7) formas normales (1NF, 2NF, 3NF, BCNF, 4NF, 5NF y DKNF). Las primeras están basadas en el concepto de dependencia funcional. La Forma Normal de Boyce-Codd (BCNF) es una versión más restrictiva a la Tercera Forma Normal (3NF) como solución a las anomalías de actualización que pueden presentar estos esquemas bajo condiciones específicas. Finalmente se definieron tres formas normales adicionales que solucionan las redundancias y anomalías que pudieran

presentarse en esquemas en BCNF, pero no están basadas en el concepto de dependencia funcional, estas son: Cuarta Forma Normal (4NF), Quinta Forma Normal (5NF) y Forma Normal Dominio-Clave (DKNF).

En un esquema de Base de Datos se pueden especificar diversas restricciones. Estas restricciones pueden ser de dominio, de clave, de integridad referencial. También existen otro tipo de restricciones de dependencias de datos, entre las cuales se encuentran las restricciones de dependencias funcionales y dependencias multivaluadas.

Las restricciones de dependencias funcionales suelen ser utilizadas para la detección de anomalías y su eliminación a través del proceso de normalización.

La teoría de normalización permite obtener estructuras de datos eficientes y evita las anomalías de actualización. Es un método formal que permite realizar un diseño flexible de relaciones, que permita extenderse al agregar nuevos atributos a las relaciones existentes y nuevas relaciones [9].

Partiendo de una relación, sea realizan sucesivas proyecciones en relaciones más pequeñas, de modo que su concatenación permite obtener la relación original.

El proceso de normalización está basado en el análisis del esquema de las relaciones, en las dependencias funcionales que surgen de dicho esquema y sus claves candidatas. Cada esquema de relación que no cumple con los preceptos de la forma normal deseada se descompone es esquemas más reducidos que sí los cumplen.

El diseño de una Base de Datos con calidad puede realizarse siguiendo alguna de las siguientes metodologías: ascendente o descendente, o también llamadas metodologías de diseño por síntesis o análisis, respectivamente.

Ambos métodos persiguen el mismo fin, están basados en el concepto de la dependencia funcional y el cálculo del recubrimiento irredundante, pero tienen diferencias entre sí. La principal diferencia radica en que el método de análisis permite obtener una relación normalizada hasta Quinta Forma Normal (5NF), en cambio el método de síntesis permite llegar a Tercera Forma Normal (3NF) sin pérdida de información, y sino hubieran varias claves candidatas y solapadas podría llegarse a Forma Normal de Boyce-Codd (BCNF). [9][13]

5 CadiDF: una herramienta de soporte para los procesos de enseñanza y de aprendizaje de la práctica del proceso de normalización

En el contexto descripto, surge la realización de un prototipo que permitiera reemplazar el rol del papel y el lápiz en la práctica del proceso de normalización de relaciones.

En este prototipo se busca integrar las características destacadas de las herramientas web disponibles actualmente en el mercado, y los comportamientos observados durante las clases prácticas de la asignatura Bases de Datos I en la Universidad Nacional de Tierra del Fuego.

Para desarrollar este prototipo de analizó el comportamiento del alumnado durante cuatro (4) ciclos lectivos. Esto permitió observar qué dispositivos utilizan durante la clase, sus técnicas de estudio, las consultas que realizan sobre el proceso de normalización de BD y su trabajo áulico.

El proceso de normalización puede realizarse siguiendo el método de análisis o de síntesis. Ambas metodologías requieren de una relación, el cálculo del recubrimiento irredundante, las claves candidatas y la forma normal en la que se encuentra esta relación. Partiendo de estas componentes, la relación universal se proyecta en relaciones más pequeñas que satisfacen formas normales más avanzadas que la forma normal en la que se encuentra la relación.

Finalmente se obtienen n relaciones en la forma normal deseada, que concaténandose representan una relación equivalente a la relación universal. Esto permite conservar la información contenida en la relación (atributos y tuplas) y las dependencias [9][13].

CadiDF discrimina el cálculo del proceso de normalización en cuatro módulos: la carga del conjunto de atributos y dependencias funcionales que conforman la relación; el módulo calculadora, donde se presenta el resultado del cálculo de los conjuntos de relaciones del proceso de normalización; el módulo tutorial, detalla cada uno de los resultados obtenidos en el módulo calculadora con sus conjuntos intermedios; y el módulo desafío que permite al alumno ingresar las respuestas que considera solución y verificar su correctitud.

5.1 Módulo calculadora y tutorial

Para utilizar los módulos de CadiDF, previamente debe ingresarse una relación en su pantalla principal. Esta pantalla permite el ingreso de un conjunto de atributos y las dependencias funcionales entre ellos, como se muestra en la Fig. 1.

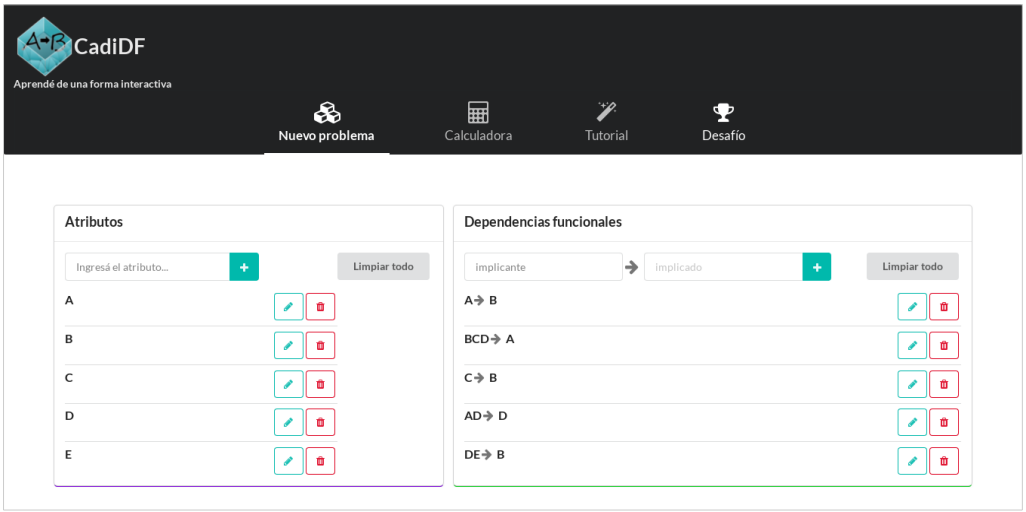


Fig. 1. Pantalla principal de CadiDF – Módulo de carga

Los módulos calculadora y tutorial están basados en los mismos algoritmos: calculan el recubrimiento irredundante, las claves candidatas y la forma normal en que se encuentra la relación para entregar un conjunto de relaciones en la forma normal deseada. En la Fig. 2 se muestra el cálculo de la clave candidata en el módulo tutorial, donde se explica su procedimiento.

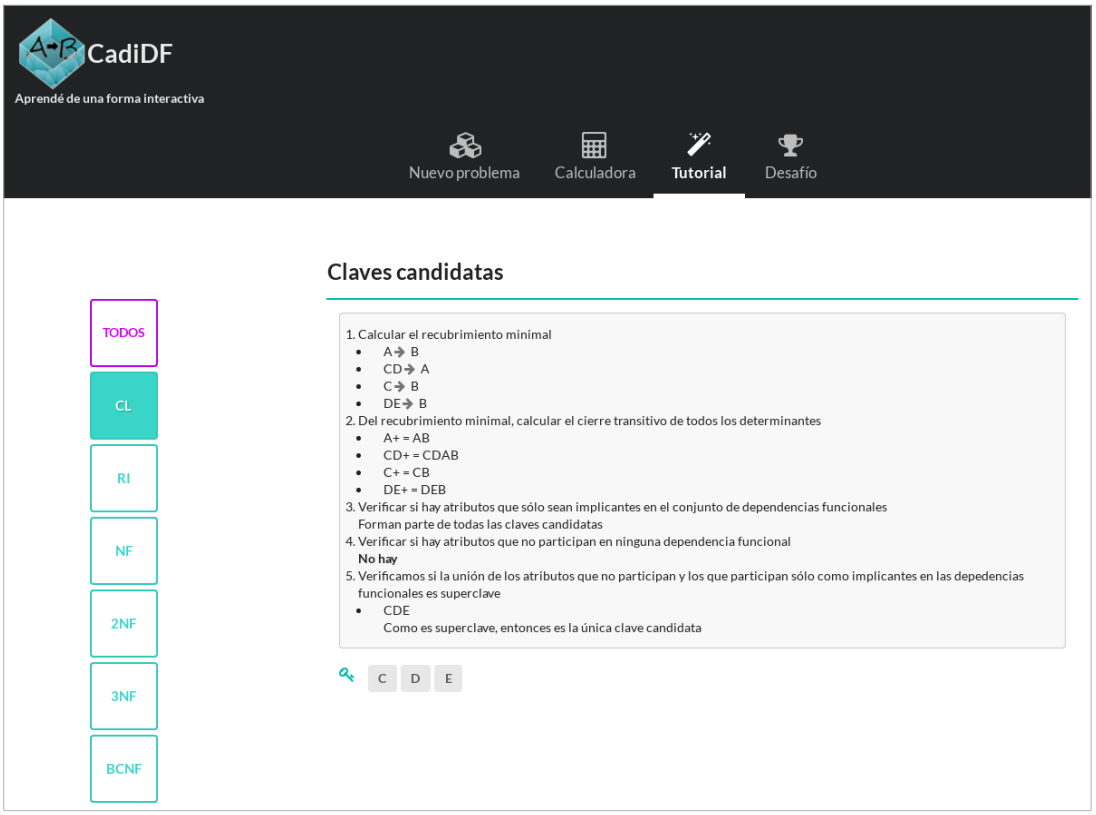


Fig. 2. Procedimiento de cálculo del conjunto de claves candidatas en CadiDF – Módulo tutorial

La posibilidad de seleccionar qué método se utilizará para normalizar la relación significó la implementación del algoritmo para dar soporte al método de análisis, realizando un proceso iterativo de descomposición de relaciones hasta lograr la forma normal deseada, y el algoritmo de síntesis que realiza un proceso de agrupación de atributos hasta llegar a tercera forma normal (3NF).

Los algoritmos para la resolución del problema, fueron seleccionados en concordancia con los contenidos teóricos de la asignatura. Esto permite a los alumnos entender el proceso que están estudiando e incorporarlo en una forma más visual.

5.1.1 Cierre de un descriptor

CadiDF utiliza el algoritmo tradicional de cálculo del cierre de un descriptor, que fuera propuesto por Ullman [15] y que recibe como entrada un conjunto de dependencias funcionales y un descriptor X.

5.1.2 Recubrimiento irredundante

El algoritmo de cálculo del recubrimiento irredundante es el que fuera propuesto por Ullman [15] y otros autores, que transforma un conjunto de dependencias funcionales elementales en un conjunto mínimo y equivalente. Este algoritmo está basado en el cálculo del cierre de un descriptor.

5.1.3 Claves candidatas

CadiDF utiliza los preceptos propuestos por Delobel para el cálculo de todas las claves candidatas de una relación [13]. Este algoritmo no utiliza el conjunto de dependencias funcionales ingresado por el usuario , sino que su entrada es el recubrimiento irredundante calculado previamente.

5.1.4 Forma normal de la relación

Para determinar la forma normal de la relación, se realizan verificaciones en orden creciente de forma normal, primero se verifica la segunda forma normal (2FN), luego la tercera forma normal (3FN) y por último la forma normal de Boyce-Codd (BCNF). No se realiza verificación alguna sobre la 1FN, dado que no pueden ingresarse atributos multivaluados en CadiDF.

5.2 Módulo desafío

Este módulo permite al alumno ingresar la solución propuesta por él y verificar su correctitud.

En virtud a que el recubrimiento irredundante de una relación puede no ser único, se calcula el cierre del conjunto de dependencias para corroborar la equivalencia con el conjunto mínimo de dependencias, y los demás algoritmos son calculados basándose en el recubrimiento ingresado.

6 Conclusiones y trabajos futuros

CadiDF proporciona a los alumnos una herramienta práctica para afianzar los conceptos teóricos concernientes al Proceso de Normalización de Bases de Datos. Permite corroborar la correctitud en las soluciones planteadas a los ejercicios prácticos, así como la descripción de los algoritmos para llegar a esas soluciones.

Este prototipo fue concebido como un recurso motivacional para el alumnado, transformándolo en hacedor de su propio aprendizaje, situándolo en una posición crítica hacia ellos mismos en la resolución de problemas.

La posibilidad del contacto con la práctica de normalización de relaciones, de estudiar el comportamiento del alumnado frente a esta disciplina han permitido y han guiado el desarrollo de CadiDF. La utilización durante el período de clases durante el segundo semestre 2018 permitirá, a través de la interacción con usuarios reales, una realimentación para la corrección de defectos, ampliar sus funcionalidades o mejorar el diseño e interoperabilidad.

Si bien CadiDF posee un diseño web responsivo, la utilización de React y Redux para su implementación permite que solo haya que reescribir partes para lograr una aplicación móvil, con React Native como librería de desarrollo en vez de React. Esto posibilita, como trabajo futuro, la conversión de CadiDF, a una aplicación móvil nativa.

Asimismo, como trabajo futuro se ha planteado a esta herramienta como un módulo de una herramienta más grande e integradora de más unidades que se dictan en la cátedra Bases de Datos I. Esto permitirá que el alumno pueda centrarse en los contenidos propios de la asignatura, en lugar de aprender a utilizar distintas herramientas.

Referencias

1. C. Domínguez Pérez, A.J. Elizondo y T.A. Pérez Fernández. "¿Todavía interesa normalizar bases de datos? Reflexionando sobre su enseñanza". XV JENUI, Barcelona, julio 2009
2. L. Ahmedi y E. Jajaga. "A database normalization tool using Semantic Web technologies", vol. 5, núm. 4, 2011, International Journal of Systems Applications, Engineering & Development
3. C. Stefanidis y G. Koloniari, (2016) "An interactive tool for teaching and learning database normalization". In Proceedings of the 20th Pan-Hellenic Conference on Informatics (PCI '16). ACM,Nueva York, artículo 18. Disponible en: https://doi.org/10.1145/3003733.3003790
4. A. Yazici y Z. Karakaya, (2007) "JMathNorm: A Database Normalization Tool Using Mathematica", International conference; 7th, Computational science, Beijing
5. UNESCO. (2010). Experiencias de formación docente utilizando tecnologías de información y comunicación. Santiago de Chile: AMF Imprenta.
6. D. Fernández Zalazar, et al. (2012). Los estudiantes universitarios y las TIC. de Moebius@cneri. Disponible en: web: http://enmoebius.com.ar/?p=1846
7. https://reactjs.org/

8. M.D. Costal Costa. El modelo relacional y el álgebra relacional. UOC/Digitalia: Universitat Oberta de Catalunya, 2001, pp. 7-31
9. R. Elmarsi y S.B. Navathe. Fundamentos de Sistemas de Bases de Datos. 5ta Ed. Madrid: Pearson - Addison Wesley, 2007
10. C.M. Ricardo. Bases de Datos. México: Mc Graw Hill, 2009, pp.123-198 11. E.F. Codd. A relational model of data for large shared data banks. IBM Research Laboratory, San José,California, Commun. ACM 13 (junio 1970)
12. J. Sánchez. (2013). Principios sobre Bases de Datos Relacionales (versión 2.4) [Online]. Disponible: http://jorgesanchez.net/manuales/viejos/gbd/gbd2012.pdf
13. A. de Miguel Castaño, M. Piattini Velthuis y E.M. Martínez. Diseño de Bases de Datos Relacionales. México: Alfaomega, 2000
14. R.C. Paré et al. Bases de Datos. Barcelona: fundació per a la Universitat Oberta de Catalunya, 2005
15. J.D Ullman y J. Widom. Introducción a los Sistemas de Bases de Datos. México: Prentice Hall, 1999
16. P. A. Bernstein. (1976). Synthesizing Third Normal Form Relations from Functional. Dependencies.

A Nomadic Testbed for Teaching Computer Architecture

Pablo D. Godoy[1,2,3], Carlos G. García Garino[1,3], and Ricardo L. Cayssials[4]

[1]ITIC, Universidad Nacional de Cuyo, Mendoza, Argentina
[2]FCEN Universidad Nacional de Cuyo, Mendoza, Argentina
[3]Facultad de Ingeniería, Universidad Nacional de Cuyo, Mendoza, Argentina
[4]Universidad Tecnológica Nacional - FRBB, Buenos Aires, Argentina
{pablodgodoy@gmail.com}

Abstract. A nomadic laboratory or testbed, based on Raspberry Pi 3 computers and Arduino microcontrollers, has been developed in order to teach subjects related to computer architecture. The testbed can be transported to the classroom. Students can access it through the available network, which can be a wireless LAN, wired LAN o a custom network. The student can access without constraints to the platforms, therefore there are a wide range of possible experiments. This laboratory was used during 2017 for practical works in the course Introduction to Technology, and during 2018 in the course Computers Architecture at Universidad Nacional of Cuyo. Some of the experiments that are been carried out by students are: to explore and analyse the architecture of the computers through Linux commands, write and run programs on different programing languages, input and output operations through memory mapped addressing and isolated addressing, write interrupt service routines in order to service interrupts, multithreading programing, explore memory maps, CPU features, etc. This paper describes the testbed architecture, experiments performed by students in the mentioned subjects, present the students feedback, and describes the possible methods in order to integrate it to a remote laboratory.

Keywords: Nomadic testbeds; Teaching platform; Computer architecture

1 Introduction

Computer architecture is a central subject in careers such as engineering or bachelor degrees in computer, information systems, electronics, information and communication technology, etc. Laboratory experiences and practical activities about computer architecture usually implies to recognize all components, their functions and features in different computer systems, such as the processors, data memories, data buses, bridges, etc., to employ operating system tools to visualize and analyze the processors workload, memory usage, peripheral status, etc., and to build programs to perform input/output operations, multithreading

task over superscalar processors or multiple core processors, etc. In order to perform laboratory experiences and practical activities, students need to access to different kind of computers, in order to understand and compare different computer architectures. As a result, it is very useful to build platforms to enable students carrying out practical works and laboratory experiences.

University computer laboratories usually have computers based on x86 processors (Intel, AMD, etc.). This limits the kinds of architectures over which students carry out experiments. As a result, it is useful to deploy platforms and facilities with different architectures, for example: ARM, in order to give students more options about different computer architectures for performing experiments.

In the literature several experimental platforms have been proposed, with educational and scientific purposes [1][2][3][4]. Horneber and A. Hergenröder [2] distinguish between two kinds experimentation platforms or testbeds, nomadic testbeds and remote testbeds. Nomadic testbed can be completely or partly movable to the place where the experiments will be performed. The main advantage of nomadic testbeds for teaching purposes is that students can put hands on the equipments. But the main disadvantage is the limited time for performing the experiments. Remote testbeds can be accessed remotely through Internet. The advantage of remote testbeds is that they can be accessed all time, but the main disadvantage is that students can not put hands on the equipments under test. Our platform can be entirely transported to the classroom in order to students carry out experiments. But also can be remotely accessed by ssh protocol. Therefore our platform have features of both kind of experimental platforms.

However, until today our platform only has been used as a nomadic testbed. As a result, we classify our platform as a nomadic testbed for teaching purposes, with capacity to be transformed in a remote and nomadic testbed.

The main contributions of this papers are two:

- The proposed architecture to build a nomadic experimental platform to perform experiments related with computer architecture based on ARM processors.
- The proposed experiments for putting in practice theoretical knowledges about computer architectures, that students acquire during theoretical classes.

2 Remote Laboratory Architecture

2.1 Equipment

The central part of the laboratory are Raspberry Pi 3 computers [5]. The election of this computer is due to they have processor of four cores, which enables multicore and multithreading experiments. These kinds of experiments are needed in the last part of the subject *Computer Architecture*, in order to students can write programs over parallel architectures. On the other hand, computer laboratories for teaching purposes usually are built with PC computers

with x86 processors. One objective of our proposed testbed is to use ARM processors, in order to extend the architecture types over which students carry out practial works. In addition, the small size of these computers are a very important feature due to the laboratory must be transported. An auxiliary board with leds and buttons is attached to the Raspberry Pi 3 computer. The leds and buttons are connected to the general purpose input/output pins of the Raspberry Pi 3, to enable performing input and output experiments

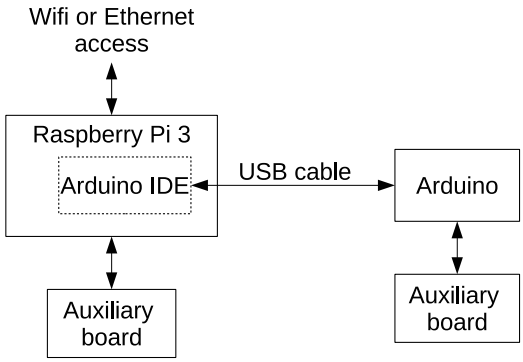
In addition, an Arduino UNO board [4, 6] is attached to each Raspberry Pi 3 computer, for enabling experiments over a simpler computer and over other kind of peripheral, like analogical to digital converters, etc. An auxiliary board with leds and buttons is attached to the Arduino board. The leds and buttons are connected to the general purpose input/output pins of the Arduino, to enable performing input and output experiments. The Arduino Integrated Development Environment (IDE) is installed on the Raspberry Pi 3 computer. Both computers are attached by a USB cable.

The equipment is completed by:

- The teacher computer: that enables monitoring the students activities through ssh or vnc protocols.
- An access point: that enables wireless access to the Raspberries.
- Ethernet cables: that enable wired access to the Raspberries.

The kind of access is selected in function of the infrastructure and network available in the classroom. The figure 1 shows a block diagram of a node.

Fig. 1. Node block diagram



2.2 Access to the Equipment Under Test

The access to the equipment under test is through SSH or VNC protocols. In some experiments students need to consult configuration files or run Linux

commands. To other experiments, students need to write and run programs. To almost all activities, both SSH and VNC can be used equally. The VNC protocol is mandatory only to employ the tools provided by the Raspberry Pi operating system. Both tools, SSH and VNC, have been used in several testbeds presented in the bibliography, for example: [7][8]. The experiences carried out during 2017 and 2018 shown that students prefer VNC, maybe due to students prefers to interact through a graphical user interface (GUI) proper of desktop operanting systems. As a result, students are encouraged to use both: SSH and VNC, in order to acquire experience in the use of a wider range of tools.

In order to access to the equipment, three methods are provided:

- Through Ethernet network.
- Through Wifi Network.
- Through a cloud services provider accessible through Internet.

The first method can be used if there is an Ethernet network deployed in the classroom. The second method, through Wifi, is always available due to the platform has an Wifi access point as a part of its equipments, and each Raspberry Pi 3 computer can be used as an access points for deploying a wireless network. The access through a cloud platform was implemented using the services of Remote-IoT [9]. Through this service, a computer can be accessed remotely through Internet, via SSH and VNC protocols. The access through a cloud platform will be investigated and analyzed in greater details in future works. The three methods were used during 2017 and 2018, in the courses of *Computers Architecture* and *Introduction to Technology*, with successful results.

3 Performed Experiments

In this section, some experiments carry out by students during 2017 and 2018 over the presented testbed are described.

3.1 Architecture Analysis

In these experiments, students analyse the architecture of a computer through Linux commands. The commands employed were:

- lshw (HardWare LiSter for Linux): tool to show information about the hardware configuration of the computer. The information shown by lshw include: (a) memory configuration, (b) CPU features and speed, (c) cache memory configuration, (d) type of buses and speeds, (e) bridges between buses, (f) massive storage devices, (g) different peripheral, etc.
- lspci: shows information about PCI buses and devices attached to these buses.
- free: shows information about memory, used and free physical and swap memory, and memory used as buffers.
- uname: displays information about the operating system (especially the options: -srpo)

– df: shows information about the hard disk (option used: -h)
– ifconfig: displays information about the network interfaces.

These commands and most of their options and output formats have been selected for performing this practical activity (see [10] for detailed information about these commands and their options). The activity carried out by the students consisted in to execute these commands, to analyze the information produced, to draw a block diagram of the computer architecture and specify their characteristics.

3.2 Input-Output Operations

The objective of this activity is to distinguish between memory mapped and isolated addressing of input/output devices [11]. The practical work consists of switching off and on leds and to read the state of input pins through two methods:

– Memory mapped I/O devices
– Isolated I/O devices

For Arduino processor, students can use both methods, and for Raspberry Pi 3 only memory mapped access is available. In order to see the differences between these two methods, students must program the processors in assembler language, since a high level language hides these details for programmers.The students write code segment in assembly language embedded in C, C ++ or Python language (known languages for students).

3.3 Writing of Interrupt Service Routines

Writing of Interrupt Service Routines in Arduino UNO The practical activities consist of writing service routines to attend interrupts caused by signals detected in input pins. For this purpose, students are provided with a guide about Arduino IDE programing (obtained from [6]) and the instruction set of the ATmega328P processor [12], that is the processor used by the Arduino UNO platform. In addition, a pre-written code that include the scheme of an Arduino program in C language and the code to embed assembler instructions are provided.

Students write service routines for the two interrupt sources of the Arduino Uno platform, the pins 0 and 1. These service routines inform when an interrupt is detected. After, students carry out a set of actions to analyze the interruption priorities.

Writing of Interrupt Service Routines in Raspberry Pi 3 A similar activity is carry out over Raspberry Pi 3 platform. Two differences with Arduino UNO activity are proposed to students:

– Write programs using Python language.

– Use the configuration files of the Linux filesystem in order to control the hardware instead of assembly language.

These changes have the purpose of give new knowledge to students.

Similar to the previous activity, students write routines of service for two interrupt sources of the Raspberry Pi 3 platform. These interrupts are triggered by signals over different input pins. These service routines inform when an interrupt is detected. After, students carry out a set of actions to analyze the interruption priorities.

In addition, students can analyze interrupts statistics through the */proc/interrupts* Linux file.

3.4 Multithreading Programing

The objective of this activity is to show to students the speed-up that can be achieved in some task using multithreading programing. For this activity, Raspberry Pi 3 computers were used due to they have processors with four cores. Because students at this points of theirs career have reduced knowledge of programing, this activity was performed using C++ language, due to it enables to write multithreading programs in an easy way. Other options (like Python, Java or C) require to use classes and objects, and students still does not have knowledge about these topics. The activity consists of writing a program that uses a 200x200 matrix of float point variables, and several subroutines that perform the following tasks:

– Modify a randomly chosen elements of the matrix with a random value every time that this function is called.
– Calculate the root mean square value of all elements of the matrix.
– Obtain the highest value inside the matrix.
– Find the row whose sum of their elements has the highest value.

After verify the proper execution of their programs, students verify some statistic of execution through Linux commands, like number of threads, cores usage, execution speed of every function, etc.

Then, students are instructed to divide their programs in several threads using the functions provided by C++ language. For simplicity, students are suggested to execute every subroutine in a different thread. After verify the proper functioning of their programs, students verify again the above mentioned execution statistics, and compare against the prior values. Students can see how the parallel execution of their functions increases the processor usage and the speed of every function.

Some students, for their own initiative, went beyond what was requested them and divided their programs in more threads (a group divided its program in 64 threads), and a competition for achieve a higher number of threads was generated between students. Students obtained higher and lower speed-ups. This result allowed showing to students that the parallelization must be performed in an efficient way (students will acquire knowledge about efficient parallelization and execution of tasks in later subjets of their career).

3.5 Interference in Wireless Networks

The objective of these experiments was to show the effect of interference on wireless networks (IEEE 802.11). For this purpose, the Raspberries Pi 3 computers were used as wireless access points of IEEE 802.11 wireless networks on infrastructure mode. Wireless access points based on Raspberry Pi 3 computers are chosen due to they enable to configure a wider range of parameters than commercial wireless access points, like the interframe time for different kind of data.

The experiments performed by students consisted of measuring the following parameters:

- Data rate while a large size file is transfered.
- Latency for a ping command during the transference.

These parameters were measured under the following conditions:

- Three wireless access points in different wireless channels.
- Three wireless access points in the same wireless channel with time interframe values similar to commercials wireless access points.
- Three wireless access points in the same wireless channel with time interframe values modified a their minimum value [13].

The measurements were performed while three users were transfering big size files between the wireless access points to their computer through SSH, being connected every user to a different wireless access point. The values of the above mentioned parameters were obtained through the values displayed by the ping and SCP (secure copy protocol) commands.

Students can see how the channel occupation affect the individual data rate and latency. The methodology of these experiments was obtained of a similar experiment, carried out with scientific purposes, using a wireless sensor network [14].

4 Answer to an Anonymous Survey about the use of the Testbed

Eighteen students that took the course "Technology Introduction" of the Computer Science career at Universidad Nacional de Cuyo, were invited to answer an anonymous survey about the use of the platform. The questions of the survey and the students answers were:

1. ¿Was the proposed platform useful in your process of acquiring new knowledge?
   Answers:
   (a) It was not useful: 0%
   (b) It was useful: 20%
   (c) It was very useful: 80%

2. ¿Were the experiments interesting and amused?
   Answers:
   (a) The experiments were boring and no interesting: 0%
   (b) The experiments were interesting and amused: 10%
   (c) The experiments were very interesting and amused: 90%
3. ¿Were the experiments easily to perform and could you understand the theoretical concepts behind each activity?
   Answers:
   (a) The experiments were complicated and I could not understand the concepts: 0%
   (b) The experiments were easy to follow and I could understand the concepts: 20%
   (c) The experiments were easy to follow and I could understand very well the concepts: 80%

In addition, student were invited to make suggestions and comments. Some of these sugerencias and comments (literally transcripted) were:

- "*It was very amused and motivating*".
- "*It was a very good experience*".
- "*It would be good to have more time to work with the platform*".

5 Future Work

Remote laboratories, also known as remote testbeds, are platforms that allow remote access to different types of equipments, devices, laboratories, etc. through a LAN networks or Internet. They are complex systems that include a large number of components with functions such as: access control, monitoring of activities, information storage, etc. They may be aimed at scientific research, application development, training, teaching, etc. [3, 15, 16].

Cloud computing is a model for allowing on demand and ubiquitous access to a configurable, virtualized and shared computational resources set (for example networks, servers, storage, applications and services), that can be quickly provided and released with minimal management effort and interaction with the cloud services provider. These resources are offered by cloud computing services providers to external customers through Internet as web services, based on negotiated agreements between the service provider and customers [17].

The immediate future work will be to integrate the proposed testbed to a cloud computing platform, similar to [18, 19]. The objective of this integration is to provide access to the testbed all time through Internet, overcoming the time limitation for accessing the nomadic testbed mentioned by students.

6 Conclusions

The experiments performed by students during 2017 and 2018 show the versatility of the proposed platform for performing different kind of experiments, including:

- Analyze the internal architecture of different computers.
- Perform input and output operations and to write interruption service routines.
- Basic parallelization of tasks.
- Analysis of wireless networks behavior.

During the experiments it can be noted that students adopted very well the proposed educational testbed, and they wanted to exploit it adding activities different to the proposed ones. In addition, the feedback obtained from students through an anonymous survey was very positive. This shows the motivation of students to work with new and different technology. For this reason, it is needed to increase the number of nodes of the platform and to add a mechanism that enable remote access to the platform. In this way, students would be able to access the platform from their homes.

Acknowledgment

The authors acknowledge the financial support received from the *Universidad Nacional de Cuyo* through the project *B041 "Implementación de laboratorios remotos basados en cloud computing"*. Also, authors acknowledge to the professors *Lucas Iacono* and *Osvaldo Marianetti*, professors of *Introduction to Technology* and *Computer Architecture* respectively, at *Universidad Nacional de Cuyo*, for allow us to use our platform on their courses.

References

1. L. Steyn and G. Hancke, "A survey of wireless sensor network testbeds," in *AFRICON, 2011*, pp. 1 –6, sept. 2011.
2. J. Horneber and A. Hergenröder, "A survey on testbeds and experimentation environments for wireless sensor networks," *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, pp. 1820–1838, 2014.
3. P. Orduna, A. Gomez-Goiri, L. Rodriguez-Gil, J. Diego, D. Lopez-de Ipina, and J. Garcia-Zubia, "wCloud: Automatic generation of WebLab-Deusto deployments in the Cloud," in *Remote Engineering and Virtual Instrumentation (REV), 2015 12th International Conference on*, pp. 223–229, Feb 2015.
4. M. Hakima, B. Abdelhalim, and B. A. Adda, "Design of a low cost remote electronic laboratory suitable for low bandwidth connection," *Computer Applications in Engineering Education*, vol. 25, no. 3, pp. 480–488.
5. Raspberry Pi Foundation website, "Raspberry pi 3 single-board computer." Retrieved in 2018.
6. Arduino s.r.l, "Arduino uno rev3 documentation." Retrieved in 2018.

7. M. N. Kabiri and M. Wannous, "An experimental evaluation of a cloud-based virtual computer laboratory using openstack," in *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, vol. 00, pp. 667–672, July 2017.
8. T. Gerhard, D. Schwerdel, and P. Müller, "A networkless data exchange and control mechanism for virtual testbed devices," in *Testbeds and Research Infrastructure: Development of Networks and Communities* (V. C. Leung, M. Chen, J. Wan, and Y. Zhang, eds.), (Cham), pp. 14–22, Springer International Publishing, 2014.
9. Remote-IoT, "Remote-iot website." Accessed in 2018.
10. Canonical Ltd, "Ubuntu manpages website." Accessed in 2018.
11. W. Stallings, *Computer organization and architecture designing for performance*. Pearson, 9 ed., 2013.
12. Microchip Technology Inc., "DS40001984A: Complete Datasheet of the ATmega328/P." Accessed in 2018.
13. IEEE Institute, "IEEE std 802.15.4-2003," pp. 1–670, 2003. Retrieved in 2015.
14. P. D. Godoy, R. L. Cayssials, and C. G. G. Garino, "Communication channel occupation and congestion in wireless sensor networks," *Computers & Electrical Engineering*, 2018.
15. P. Godoy, R. Cayssials, and C. Carcía Garino, "A WSN testbed for teaching purposes," *IEEE Latin America Transactions*, vol. 14, pp. 3351–3357, Jul. 2016.
16. M. Waldrop, "Campus 2.0," *Nature*, vol. 495, no. 7440, pp. 160–163, 2013.
17. R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599 – 616, 2009.
18. P. Godoy, R. Cayssials, and C. García Garino, "A cloud based WSN remote laboratory for user training," in *Proceedings of TE&ET (Technology in Education and Education in Technology)* (U. of Morón, ed.), 2016.
19. P. Godoy, R. Cayssials, and C. García Garino, "Laboratorio remoto para la formación de usuarios basado en el cloud," *Revista Iberoamericana de Educación en Tecnología y Tecnología en Educación*, pp. 7–18, 12 2016.

Análisis de uso de un juego de realidad aumentada en la creación del hábito de la caminata en niños con necesidades especiales

Cruz Alejandro[1,3] , Acosta Nelson[2,4]
[1] Universidad Nacional de Chilecito, Chilecito - La Rioja
[2] Facultad de Ciencias Exactas, UNICEN, Tandil - Buenos Aires
[3] Becario Doctoral Agencia
[4] Departamento de Ingeniería, UNTREF, Caseros - Buenos Aires
acruz@undec.edu.ar, el.nelson.acosta@gmail.com

Abstract. El estilo de vida sedentario es peligroso para la salud, y jugar videojuegos en muchos casos contribuye a este problema. La evolución de los teléfonos inteligentes y el avance de la tecnología de realidad aumentada, permiten el desarrollo de nuevos juegos combinando la diversión y el entretenimiento con el ejercicio físico. Este trabajo presenta un análisis del comportamiento de los usuarios de un juego de realidad aumentada para incentivar la actividad física en los niños, especialmente aquellos con diagnósticos dentro del espectro autista, diabetes, parálisis cerebral, sobrepeso, obesidad infantil; donde el sedentarismo tiene graves consecuencias.

Keywords: Juegos Serios, Exergaming, Juegos, Realidad Aumentada, Discapacidad.

1 Introducción

La inactividad física se ha convertido en uno de los problemas de salud más importante tanto en los países desarrollados como en los países en desarrollo. Según la Organización Mundial de la Salud (OMS), el 60% de la población mundial no realiza actividad física para obtener beneficios para la salud y el 6% de las defunciones a nivel mundial son ocasionadas por la inactividad física, siendo superada solamente por la hipertensión, el consumo de tabaco y el exceso de glucosa en la sangre [1].

La actividad física influye sobre el funcionamiento psicológico de los individuos [2]. El ejercicio activo y habitual influye positivamente en la calidad de vida de las personas que padecen diversas condiciones de salud. La mayoría de las personas con discapacidades o características especiales pueden beneficiarse de la práctica del ejercicio físico, adaptando la misma a los diferentes diagnósticos de cada individuo [3].

La inactividad física durante los primeros años de vida es un factor que colabora en el incremento de los niveles de obesidad y de otros trastornos médicos que se observan en niños, niñas y adolescentes [4, 5].

La naturaleza de las actividades de entretenimiento de los niños ha cambiado drásticamente en las últimas décadas. En el pasado, la infancia dedicaba gran parte de su tiempo de ocio a practicar juegos activos al aire libre. En la actualidad, la evolución de la tecnología y la aparición de consolas de juegos, pc y teléfonos inteligentes cada vez más potentes, ha provocado que los niños dediquen una parte mucho mayor de su tiempo libre a actividades de tipo sedentarias.

El estilo de vida sedentario se ha vuelto bastante común hoy en día y jugar videojuegos, se puede considerar como uno de los comportamientos que contribuyen a este problema, así como a muchos otros problemas de salud. La industria de los videojuegos cada vez cobra una mayor importancia en el ámbito del ocio.

Durante los últimos años, el uso de dispositivos móviles se ha incrementado considerablemente. La ventaja de movilidad de los teléfonos inteligentes combinada con el uso de tecnologías como la Realidad Aumentada (RA), permite considerar a estos dispositivos como un medio para incentivar la actividad física en niños que lleven un estilo de vida predominantemente sedentario.

Los juegos móviles que se basan en el uso de la RA permiten ampliar el universo del juego a través de modificaciones virtuales en el mundo real, incentivando el movimiento por lo que los juegos dejan de ser sedentarios y donde la interacción con el juego se realiza en el mismo mundo donde el participante se mueve habitualmente.

La combinación de la RA y los teléfonos inteligentes modernos permiten desarrollar juegos que requieran que el jugador tenga que caminar, recorrer y explorar determinadas áreas, ayudando de esta forma a aumentar el nivel de actividad física. Caminar es una forma común, accesible y económica de actividad física. Es aeróbico y estimula el uso de grandes músculos esqueléticos, proporcionando numerosos beneficios para la salud [6]. Por lo tanto se debe aprovechar las ventajas del uso de esta tecnología para incentivar el hábito de la caminata a través de videojuegos, mejorando el bienestar de las personas que los juegan.

2 Antecedentes

La industria de los videojuegos es una industria dedicada al ocio. Sin embargo en los últimos años ha tomado una importante relevancia investigaciones relacionadas a la creación de videojuegos orientados a la salud. Si bien estos juegos, en la mayoría de los casos, no tienen el presupuesto ni los recursos comparados con los juegos comerciales que encontramos en el mercado, logran aportar al jugador además de diversión y entretenimiento, algún tipo de beneficio o información sobre la salud de las personas.

2.1 Serious Games

Un serious games es un juego en el que la educación, en sus diversas formas, es el principal objetivo [7]. Estos juegos fomentan el aprendizaje y los cambios conductuales [8]. De una u otra manera, cuando se diseñan este tipo de videojuegos, se persigue la educación del jugador, tanto en la adquisición de hábitos saludables, como en la comprensión y entendimiento de procesos patológicos complejos.

Investigadores en el campo de la salud han comenzado a aprovechar el poder de los videojuegos para motivar a los pacientes y, en última instancia, para mejorar sus resultados de salud [9].

Con respecto a los videojuegos relacionados a la mejora de la salud, encontramos aquellos cuya finalidad no es sólo enseñarles a los pacientes los procesos que se están llevando a cabo dentro de su cuerpo, sino también reducir el estrés o la ansiedad derivados de los tratamientos. Estos videojuegos tratan de transmitir la desmitificación de las enfermedades, convirtiéndolas en algo más natural y menos monstruoso de lo que a menudo nos parecen, logrando de esta manera una mejor aceptación de la enfermedad por parte de los pacientes. Algunos que podemos mencionar son:

- Re-Mission [10], fue creado específicamente para jóvenes con cáncer infantil. Los jugadores pilotan a Roxxi, un nanobot que viaja a través de los cuerpos de los pacientes con cáncer para destruir las células cancerosas, luchando contra las infecciones bacterianas y la gestión de efectos secundarios derivados del cáncer y su tratamiento.
- Zak vs Mutantes, el protagonista del juego, Zak y tiene diabetes. Durante la aventura que vivirá Zak, necesitará vigilar sus niveles de glucosa en sangre, comer para recuperar energías cuando lo necesite, hacer un poco de ejercicio y encontrar a los mutantes que invaden la ciudad. El juego pretende que tanto los chicos con diabetes como sus familias sean conscientes de que pueden vivir y jugar como los demás chicos si se hacen responsables de su enfermedad [11].

Dentro de los videojuegos relacionados a la prevención y promoción de la salud, encontramos aquellos que nos acercan a conductas y conocimientos sobre salud:

- DigesTower, tiene como escenario el sistema digestivo humano. El juego está enfocado para niños en edad escolar. El personaje se llama Elise. El objetivo principal del juego es comprender la importancia de una alimentación saludable [8].
- Virtual Drug Use, intenta hacer conscientes a los jóvenes de los problemas de la adicción a las drogas, en concreto a la cocaína. El juego permite la opción de correr una carrera de motos, con drogas o sin drogas. Los investigadores propusieron el siguiente modo de juego: primero jugar una carrera tomando drogas. La ventaja de este modo es que eres casi invencible (no te puedes caer de la moto, saltas a enorme altura, no hay sensación de peligro) pero cada poco tiempo hay que parar a tomar una dosis, lo que interrumpe el juego dificultando la carrera. La segunda partida será sin consumir drogas, lo que permite al jugador medir su propia habilidad, ya que el motociclista se puede caer de la moto y no conduce tan deprisa, pero el juego es continuo [12].
- Pamoja Mtaani, centra sus esfuerzos en la prevención y control de enfermedades de transmisión sexual. Pamoja Mtaani (Juntos en el barrio) es un videojuego multijugador para PC basado en un mundo 3D abierto para jugar on-line. Contiene mensajes de prevención sobre el virus del SIDA que pretenden cambiar el comportamiento de los jóvenes frente a esta enfermedad [13].

Los dispositivos móviles y la RA pueden ayudar en el proceso de aprendizaje al crear entornos más interactivos y atractivos. La RA es una tecnología que tiene la capacidad de captar la imaginación de los niños permitiéndoles interactuar con elementos virtuales en el mundo real. Existen numerosas investigaciones relacionadas con la RA principalmente en el sector salud:

- En [14] investigadores desarrollaron un juego de realidad aumentada para dispositivos compatibles con el sistema operativo Android, que permite apoyar la educación terapéutica para niños (de 5 a 14 años) con diabetes. El juego ayuda a los niños a aprender el contenido de carbohidratos de diferentes alimentos, mostrando comidas virtuales en un plato real.
- En [15] se desarrolló un GameBook llamado "Tobias en el zoológico", utiliza la tecnología de la RA con el objetivo de ayudar a niños con trastornos del espectro autista a reconocer y adquirir emociones. El GameBook contiene una historia que se puede leer o escuchar, describiendo algunos escenarios con los cuales el niño debe interactuar. El juego se puede jugar en cualquier dispositivo móvil (compatible con Android) o una computadora portátil, ya sea con una cámara web externa o incorporada.

Una característica de los juegos mencionados anteriormente, es que no están orientados al movimiento del participante sino más bien a un juego pasivo, de movimiento reducido que obliga al jugador a estar sentado para ejecutarlo imposibilitándolo de explorar su entorno y objetos que lo rodean.

2.2 Exergaming

El término "ExerGaming" consiste en dos palabras fusionadas: ejercicio y juego. El rol de los exergames no es sólo proporcionar una forma entretenida de entrenar y facilitar la actividad física, sino también promover un estilo de vida saludable y crear conciencia sobre la salud y la condición física [16].

Dance Dance Revolution (DDR) fue presentado por Konami en 1998 y es el juego a partir del cual se considera que se popularizó el termino exergaming. DDR es un simulador de baile rítmico, donde el jugador debe seguir los indicadores de la pantalla y realizar los movimientos correspondientes mientras se encuentra sobre el panel de control. Este tipo de juego requiere una gran cantidad de esfuerzos físicos por parte del jugador.

Uno de los ejemplos más conocidos de exergames contemporáneos es la serie de juegos presentados por Nintendo en 2006 para la consola Wii (Wii Sports, Wii Play, Wii Fit y otros). La característica principal de estos juegos es usar un equipo especial incluido con la consola. El control remoto Wii (Wii Sport y Wii Play) es de mano y se usa para diversas actividades de juego, por ejemplo para imitar el boxeo o jugar al tenis. Wii Balance Board (Wii Fit) se asemeja a una báscula corporal; mide el peso y el centro de equilibrio del jugador y proporciona juegos para trabajar en una postura corporal. La mecánica básica de los juegos de Wii hace uso del cuerpo del jugador y permite el control del juego a través de los movimientos físicos. La filosofía general de los juegos de Wii es sacar a los jugadores de sus sofás e involucrarlos en un juego activo.

En 2009, Microsoft presentó una cámara con sensor de movimiento llamada Kinect, para su consola de juegos Xbox 360. A diferencia de Wii, los juegos que trabajan con la Kinect no requieren de controles o mandos especiales para ser jugados. Los juegos son controlados por gestos, movimientos y voz, lo cual brinda más libertad a los jugadores y facilita la interacción natural e intuitiva.

A través de los exergaming, los jugadores realizan actividad física al mismo tiempo que gozan del entretenimiento. Sin embargo, investigadores ven varias limitaciones a esta clase de juegos, ya que requieren una buena cantidad de espacio libre para que los jugadores no sufran lesiones accidentales mientras juegan. Todos los videojuegos activos están vinculados a una consola y televisión, y todas las actividades de exergaming están limitadas al uso en interiores, la sala de estar en particular. La sala de estar generalmente es un espacio inactivo y estático con grandes muebles pesados que ocupa la mayor parte de la sala [17].

Por este motivo, el uso de dispositivos móviles modernos junto con la tecnología de la RA, puede ser una herramienta para incentivar la actividad física al aire libre y de esta forma eliminar las limitaciones de los juegos basados en ubicaciones fijas.

El avance tecnológico que han experimentado los dispositivos móviles permite desarrollar juegos con experiencias increíbles para los jugadores que pueden disfrutar simplemente desde la palma de sus manos. Los dispositivos móviles no sólo incluyen CPU más rápidas, pantallas grandes y cámaras con mejores resoluciones, sino que también contienen una amplia gama de sensores que nos brindan la posibilidad de combinarlos con la tecnología de la RA generando entornos más interactivos y atractivos. De esta forma, lo que se busca es dar movilidad al exergaming, es decir crear juegos que permitan ser jugados al aire libre incentivando la actividad física.

Un grupo de investigadores [18] diseñaron un juego RA llamado GeoBoids. El jugador está rodeado de criaturas virtuales que debe encontrar y capturar. Para ello tiene que caminar a ubicaciones del mundo real, guiados por un mapa que le permite ver dónde están los GeoBoids más cercanos.

SpecTrek es un juego de RA donde el jugador debe localizar y atrapar fantasmas virtuales en el mundo real dentro de un tiempo limitado. El juego posee un mapa que muestra el radio de caza, cuando el jugador se acerca a la posición de un fantasma, se activa la cámara mostrando el fantasma que debe atrapar. Los jugadores tienen que localizar los fantasmas en un área bastante grande por lo que a veces deben apresurarse para completar el juego a tiempo.

La empresa Niantic en Julio de 2016 lanzó Pokemon GO para Smartphones compatibles con iOS y Android. Con más de 100 millones de descargas este juego de RA ha logrado alentar a grandes secciones de la comunidad global a participar en un juego que respalda los cambios en el estilo de vida. El juego consiste en buscar y capturar personajes de la saga Pokémon escondidos en ubicaciones del mundo real y luchar con ellos, lo que implica desplazarse físicamente por las calles de la ciudad para progresar. La aplicación se comporta como un elemento de interacción social, ya que promueve reuniones físicas de los usuarios en distintas ubicaciones de sus poblaciones.

Un estudio reciente ha demostrado que Pokémon Go alienta a los usuarios a aumentar sus actividades físicas en 1437 pasos por día en promedio, además también parece estimular nuevos lazos comunitarios que entregan una mayor socialización, especialmente entre ciertos grupos vulnerables como adolescentes y adultos jóvenes

que a menudo han sido retratados como socialmente aislados e inactivos físicamente, en parte por su interés y compromiso con el entretenimiento tecnológico [19].

3 Juego de Realidad Aumentada

En la actualidad existen una amplia variedad de juegos enfocados en el aprendizaje y en incentivar la actividad física de los jugadores. Sin embargo, son casi inexistentes los juegos que combinan ambas perspectivas para el desarrollo de juegos de RA orientados a niños con discapacidad.

El presente trabajo tiene como objetivo incentivar el hábito de caminata en los niños con diagnósticos dentro espectro autista, trastornos de marcha, diabetes, sobrepeso, obesidad y aquellos en donde sin tener una discapacidad con sintomatología general se recomiende el ejercicio de caminata controlado.

La RA permite la interacción del participante con el mundo real y los objetos que lo rodean, invitándolo a explorarlos e interactuar con ellos mediante el movimiento y el desplazamiento. Esto permitirá a los niños adquirir nuevas experiencias a nivel motor abandonando el sedentarismo que hasta ahora fueron propuestos por otros videojuegos orientados a la salud.

Las aplicaciones de RA permiten ampliar el universo del juego a través de modificaciones virtuales en el mundo real, que sólo ocurren a través del dispositivo con el que se está observando el escenario.

El juego permite diferentes tipos de configuraciones, en donde cada usuario podrá optar de acuerdo a sus necesidades y/o impedimentos, la que más se adecue a sus capacidades, destacando un lector de voz y la posibilidad de modificar los textos a mayúscula para jugadores disminuidos visuales o que no poseen lectoescritura. También se destaca el control parental, que es una característica especialmente útil para padres, permite configurar los temas sobre los cuales se realizaran las diferentes preguntas en el juego (geometría, animales, sumas, restas, colores, frutas, verduras, entre otros) y delimitar el tiempo de uso del juego por parte del participante para que el mismo sea consumido en forma óptima y regulada.

El adulto que está a cargo del niño es el encargado de definir el escenario a jugar, colocando las pistas que considere necesario a través de la utilización de Google Maps integrado en el juego. Además puede restringir el área del mismo a determinada cantidad de metros alrededor de la posición actual, sonando una alarma al salirse del perímetro establecido. Por otra parte, permite tener abierta múltiples sesiones o juegos, de tal forma que si el usuario deja de jugar o suspende el juego, cuando desee y esté en la zona del escenario puede retomarlo.

“Cofre del Tesoro Perdido” (Fig. 1), es un juego de RA compatible con el sistema operativo Android. Este juego se basa en la búsqueda del tesoro, en donde el jugador seguirá pistas que le permitan encontrar los tesoros olvidados por un pirata y también deberá responder preguntas para obtener puntos para avanzar de nivel y alcanzar un mayor rango dentro de la tripulación pirata. Además cerca de las pistas pueden aparecer monedas y algunos miembros de la tripulación que tendrá que capturar.

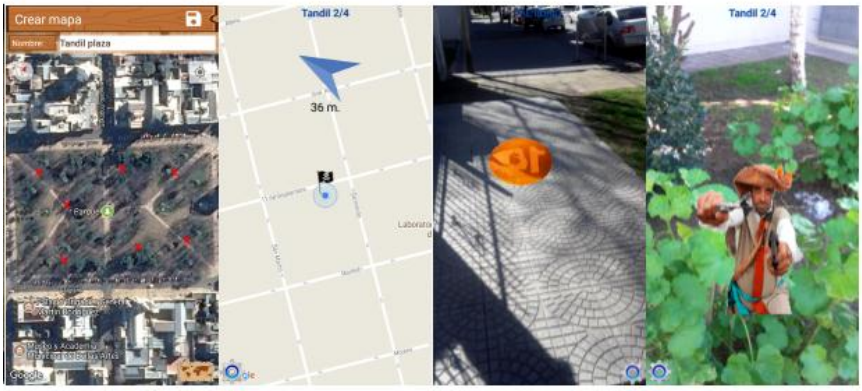


Fig. 1. Capturas de pantalla de “Cofre del Tesoro Perdido”.

Para la codificación del juego se utilizó como entorno de desarrollo Android Studio, lenguaje de programación Java, motor de bases de datos SQLite y la herramienta Android SDK de Google que nos proporciona el acceso a los diferentes componentes de los dispositivos.

El juego se basa en un motor de diseño propio y permite administrar servicios del equipo tales como posicionamiento preciso, cámara, acelerómetro, giróscopo, y compás; siempre orientado a juegos de realidad aumentada. Se aplica un conjunto de filtros y funciones de tratamiento de señales para trabajar con señales estables y que representen la realidad. Brinda un conjunto de servicios que facilitan la construcción del juego, permitiendo mantener objetos geo-localizados, el cálculo de ángulos, posiciones y tamaños en los que se verían dichos objetos de acuerdo a los 6 grados de libertad con los que cuenta el dispositivo móvil. Este motor de juego ha sido desarrollado totalmente a medida para este tipo de aplicaciones.

4 Análisis de Uso

De un total de 78 instalaciones del Google Play Store, hay un total de 38 que han sido completamente dirigidas y rastreadas por diagnóstico, el resto fueron descartadas por considerarse instalaciones recientes, desconocimiento del diagnóstico o por tener menos de 50 escenarios jugados. Con totales de 11 para Diabetes, 4 para Parálisis Cerebral, 12 para Autismo, 6 para Sobrepeso, 2 de Obesidad y 3 niños con otros diagnósticos. Se han analizado un total de más de 4 mil kilómetros contando todos los diagnósticos en un periodo de 9 meses de uso.

El sistema registra el total de metros recorridos por cada juego, el lugar donde fue realizado el escenario y la cantidad de veces que se ha jugado.

Si analizamos el diagnóstico de acuerdo a la cantidad de metros caminados en promedio por cada jugada, tanto diabetes como autismo son los que más caminan ya que superan los cuarenta mil metros, mientras que obesidad y parálisis cerebral registran una menor cantidad de metros (Fig. 2).

La cantidad de caminatas por diagnóstico muestran que tanto diabetes, autismo y sobrepeso son los que más veces han jugado (Fig. 3).

La cantidad de metros recorridos por diagnóstico en el espacio de tiempo analizado, muestra que tanto diabetes como autismo han superado el millón de metros (Fig. 4). Hay que destacar que ambos diagnósticos también son los que más usuarios tienen, y son los que más seguido juegan. En la Fig. 5 se muestra en detalle la cantidad de escenarios jugados por cada jugador en relación a los diferentes diagnósticos, donde horizontalmente se han dispuesto los identificadores de los niños jugadores y verticalmente se muestra la cantidad de escenarios jugados.

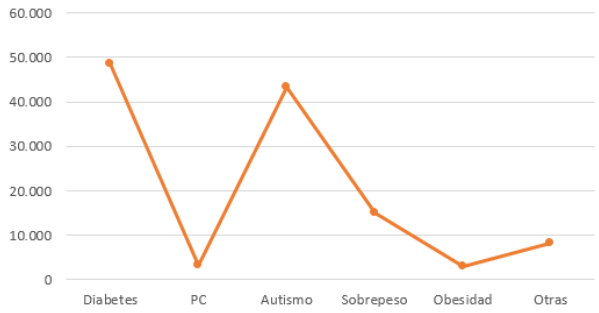


Fig. 2. Metros acumulados por diagnóstico.

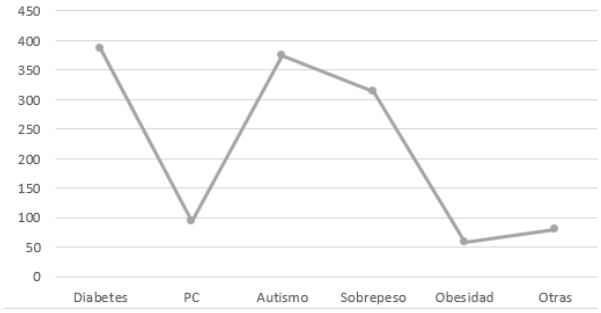


Fig. 3. Caminatas por diagnóstico.

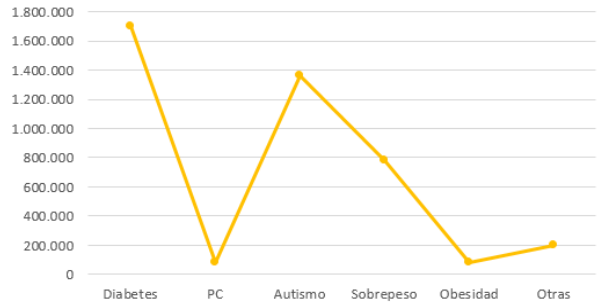


Fig. 4. Metros totales por diagnóstico.

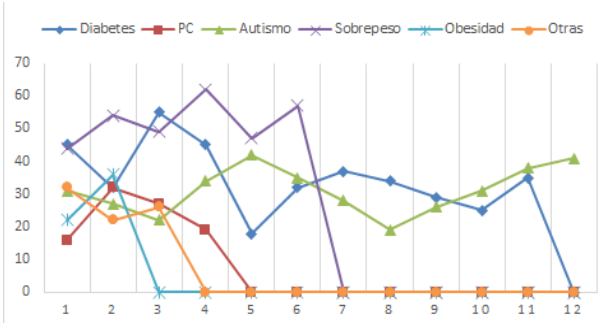


Fig. 5. Escenarios jugados por diagnóstico.

5 Conclusiones y trabajos futuros

En este trabajo se presenta un análisis de un juego de realidad aumentada para incentivar el hábito de caminata en niños con diagnósticos dentro espectro autista, trastornos de marcha, diabetes, sobrepeso, obesidad y aquellos en donde sin tener una discapacidad con sintomatología general se recomiende el ejercicio de caminata controlado.

En base a los resultados obtenidos podemos afirmar que los jugadores con diabetes, autismo y sobrepeso son los que más veces jugaron, recorriendo las mayores distancias. Por otra parte, los niños con parálisis cerebral han jugado menos y con mínimas distancias recorridas, ya que las dificultades motoras hacen que desplazarse les resulte más difícil.

La realidad aumentada es una herramienta innovadora que debemos aprovechar para el desarrollo de juegos que permitan incentivar la actividad física en los niños con diferentes diagnósticos. Con este trabajo pretendemos ser el puntapié inicial para impulsar el desarrollo de juegos que permitan mejorar la calidad de vida de los niños.

Como trabajos futuros se plantea agregar nuevas funcionalidades con la finalidad de añadir un comportamiento social al juego, es decir, que los jugadores puedan compartir escenarios y sus progresos. De esta manera se pretende generar una competencia, buscando aumentar el uso del juego al mismo tiempo de incrementar la actividad física en los jugadores. Además, se analizará las variaciones en el uso del juego a lo largo del tiempo, permitiendo determinar cómo afecta el estado físico al niño jugador.

6 Referencias

1. Organización Mundial de la Salud.: Recomendaciones mundiales sobre actividad física para la salud (2010).
2. Caracuel J. C, Arbinaga F.: Repercusión del ejercicio físico sobre la salud. Apuntes de Psicología. 2012, Vol. 30 (1-3), pp. 547-554.
3. Casajús J. A, Vicente-Rodriguez G.: Ejercicio físico y salud en poblaciones especiales. Exernet, 2011.

4. Speiser PW, Rudolf MC, Anhalt H, Camacho-Hubner C, Chiarelli F, Eliakim A, Freemark M, Gruters A, Hershkovitz E, Iughetti L et al.: Childhood obesity. J Clin Endocrinol Metab 2005, 90:1871-1887.
5. Livingstone MB.: Childhood obesity in Europe: a growing concern. Public Health Nutr 2001, 4:109-116.
6. Hallal, P. C., Andersen, L. B., Bull, F. C., Guthold, R., Haskell, W., Ekelund, U. Wells, J. C.: Global physical activity levels: Surveillance progress, pitfalls, and prospects. The Lancet, 380: 247–257 (2012).
7. Deterding S, Dixon D, Khaled R, Nacke L.: From game design elements to gamefulness: defining "gamification". Proceedings of the 15th International Academic MindTrek Conference, Tampere, Finland (2011).
8. Dias JD, Mekaro MS, Lu JKC, Otsuka JL, Fonseca LMM, Zem-Mascarenhas SH.: Serious game development as a strategy for health promotion and tackling childhood obesity. Rev.Latino-Am Enfermagem, Vol. 24, Ribeirão Preto (2016).
9. Kato, P. M.: Video games in health care: Closing the gap. Review of General Psychology 2010, Vol. 14, No. 2, pp. 113–121.
10. Kato, P. M., Cole, S. W., Bradlyn, A. S., & Pollock, B. H.: A video game improves behavioral outcomes in adolescents and young adults with cancer: A randomized trial. Pediatrics, 122, Number 2, (2008).
11. Buena Salud. "Un juego de video para niños con Diabetes", Octubre 2010. http://www.revistabuenasalud.cl/un-juego-de-video-para-ninos-con-diabetes/
12. Adicción Drogas, Prevención de la Adicción a las Drogas. https://www.tavad.com/juego-adiccion-virtual/
13. Virtual Heroes, 2008. http://www.virtualheroes.com/portfolio/Our%20Portfolio/Pamoja-Mtaani
14. Calle-Bustos A-M, Juan M-C, García-García I, Abad F.: An augmented reality game to support therapeutic education for children with diabetes. PLoS ONE 12(9): e0184645. https://doi.org/10.1371/journal.pone.0184645
15. Carvalho V., Brandão J., Cunha P., Vasconcelos J., Soares F.: Tobias in the Zoo – A Serious Game for Children with Autism Spectrum Disorders. International Journal of Advanced Corporate Learning, Vol 8, No 3 (2015).
16. Brauner, P., Calero Valdez, A., Schroeder, U., & Ziefle, M.: Increase Physical Fitness and Create Health Awareness through Exergames and Gamification: The Role of Individual Factors, Motivation and Acceptance. Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7946 LNCS, pp. 349–362 (2013). http://doi.org/10.1007/978-3-642-39062-3_22
17. Bogost, I.: Persuasive Games: The Expressive Power of Videogames. Literary Linguistic Computing, 23, pp.502–504 (2007). doi:10.1093/llc/fqn029
18. Lindeman, R. W., Lee, G., Beattie, L., Gamper, H., Pathinarupothi, R. & Akhilesh, A.: GeoBoids: Mobile AR for Exergaming. Proceedings of the 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction, 100. (2012). http://doi.org/10.1145/2379256.2379284
19 Wong, Ming Chao et al. "Pokémon-Go: Why Augmented Reality Games Offer Insights for Enhancing Public Health Interventions on Obesity-Related Diseases." Studies in health technology and informatics 241 (2017): pp. 128-133.

Los e-Portfolios como Metodología de Evaluación y Seguimiento de Aprendizajes en Ámbitos Educativos

Natalia Correa

LIFIA – Facultad de Informática de la Universidad Nacional de La Plata
La Plata, Pcia. de Buenos Aires, Argentina
nataliac@lifia.info.unlp.edu.ar

Abstract. Los portfolios educativos son una herramienta y, a su vez, una estrategia de evaluación que reúne una serie de evidencias seleccionadas por los alumnos que reflejan su aprendizaje y que permite ser evaluado ya que demuestra comprensión, avances y logros de los alumnos así como dificultades, estancamientos y necesidad de revisión. Se presentan como un aporte valioso para el docente, con posibilidades de ser explotados y explorados en diferentes ámbitos educativos y con diferentes objetivos de desarrollo de competencias según el grupo de alumnos al que esté dirigido.

Este trabajo se centra en los portfolios de alumnos presentando nociones generales de los mismos así como su sentido de sistema de enseñanza aprendizaje. Se presentan además algunas experiencias, de modalidad innovadora, realizadas en diversos ambientes educativos.

Keywords: Portafolios electrónicos, Sistemas de evaluación, Procesos de Enseñanza y Aprendizaje

1 Introducción

Los portfolios -o portafolios- suelen ser utilizados por muchos profesionales, en especial de las artes, para distintos tipos de producciones artísticas: fotografía, escultura, video, pintura, etc.; así como también de ámbitos como son la publicidad, la narrativa, el diseño, la programación web, entre otros.

Cuando alguien diseña su portfolio tiene el claro propósito de seleccionar los mejores trabajos de su producción para ser mostrados a otras personas. En esta selección, el autor reflexiona y valora sus trabajos muchas veces agregando explícitamente, información sobre su ponderación de la propia producción. En definitiva, un portfolio permite demostrar el conocimiento, las habilidades y competencias del autor en un área o contexto determinado.

Barberá define en [1] al portfolio como un "instrumento que tiene como objetivo común la selección de muestras de trabajo o 'evidencias' de consecución de objetivos personales o profesionales que, ordenados y presentados de un determinado modo, cumplen la función de potenciar la reflexión sobre cada una de las prácticas (educativas, profesionales o civiles)".

Por ello mencionamos que el autor de un portfolio 'valora' su propia producción, ya que ésta es una parte del propio instrumento, y que además habilita la valoración de otros ya que el resultado tiene un objetivo: mediante la selección de evidencias o mejores producciones, realizadas en un tiempo determinado y ordenadas, documentadas se puede buscar un trabajo, solicitar una promoción, ser evaluado para ocupar un puesto, etc.

Por la definición brindada, el sentido explicado a continuación y el uso que se le han dado, los portfolios en el ámbito educativo configuraron un nuevo instrumento de evaluación (y una estrategia) y se ha sumado a las varias herramientas de las que los docentes se sirven a la hora de diseñar situaciones educativas contemplando además la evaluación.

Dado que hoy en día y cada vez más, los ámbitos educativos, las prácticas y situaciones educativas, los procesos de enseñanza aprendizaje se encuentran atravesados y conformados por las TIC, hablamos de e-portfolios o portfolios electrónicos haciendo referencia al mismo concepto presentado pero en un espacio puramente digital.

El presente trabajo se centrará en portfolios electrónicos de alumnos cuyo objetivo primario es el desarrollo de competencias que éste necesitará para su estudio, su práctica profesional, su vida y que se enmarcan en diversos ambientes educativos. La organización del mismo es la siguiente: en la sección 2, se menciona al portfolio dentro del proceso de enseñanza aprendizaje. La sección 3 presenta algunas experiencias en diferentes ámbitos educativos –formales/ no formales, públicos/ privados- para finalizar con la sección 4 de conclusiones finales del artículo.

2 Los e-Portfolios en los Procesos de Enseñanza Aprendizaje

Puede decirse que el uso del portfolio electrónico responde, en definitiva, a un cambio –o ampliación- de la concepción de la evaluación y a un cambio más amplio dentro del paradigma en la educación, hoy atravesado y conformado por las TIC.

Según Barberá en [2], [3] la evaluación es un mecanismo central en la buena marcha de los procesos de enseñanza y aprendizaje; bien entendida debe ser formativa y formadora, no circunscripta a un momento en particular sino que involucra a todo el proceso de enseñanza y aprendizaje; da espacio a la retroalimentación continua entre docentes y alumnos y alumnos con sus pares, a la reflexión sobre el propio aprendizaje.

En este marco, el e-portfolio permite al alumno demostrar que está aprendiendo, cómo lo está aprendiendo, posicionándolo activamente en ese/ su proceso con una actitud crítica y dialógica respecto de sus conocimientos. Por otra parte, posibilita al docente un seguimiento de ese proceso de aprendizaje, le informa sobre el nivel de competencia, le permite intervenciones en tiempo y forma –si fuesen necesarias- además de habilitar implícitamente un feedback de su propia actuación docente.

De lo explicitado anteriormente se desprende en forma general, que este tipo de evaluación se aleja por su concepción de la evaluación más tradicional y estandarizada así como que requiere, por parte del docente, implicarse más en el proceso de aprendizaje del alumno, estar atento a sus avances individualmente, a cómo aprende o qué necesita como soporte del y para el aprendizaje.

Al pensar entonces en la implementación de un e-portfolio educativo, se deberían tener en cuenta las siguientes características del mismo como instrumento de evaluación que retomamos de [4]:

- como Reflexión, que se sustente en la descripción de las producciones con documentos significativos.
- como Autoevaluación, que permita analizar objetivamente las habilidades y competencias alcanzadas al momento, en forma individual o grupal, de forma dinámica y en diferentes momentos del aprendizaje
- como Evaluación auténtica, que permita advertir al estudiante sus problemas y progresos respecto de los objetivos de aprendizaje propuestos
- como Aprendizaje del aprendizaje, que permita a los docentes comprender mejor cómo aprenden los estudiantes
- como Atención a los problemas específicos, que al detectar errores, en las producciones, se deban llevar a cabo estrategias específicas para su solución

Cabe destacar que todos los aspectos mencionados anteriormente reflejan la visión del portfolio como un sistema de evaluación que se encuentra plenamente integrado en el proceso de enseñanza aprendizaje.

3 Experiencias con Portafolios Electrónicos en Ámbitos Educativos

En esta sección, se mencionan tres experiencias realizadas en el uso de e-porfolios en diversos ambientes educativos, con diferentes grupos de alumnos.

En cada caso se menciona el e-portfolio definido y las características de las experiencias.

3.1 Experiencia en un Ámbito Educativo no Formal

La siguiente experiencia es reciente, de los últimos dos años, y sucedió en el contexto de una ONG (Organización No Gubernamental), con adolescentes participantes del Taller de Programación Web "ProgrAmando".

Para la definición de este e-portfolio, se consensuó entre docentes y alumnos, una rúbrica que contenía aspectos considerados elementales desde la perspectiva técnica/ tecnológica, visual y comunicacional y de la usabilidad.

Además de los conocimientos tecnológicos para la creación de páginas web, los docentes estaban interesados en incentivar habilidades y competencias relacionadas a la expresión y comunicación, escritura, justificación oral y escrita, apropiación de vocabulario propio de la especialidad. Para ese planteamiento de la forma de evaluar

se priorizaron las habilidades que se buscaban potenciar y no tanto los contenidos 'técnicos'.

Los e-portfolios primeramente se configuraron como internos y personales: cada participante del taller tenía su propia página web con todas sus producciones; y cada una de estas últimas era una página web en la que se reflejaban los conocimientos adquiridos al momento. La página principal presentaba al autor, indexaba todas las producciones y mostraba sus avances. A medida que iba avanzando el año, sólo las producciones que cada joven consideraba como las 'mejores' quedaban en su página que era su e-portfolio personal.

Al finalizar el ciclo lectivo, cada uno generó una página web como trabajo final integrador que fue evaluado por docentes y pares. Las producciones finales son parte del portfolio del Taller de Programación de la ONG [5] y se presentaron a toda la comunidad educativa de la ONG con una actividad interactiva con celulares para descubrir y ver cada una de las páginas web finales –e-portfolios- de los alumnos.

3.2 Experiencia en Ámbito Educativo de Formación a Empresas/ Instituciones

La siguiente experiencia también es reciente, del año en curso. Se dio con la inclusión de un e-portfolio en el dictado de un curso en línea para personal de una institución que buscaba capacitarse en un tema específico. Todos los participantes eran adultos, con formación terciaria o universitaria y contaban con conocimientos tecnológicos.

La definición del e-portfolio se realizó de forma previa, junto con la definición del programa y currículum: selección de contenidos, preparación de actividades, estructuración del curso en línea.

En este caso y por darse dentro de un EVEA (Entorno Virtual de Enseñanza Aprendizaje) como Moodle [6] se utilizó una extensión de portfolio provista por Moodle que permite configurar sistemas remotos para que los alumnos exporten contenido en ellos. De los portfolios habilitados se usó: G Drive [7].

El e-portfolio definido se presentó como "Registro Individual del Aprendizaje", individual y privado, en el que se solicitaba un análisis textual posterior a cada unidad temática. Se presentó con la consigna de realizar el análisis, escribiendo una síntesis de lo 'central' de la unidad y formular dos preguntas sobre lo aprendido (con una breve introducción) que los participantes creyesen estar en condiciones de responder por ellos mismos.

Para ayudar a completar esas dos preguntas que quedarían alojadas en el documento compartido con el docente y gracias a la extensión de Moodle, los participantes podían exportar sus propias intervenciones en foros si lo consideraban adecuado/ relevante para responder a la consigna. Con lo cual, no sólo se definió un e-portfolio particular, sino que en este caso además se pudo relacionar e-actividades [8] con e-portfolios.

Además de los conocimientos propios del curso en cuestión, se tenía el interés en incentivar o propiciar el desarrollo de habilidades como relacionar y comparar, formular preguntas, organizar y clasificar.

La experiencia fue exitosa tanto para docente como alumnos.

3.3 Experiencias en Ámbito Educativo de Nivel Superior

La última experiencia a compartir, se da en el ámbito universitario, con alumnos de la cátedra "Desarrollo de Software Basado en Modelos", materia optativa de 4to y 5to. año del área de Ingeniería de Software y Bases de Datos de las carreras Licenciatura en Sistemas y Licenciatura en Informática.

En este caso, el trabajo práctico definido en la materia consta, entre otros ítems, de una sección de informe sobre la herramienta CASE utilizada para generar los modelos solicitados durante la cursada.

Ese informe fue puntapié para definir el e-portfolio en este caso. En particular, se trabajó en grupos o individualmente, compartiendo el portfolio con la docente. El mismo constó de las siguientes secciones, como se menciona en [10]:

- Un apartado introductorio: propósitos del portafolio, el autor o autores
- Una guía de contenidos (una especie de índice), para que el lector pueda orientarse dentro de las secciones del documentos;
- La/s sección/es de desarrollo del portfolio, donde estuviese aquello relevante para el alumno, pudiendo ser en este caso, una producción propia (comentario desde el uso de la herramienta, capturas de pantalla) o ajena (video, tutorial, pregunta en un foro de ayuda externo, etc.) justificando en este caso su incorporación
- Una sección final, de conclusiones finales y valoración del proceso;

y se desarrolló durante todo el cuatrimestre.

Entre las habilidades que se buscaron potenciar se destacan, entre otras: representar mediante el uso de modelos e instrumentos, argumentar: justificar acciones y decisiones, evaluar/ atribuir valores.

En este caso, fue difícil para los alumnos llevar un ritmo acorde: por un lado la tarea del modelado de un sistema y por otro, la incorporación de evidencias del uso de esas herramientas de modelado. En su gran mayoría, se dio una participación regular al inicio y alta al final del período, registrando y documentando más bien 'de memoria', dificultando la posibilidad de valorar con amplitud el proceso de aprendizaje completo.

4 Conclusiones

La evaluación debería contribuir a generar procesos reflexivos, toma consciente de decisiones por parte de los estudiantes, ser formativa y darse a lo largo de todo el proceso de aprendizaje.

Clara y vivencialmente constatamos que se marca como tendencia el cambio en el rol de los alumnos: de consumidores a prosumidores, cambio que se debe, entre otros factores humanos, a la gran disponibilidad y penetración de dispositivos y herramientas TIC en usos cotidianos; lo cual conlleva la necesidad de transformación

de las prácticas pedagógicas, el fomento de la creatividad y la adquisición de habilidades productivas.

Un portfolio, en su estructura, puede ser muy variado dependiendo de su propósito, del público al que se dirige, del soporte donde es implementado y, también y no menos importante, de la creatividad de cada autor.

Los beneficios de adoptar la práctica de planear la evaluación junto con la planificación del currículum, se mencionan en [10] y de ese trabajo destacamos: el desarrollo de habilidades tecnológicas, la promoción del aprendizaje, la contribución a la reflexión de lo producido, la propia revisión del proceso. En este sentido, resulta muy valioso que además de los contenidos, la propuesta evaluativa aborde una mirada del proceso de aprendizaje de cada uno.

Para finalizar y concluir, en este trabajo se han presentado las características y beneficios del uso de portafolios con fines educativos y su potencial como herramienta y estrategia de aprendizaje.

Se han presentado casos de ejemplo como ricas experiencias donde se comprende que su implementación requiere una planificación educativa adecuada que atraviesa todo el proceso, que tiene pocas posibilidades de reutilización y que ofrece, mediante un proceso constante de diálogo con el alumno, una invitación a la reflexión del proceso de aprendizaje vivido.

Bibliografía

1. Barberá, E. (2005). “La evaluación de competencias complejas”. Revista Educere, 31, 497-504.
2. Barberá, E., Bautista, G., Espasa, A. y Guasch, T. (2006). “Portafolio electrónico: desarrollo de competencias profesionales en la Red”. En Antoni Badia (coord.). Enseñanza y aprendizaje con TIC en la educación superior. Revista de Universidad y Sociedad del Conocimiento (RUSC). Vol. 3, n.° 2. UOC.
3. Barberá, E., Gewerc, A. y Rodríguez, J.L. (2009). “Portafolios electrónicos y educación superior en España”. RED: Revista de Educación a Distancia. Número monográfico VII.-30 de abril de 2009. Número especial dedicado a Portafolios Electrónicos y educación superior. Disponible en http://www.um.es/ead/red/M8. Última visita: 11/07/2018
4. Cataldi; Z. y Lage F. (2010) E-Portafolio. Una opción metodológica más auténtica para evaluación de aprendizajes autónomos en educación superior. TE&ET 2010. El Calafate.
5. ONG Centro Educativo Integral "Camino a la Casita". Taller de Programación ProgrAmando (2018). http://caminoalacasita.org/programando2017. Última visita: 11/07/2018
6. Moodle. Herramienta de gestión de aprendizaje. https://moodle.org. Última visita: 11/07/2018
7. Google Drive. Servicio de alojamiento de archivos. https://www.google.com/drive/
8. Sanz, C., Zangara, A. (2013). "Las e-actividades como elemento central en el diseño de propuestas de educación mediada". In I Jornadas Nacionales de TIC e Innovación en el Aula. Recuperado de: http://sedici.unlp.edu.ar/handle/10915/26547. Última visita: 11/07/2018
9. Barberá, E. "Estado y tendencias de la evaluación en educación superior" Revista electrónica de la Red Estatal de Docencia Universitaria. Volumen 3. Nº 2.
   Disponible en: http://revistas.um.es/redu/article/view/10161. Última visita: 11/07/2018
10. Barberá, E, De Martín, E. (2009). Portafolio electrónico: aprender a evaluar el aprendizaje. Barcelona: Editorial UOC.

Propuesta pedagógica para la comprensión del espacio tridimensional utilizando Realidad Aumentada

Álvaro Saldivia Obando[2], Tatiana Gibelli[1,2], Cecilia Sanz[3,4],

[1]Centro Interdisciplinario de Derechos Inclusión y Sociedad
[2]Sede Atlántica, Universidad Nacional de Río Negro
[3] III-LIDI Facultad de Informática Universidad Nacional de La Plata
[4]Investigador Asociado de la Comisión de Investigaciones Científicas de la Pcia. de Bs. As.
{asobando,tgibelli}@unrn.edu.ar, csanz@lidi.unlp.edu.ar

Resumen. El aprendizaje de conceptos de geometría tridimensional como puntos, vectores, planos y rectas en el espacio suele presentar dificultades por parte de los estudiantes universitarios de primer año. Dichas dificultades pueden estar asociadas a que se concentran en cuestiones técnico-algebraicas descuidando la visualización geométrica que permitiría una mejor comprensión de dichos conceptos. Ante esta situación, se diseña una propuesta pedagógica consistente en actividades que incluyen el uso de realidad aumentada, considerando que esta tecnología facilita el desarrollo del pensamiento espacial. En este trabajo se presenta la fundamentación teórica, y detalles de la propuesta, junto con comentarios sobre resultados de una implementación piloto de la misma.

Palabras clave: Realidad aumentada, geometría tridimensional, propuesta pedagógica.

1 Introducción

El aprendizaje de conceptos de geometría tridimensional, tales como puntos, vectores, planos y rectas en el espacio, suele presentar dificultades por parte de los estudiantes universitarios. En las materias en las que se desempeñan como docentes algunos de los autores de este trabajo se observa esta situación, en particular en las materias Matemática I, correspondiente al primer año de la carrera de Licenciatura en Sistemas e Ingeniería Agronómica y en Álgebra correspondiente al tercer año de la carrera de Contador Público, todas ellas de la Sede Atlántica de la Universidad Nacional de Río Negro (UNRN). Se considera que esta problemática suele estar asociada a las dificultades para la visualización del espacio tridimensional que permitiría una interpretación geométrica de dichos conceptos, y facilitaría la comprensión y el trabajo técnico-algebraico asociado [1]. De acuerdo a la teoría de Representaciones Semióticas, formulada por Raymond Duval, los objetos matemáticos no son accesibles físicamente, a través de evidencias sensoriales directas o mediante el uso de instrumentos, la única forma de acceder y trabajar con ellos es a través de signos y representaciones semióticas

[2]. Duval enfatiza en la importancia de que el alumno pueda representar y reconocer a un mismo objeto matemático en diferentes registros de representación semiótica. Para el aprendizaje de conceptos de geometría tridimensional, la integración de estos dos registros semióticos (algebraico y geométrico) permite una comprensión más acabada de los conceptos. Para la interpretación geométrica los estudiantes necesitan imaginar objetos tridimensionales en diferentes orientaciones, trasladar mentalmente la representación en el plano a tres dimensiones. Esta habilidad de pensamiento espacial es esencial para las carreras de Ingeniería y otras del ámbito artístico y científico, pues se utiliza para representar, manipular información en el aprendizaje y en la resolución de problemas [3].

Esta problemática fue abordada desde la investigación, en el marco de un proyecto de investigación denominado "La mediación de las tecnologías de la información y la comunicación en procesos educativos. Innovaciones para mejorar los procesos de enseñanza y aprendizaje" (PI-UNRN-40C486), financiado por la Universidad Nacional de Río Negro. El objetivo principal del mismo, consiste en investigar metodologías y estrategias innovadoras que favorezcan procesos educativos mediados por TIC, en particular por aquellas consideradas emergentes como la Realidad Aumentada (RA). En este contexto se consideró apropiado el diseño de propuestas pedagógicas que incorporen el uso de la RA para facilitar el desarrollo de las habilidades de pensamiento espacial.

En este trabajo se presenta una propuesta diseñada como cierre de la unidad de vectores y sus aplicaciones geométricas. Además se describe una experiencia piloto realizada a partir de dicha propuesta en la materia Álgebra correspondiente al tercer año de la carrera de Contador Público, durante el primer cuatrimestre de 2018.

2 Marco teórico

Una de las tecnologías emergentes con posibilidad de inclusión en Educación es la Realidad Aumentada (RA). Según Azuma et al. [4] la RA se caracteriza por: a) una combinación de objetos virtuales y reales en un escenario real, b) usuarios interactuando en tiempo real y c) una alineación entre los objetos reales y virtuales.

Respecto a su potencialidad en el plano educativo, la RA permite incorporar multimedia a los procesos de enseñar y aprender, innovar en la práctica docente y promover el diseño de materiales educativos atendiendo a los requerimientos didácticos [5]. La RA puede proporcionar grandes oportunidades en distintas áreas como la ciencia o la ingeniería, puesto que estas disciplinas conllevan en su currículum un enfoque práctico en el aula [6]. Sin embargo, como señala Prendes Espinosa [7], aun siendo la tecnología lo suficientemente madura para ser útil en los entornos educativos, los

nuevos desarrollos de RA llevarán a una nueva generación de aplicaciones que desplegarán mucho más su potencial.

Específicamente para la enseñanza de la matemática, Bujak et al. [8], analizan el uso de la RA desde tres perspectivas: física, cognitiva y contextual, presentando para cada una de ellas ejemplos de aplicaciones existentes de RA, así como pautas para el diseño de futuras experiencias considerando cuestiones prácticas y tecnológicas. En la dimensión física destacan que las interacciones naturales con objetos físicos fomentan la comprensión de los conceptos en contexto. En la dimensión cognitiva mencionan que la alineación espacio-temporal de la información puede servir como andamiaje en la comprensión de los conceptos abstractos. Finalmente, en la dimensión contextual plantean las posibilidades que brinda la RA para el aprendizaje colaborativo en entornos no tradicionales, generando experiencias de aprendizaje significativas.

En cuanto al uso de RA para la comprensión del espacio tridimensional existen algunas experiencias. Se destaca la que desarrollan De la Torre et al. [3] donde proponen el uso de realidad aumentada y tabletas digitales multitáctiles para ofrecer la posibilidad de manipular un modelo digital en tres dimensiones, de forma similar a como se hace con un modelo físico. Estos autores presentan resultados de una prueba piloto con 62 estudiantes de tres ámbitos educativos, y realizan una valoración global y una específica sobre las tecnologías utilizadas mostrando evidencias de que la RA con tabletas multitáctiles es una alternativa válida para la sustitución de los modelos físicos en entornos digitales.

3 Descripción de la propuesta

En base a los antecedentes indagados en la sección previa, se presenta a continuación una breve fundamentación de la propuesta con RA para el aprendizaje en la materia Álgebra, junto con descripción de las actividades diseñadas y los recursos tecnológicos utilizados para su implementación.

3.1 Fundamentación

El contenido conceptual a abordar con esta propuesta son Aplicaciones de la Geometría en Espacio Tridimensional. Se considera como hipótesis que las actividades con RA facilitan el aprendizaje de estos contenidos, de acuerdo a lo mencionado en el marco teórico.

Para la actividad diseñada el uso de la RA permitió superponer a objetos físicos ciertos objetos virtuales (videos con información adicional y explicaciones). Esto facilitó la asociación de manera natural de los conceptos matemáticos (vectores, puntos, planos y rectas en el espacio tridimensional) con objetos físicos concretos (paredes,

rampas, barandas, etc.). El escenario real aporta la posibilidad de distinguir, con el soporte de objetos concretos y cotidianos, los elementos constitutivos (vector normal, vector director) de los conceptos matemáticos (plano, recta), y esto puede favorecer así, la comprensión de la posición relativa de ellos en el espacio tridimensional.

La propuesta tiene como objetivo facilitar la comprensión de los conceptos trabajados previamente y su integración desde el punto de vista teórico-práctico. Durante el desarrollo de estas actividades, se espera que el alumno ponga en juego sus conocimientos, aplicándolos y confrontando sus producciones con los objetos reales del entorno, pudiendo detectar así, confusiones conceptuales y/o reforzar la comprensión del tema. Las actividades propuestas, siguiendo la clasificación de Marzano [9], promueven habilidades de Nivel 1 (uso del conocimiento) tales como integrar, inferir, aplicar, interpretar y de Nivel 2 (metacognición) como valorar y explicar, etc; como así también se promueven procesos cognitivos de síntesis y evaluación.

3.2 Actividad propuesta

La propuesta se basa en establecer asociaciones entre objetos cotidianos con conceptos de geometría tridimensional. Para ello se seleccionaron tres espacios físicos que los alumnos frecuentan, y a partir de ellos se planificaron las actividades con RA, en donde se asoció a dichos espacios información virtual. A continuación se describen los distintos espacios físicos y la actividad prevista en cada uno de ellos.



Fig. 1. Objetos virtuales (vectores) asociados mediante RA a los objetos físicos sobre el puente ferrocarretero de la comarca Viedma-Carmen de Patagones

El primer objeto utilizado, considerado como soporte físico, fue el puente ferrocarretero de la Comarca Viedma - Carmen de Patagones, cercano al campus universitario. Allí se asocian las barandas de la pasarela del mismo con rectas en el

espacio (figura 1). La información virtual se complementa con una explicación sobre la forma de encontrar la ecuación en base a un punto conocido y su vector asociado o director, referenciando dichos elementos sobre la estructura del puente.

El segundo elemento utilizado es la fachada del campus (figura 2), sobre el cual se adiciona información virtual sobre vectores normales, puntos sobre la pared, vectores asociados y un plano asociado a la pared del frente del edificio. Se complementa con una explicación (en video) sobre los lineamientos generales para encontrar las ecuaciones de planos y rectas y también el concepto de producto vectorial en base a los elementos identificados.



Fig. 2. Objetos virtuales asociados a los objetos físicos sobre fachada del campus universitario

El tercer caso (figura 3) se realiza tomando tres paredes del aula de clases. La información virtual que se adiciona a esta escena se realiza mediante dos videos:

- El primero explica cómo encontrar la ecuación de una recta paralela a dos planos que se intersectan y que forman un ángulo de noventa grados.
- El segundo video desarrolla cómo encontrar la ecuación de un plano a partir de tres puntos del mismo, para ello se sitúa el video en una intersección de dos paredes del aula.

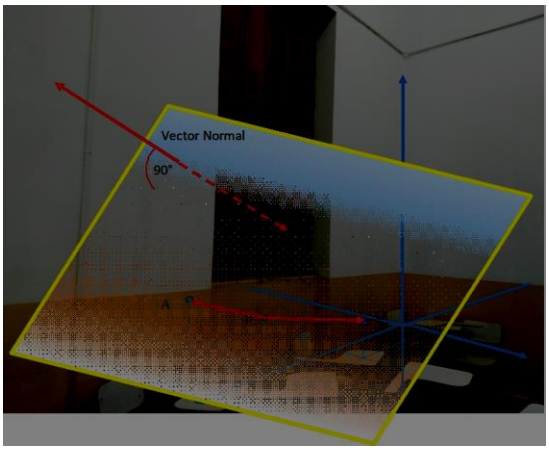


Fig. 3. Objetos virtuales asociados a los objetos físicos en el espacio áulico

3.3 Diseño tecnológico

Para el desarrollo de la actividad con RA se hizo uso del aplicativo HPReveal (nueva versión de Aurasma) que brinda herramientas para usar Realidad Aumentada y convertir objetos cotidianos, imágenes y lugares en objetos para interactuar a través de gráficos, animación, video, audio y contenido 3D. Esta aplicación puede ser utilizada de diferentes maneras:

- Para crear contenido de RA: mediante la aplicación web que permite crear, administrar y realizar un seguimiento de las experiencias.
- Para visualizar contenidos de RA creados por otros: mediante el aplicativo para móviles con sistema IOS o Android a través del cual se pueden acceder e interactuar con los desarrollos generados con esta aplicación.

El aplicativo para creación fue utilizado para generar "*Auras*" a través de los cuales se agregó la información virtual a los objetos físicos considerados. Ellos se realizaron a partir de una foto tomada en el lugar utilizada como *trigger* (imagen que reconocerá el dispositivo para aumentar la escena) y asociando sobre ella información virtual como *overlay* (es donde se agrega el objeto virtual que se va a incorporar a la escena). La información virtual adicionada sobre la imagen del objeto físico se realizó a través de videos explicativos donde se asoció información virtual a dicha imagen. Así, por ejemplo, para el Aura asociado al puente ferrocarretero (Figura 1), en el video se presenta una explicación sobre cómo encontrar la ecuación de una recta en el espacio, considerando la información virtual que se agrega sobre la imagen del espacio físico (pasarela del puente). De manera análoga se generaron otros Auras, agregando sobre el aula y el frente del campus, videos que describen elementos y estrategias que permiten

hallar ecuaciones de planos y rectas. A partir de la visualización de estos videos los alumnos debieron realizar un trabajo práctico basado en diversas configuraciones entre planos y rectas.

En el caso de los alumnos se propuso el uso de la aplicación para la visualización que les permite acceder a la información virtual (videos) adicionada a partir de las imágenes. La visualización se alcanza a partir del uso de la cámara de dispositivos móviles enfocando el objeto real (aula, puente o frente campus) o una imagen de estos espacios reales.

4 Experiencia y resultados

Se realizó una primera implementación piloto de esta propuesta durante el primer cuatrimestre de 2018 con los alumnos que se encuentran cursando la materia Álgebra, que se dicta para la carrera de Contador Público. La cantidad de inscriptos en la materia es de **45 alumnos**, cuyas características generales son: edad promedio de 23 años, un 57% desempeña actividades laborales entre 4 a 6 hs. por día, el 100% tiene acceso a internet, utilizan principalmente dispositivos móviles, sólo un 28% hace referencia a uso de PC de escritorio.

Esta primera experiencia tuvo por objetivo determinar fortalezas y debilidades en cuanto al uso de la aplicación y la propuesta en sí misma. Para ello, se realizó con todos los alumnos del curso durante el espacio de práctica una actividad para reforzar los conceptos de la unidad, previa al examen parcial. Se hizo referencia a que en el marco de una investigación se diseñó una actividad de RA a partir de la aplicación HPReveal y que el desarrollo de la actividad tendría un carácter exploratorio de los resultados obtenidos en la misma para mejorar la propuesta y eventualmente, convertirla en actividad obligatoria en años posteriores.

La clase anterior a la experiencia se indicó a los estudiantes que deberían contar con un dispositivo móvil con cámara y acceso a internet, en el cual previamente, tendrían que instalar la aplicación y crear un usuario. Se les solicitó además descargar las imágenes que se utilizarían como *trigger* o traerlas impresas (las mismas se pusieron a disposición con anterioridad a través del aula virtual de la materia junto con el trabajo práctico). Se propuso desarrollar la tarea en forma grupal de manera de contar con al menos un dispositivo por cada grupo (3 o 4 estudiantes). La segunda parte de la actividad fue realizar los problemas propuestos en el trabajo práctico y cuando surgieron dudas en su resolución se invitó a los alumnos a mirar nuevamente los videos que estaban relacionados a los ejercicios.

Durante la experiencia, se pudo observar una gran participación de los estudiantes. Los distintos grupos no tuvieron inconvenientes en cuanto al uso de la aplicación, tanto

en acciones previas (descarga y creación de usuario) así como durante la actividad. Después de unas breves explicaciones sobre el uso de la aplicación todos los grupos pudieron ver los videos explicativos, y posteriormente desarrollar los problemas. Asimismo se pudo detectar a partir de comentarios de los propios alumnos, que estos videos los ayudaron a resolver y disipar dudas en el trabajo práctico tradicional que posee la materia.

Además de observación directa, finalizada la actividad, se propuso en forma voluntaria responder a un cuestionario *online* de valoración de dicha actividad. El cuestionario fue diseñado por el equipo de investigación para conocer las percepciones de los estudiantes que usan materiales educativos digitales (MED) respecto a su utilidad, motivación y cómo estas influyen en el proceso de enseñanza y aprendizaje, luego de una revisión de los instrumentos disponibles [10]. El mismo está compuesto por ítems con afirmaciones respecto al MED en los cuáles el alumno debe indicar su grado de acuerdo en escala Likert de 5 puntos: 1-Muy en desacuerdo, 2-En desacuerdo, 3-Indeciso, 4-De acuerdo, 5-Muy de acuerdo. En esta experiencia en particular no se usa un material educativo digital, sino una aplicación basada en RA, para no entrar en tecnicismo los alumnos la denominaron MED. En base a las respuestas obtenidas (en tabla 1 se muestran medidas de resumen) se observó que los alumnos consideraron la actividad como una experiencia que aportó a su aprendizaje. Esto se evidencia en sus respuestas, donde podemos destacar:

Un acuerdo (valoración promedio próxima a 4) con afirmaciones de percepción positiva de la actividad para el aprendizaje tales como: “Considero que los recursos digitales pueden ser útiles para el aprendizaje”, “El uso de los MED puede mejorar mi comprensión del tema”, “Volvería a utilizar este MED si tuviera oportunidad”, entre otras.

Un desacuerdo (valoración promedio próxima a 2) con afirmaciones de percepción negativa del material, tales como: “No me atrae utilizar este tipo de materiales”, “Creo que haber trabajado con este MED no aportó demasiado a mi aprendizaje”, “La cantidad y repetición de las actividades me resultó aburrida”, “El MED me resultó tan abstracto que fue difícil mantener mi atención en él”.

Tabla 1. Valores medios y desvío estándar en valoración de afirmaciones del cuestionario (Escala:1-Muy en desacuerdo, 2-En desacuerdo, 3-Indeciso, 4-De acuerdo, 5-Muy de acuerdo)

Afirmación	Valor medio	Desvío estándar
Considero que los MEDs pueden ser útiles para el aprendizaje	4.57	0.53
El uso del MED puede mejorar mi comprensión del tema	4.43	0.79
Volvería a utilizar este MED si tuviera oportunidad	4.29	0.49
Realizar las actividades del MED con éxito fue importante para mí	4.14	0.38
Aprender a utilizar otros MED similares me resultaría fácil	4.14	0.69
Me gustaría utilizar MED para aprender otros temas	4	0.58

Luego de esta experiencia considero que los MED son fáciles de usar	4	0.58
Logré relacionar el contenido del MED con cosas que conozco	3.86	0.69
Me resultó fácil la interacción con este MED	3.71	0.49
El diseño del MED logró captar mi atención	3.71	0.49
El diseño del MED me resultó amigable	3.71	0.49
No me atrae utilizar este tipo de materiales	2.29	1.11
Creo que trabajar con este MED no aportó demasiado a mi aprendizaje	2.29	1.25
El uso de este tipo de MED para aprender/estudiar, me genera estrés	2	0
La cantidad y repetición de las actividades me resultó aburrida	2	0.82
El MED me resultó tan abstracto que fue difícil mantener mi atención en él	2	0.58

5 Conclusiones preliminares y trabajo futuro

A pesar de lo breve de la actividad resultó motivadora para los estudiantes, generando otra actitud para el trabajo con los contenidos. Teniendo en cuenta su valoración de la actividad a través del cuestionario, se observó que fue considerada como una experiencia positiva que aportó a su aprendizaje, mostrando predisposición para usar materiales de este tipo.

Se considera entonces que con actividades de este tipo se puede favorecer la comprensión del espacio tridimensional al permitir interactuar de otra manera con los conceptos al asociarlos a objetos cotidianos, facilitando así la interpretación geométrica de los mismos y complementando el trabajo algebraico que se realiza usualmente.

Como trabajo futuro se propone abordar a través de propuestas didácticas mediadas por tecnologías digital, en especial RA como en este caso, para otros contenidos y realizar nuevas implementaciones con un seguimiento más profundo de la experiencia que permita obtener conclusiones sobre el impacto de estas actividades en el aprendizaje de los alumnos.

Referencias

1. Del Río, L. S., Búcari, N. D. y Sanz, C. V. Uso de recursos hipermediales para la enseñanza y el aprendizaje de la matemática. In II Congreso Internacional de Enseñanza de las Ciencias y la Matemática. Tandil. (2016)
2. Duval, R. Un tema crucial en la educación matemática: La habilidad para cambiar el registro de representación semiótica. Gaceta de la RSME, 9(1), 143-168. (2006)
3. De la Torre Cantero, J., Martin-Dorta, N., Pérez, J. L. S., Carrera, C. C., & González, M. C. Entorno de aprendizaje ubicuo con realidad aumentada y tabletas para estimular la comprensión del espacio tridimensional. Revista de Educación a Distancia, (37). (2015)
4. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. Recent advances in augmented reality. Computer Graphics and Applications, IEEE, 21(6), 34-47. (2001)
5. Avendaño, V. y Domínguez, L. A. Realidad aumentada: Una exploración al escenario de la virtualidad educativa. Madrid: Editorial Académica Española. (2012)

6. Andújar, J. M.; Mejías, A.; Márquez, M. A. Augmented Reality for the Improvement of Remote Laboratories: An Augmented Remote Laboratory. IEEE Transactions on Education, 54 (3), (492-500). (2011)
7. Prendes Espinosa, C. Realidad aumentada y educación: análisis de experiencias prácticas. Pixel-Bit. Revista de Medios y Educación, (46), 187-203. (2015)
8. Bujak, K. R., Radu, I., Catrambone, R., Macintyre, B., Zheng, R., & Golubski, G. A psychological perspective on augmented reality in the mathematics classroom. Computers & Education, 68, 536-544. (2013)
9. Marzano, R. J. y Kendall, J.S. The new taxonomy of educational objectives. California, Estados Unidos. (2007)
10. Lovos, E., Gibelli, T. I., & Sanz, C. V. Evaluación de materiales educativos digitales que incorporan realidad aumentada: revisión de variables e instrumentos. In *XII Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET, La Matanza 2017).* (2017)

Workshop Computación Gráfica, Imágenes y Visualización (WCGIV)

Coordinadores

Silvia Castro (UNS) - Roberto Guerrero (UNSL) - Oscar Bría (INVAP)

Temas

Modelado Geométrico
Modelado de Fenómenos Naturales
Iluminación y Rendering
Animación
Realidad Virtual y Aumentada
Interacción Humano-Computadora
Diseño Asistido por Computadora
Interfaces Gráficas
Interfaces con Dispositivos No Convencionales
Juegos y Tecnologías del Entretenimiento
GPU y Sistemas Gráficos
Visualización de Información
Visualización Científica
Visualización de Software
Visualización de Algoritmos
Visualización de Terrenos
Visualización de Flujos
Visual Analytics
Generación, Adquisición, Registración, Manipulación, Segmentación, Restablecimiento y Compresión de Imágenes
Correcciones de Color
Codificación y Encriptación de Imágenes
Recuperación de Imágenes
Clasificación y Clusterizado

Inspección Automática de Defectos de Superficie en Baldosas Cerámicas

Lucas Echeverz[1], Matías Melograno[1], Lucas Leiva[1,2]

[1]Universidad Nacional del Centro de la Pcia. de Buenos Aires, Tandil, Argentina,
lucas.echeverz@gmail.com, ds.matias@gmail.com, lleiva@exa.unicen.edu.ar
[2]Universidad FASTA, Facultad de Ingeniería, Mar del Plata, Argentina

Abstract. La industria de fabricación de baldosas cerámicas se sustenta fuertemente en la calidad de la producción lograda. En este sentido, los productos son clasificados en diferentes categorías, y su precio (o descarte) está sujeto a esto. Actualmente este proceso es generalmente realizado por operadores humanos, sin embargo muchas de las industrias pretenden aumentar el rendimiento de producción, así como los rindes, a través de la automatización del proceso. En este trabajo se propone un conjunto de algoritmos que permiten identificar defectos sobre baldosas con texturas aleatorias. El conjunto de defectos está comprendido por blobs (gotas), scratchs (rayaduras), pinholes (agujeros), esquinas, dimensión y bordes. Los resultados obtenidos demuestran que el conjunto de algoritmos desarrollado pueden ser aplicados al problema, y en el futuro ser integrados en un sistema de clasificación automático completo.

Keywords: inspección de superficie, procesamiento de imágenes, industria cerámica.

1 Introducción

La calidad del producto final define el futuro de la industria y sólo una inspección objetiva y estricta puede prometer la mejor calidad del producto [1]. La incorporación de sistemas de inspección automática en la industria (Machine Vision) [2-4] tiene como objetivo recopilar la información necesaria para determinar si el producto se encuentra dentro de los parámetros correctos utilizando bases de conocimiento. Estos sistemas son robustos y operan en tiempo real sobre las líneas de producción, brindando una salida determinística y pueden ser incorporados en ambientes hostiles para los humanos (radiación, temperatura, sustancias tóxicas, entre otros).

El éxito de la implementación de un sistema de inspección visual automática se basa en: la justificación de la aplicación, las consideraciones sistemáticas de la eficiencia y el rendimiento del sistema (considerando la capacidad de integración, robustez, facilidad de uso y la capacidad de incluir inteligencia), la adaptación a las condiciones ambientales existentes, la calidad y resolución de la adquisición de imágenes, el rendimiento de la línea de producción, las técnicas de preprocesamiento de imágenes y el estudio de opciones en relación con el análisis de imágenes.

El control de calidad en la fabricación de baldosas se realiza en entornos no adecuados para los operadores al estar sujetos a altos niveles de ruido, polvo, temperaturas extremas, entre otros. Generalmente estas industrias incluyen en sus líneas de producción un alto grado de automatización [5-6], pero el control de calidad sigue siendo realizado por operarios. En este sentido, la capacidad de un ser humano para llevar a cabo este proceso depende de su entrenamiento, conocimiento y experiencia, pudiendo resultar además en clasificaciones erróneas de los productos sujetas al nivel de fatiga y la subjetividad [7].

El origen de los defectos pueden ocurrir en cualquiera de las fases de producción: desde el comienzo del proceso, reflejándose como irregularidades de la estructura en la composición geométrica del mismo, hasta el final del proceso observándose como fisuras. Durante todo el proceso de fabricación, existe la posibilidad que las baldosas se dañen mientras se transportan de una fase de producción a otra [8]. En [9] se definen ocho tipos de defectos posibles para la industria de fabricación de baldosas:

- *Scratch*: ruptura de baldosa
- *Pinhole*: punto blanco-negro aislado disperso
- *Blob*: gota en la superficie de la baldosa
- *Spot*: discontinuidad del color en la superficie
- *Corner*: ruptura de esquina
- *Border*: ruptura de borde
- *Scratch*: rayadura en la superficie
- *Glaze*: superficie borrosa en baldosa

A nivel regional, existe interés en las industrias cerámicas por integrar en sus líneas de producción sistemas de inspección y clasificación automática de productos en sus plantas. Particularmente, Loimar SRL [10] es una de las fábricas radicadas en la ciudad de Tandil desde hace más de 40 años constituida como una de los principales productos de cerámica roja, dedicándose a la fabricación de tejas, baldosas y ladrillos (Fig. 1). Posee una producción mensual de 60,000 m2 de pisos rústicos.



Fig. 1 Fábrica Loimar.

Si bien existen soluciones comerciales que abordan el problema, como las ofrecidas por Sacmi Ibérica [11] (FlawMaster y ProcessMaster), su costo es elevado y la calibración de estos sistemas para un correcto desempeño puede alcanzar los seis meses. En el ámbito académico, puede encontrarse literatura que aborda la detección de defectos en la industria de baldosas y cerámicos utilizando sistemas de visión [12.]. En general, las investigaciones proponen limitaciones a subgrupos de defectos a detectar y su funcionamiento puede estar basado en reconocimiento de patrones estadísticos, extracción de vectores de características o clasificación de texturas.

En este trabajo se presenta un conjunto de algoritmos que permitan detectar defectos en productos basado en el reconocimiento de patrones estadísticos, con un

nivel de precisión que se ajuste a la necesidad del problema y que puedan ser incorporados en tecnologías de bajo costo para su implementación. Los algoritmos desarrollados comprenden un subconjunto de defectos seleccionados a partir de las necesidades de la industria particular sobre la cual se basa este trabajo: scratch, pinholes, blobs, corners y borders. Además se incorpora un nuevo parámetro de dimensión que es determinado por el área del producto. En la sección 2 se presentan los algoritmos desarrollados para la detección de los defectos y la sección 4 presenta los resultados experimentales de aplicar los algoritmos a los productos provistos por Loimar SRL. Finalmente se detallan las conclusiones y los trabajos futuros.

2 Desarrollo

El desarrollo de los algoritmos para la detección de los diferentes defectos fue realizado para borders, corners, scratchs, pinholes, y blobs. Siendo que los algoritmos de identificación de esquinas, bordes y dimensión comparten la secuencialidad de los pasos estos se presentan en conjunto. No sucede así con los restantes (scratchs, pinholes y blobs).

Los algoritmos parten desde una imagen capturada de una baldosa sobre un fondo con características absorbente, desde una cámara con ubicación cenital. Se utilizó iluminación natural, que puede ser recreada de forma artificial. El entorno de desarrollo para la implementación de los algoritmos fue Matlab R2016b, siendo este un entorno apto para el prototipado de algoritmos de procesamiento de imágenes. Las siguientes secciones describen los algoritmos desarrollados para el reconocimiento de cada uno de los defectos analizados en este trabajo.

2.1 Algoritmo de detección de Borders, Corners y Dimensión

Los algoritmos de verificación de área de la baldosa, así como también el análisis de bordes y la detección de esquinas, están basados en un detector de corners. En particular, el algoritmo seleccionado fue el algoritmo de Harris-Stephens, el cual detecta los cambios de intensidades en una ventana. Para su correcta ejecución, fue necesario incorporar una etapa de preprocesamiento de la imagen, basado en una operación de umbral. Durante esta etapa se descartan los puntos del fondo de la imagen. Sin embargo esta operación puede producir puntos internos dentro de la superficie del producto, que no son relevantes al análisis. Este problema se soluciona mediante el llenado de huecos.

El conjunto de corners detectado es analizado, determinando los vértices de la baldosa. Estos vértices, permiten además realizar un cálculo de área de la superficie permitiendo la verificación de este aspecto. Si bien, se descartan otras verificaciones, como pueden ser anchos y altos, la implementación de estos aspectos carece de complejidad. Por otra parte, el conjunto de puntos obtenido, permite verificar que el producto cumpla con los estándares definidos para su clasificación respecto a los bordes y esquinas.

La secuencia de pasos involucrada para la determinación de la calidad en cuanto a dimensión, bordes y esquinas, se trascribe a continuación.

Paso 1. Umbralado entre un rango [*min,max*], pudiendo ser parametrizado a través de *cornerThreshold*.
Paso 2. Eliminación de elementos con superficie menor a minsize píxeles (*bwareaopen*).
Paso 3. Llenado de huecos (*imfill*).
Paso 4. Detección de esquinas (*Harris-Stephens*).
Paso 5. Clasificación de puntos en vértices de baldosa (*v0,v1,v2,v3*), y puntos erróneos.
Paso 6. Cálculo de la dimensión basada en (*v0,v1,v2,v3*).

2.2 Detección de Blobs

El algoritmo de detección de blobs se basa en dos aspectos: el contraste de las manchas en la superficie, y su aspecto circular. Esto motiva a aplicar, como base de ejecución algoritmos de segmentación que permitan detectar elementos circulares. En este sentido el conjunto de pasos que comprende la técnica desarrollada es el siguiente:

Paso 1. Umbralado (*threshold*).
Paso 2. Eliminación de elementos (*bwareaopen*) con superficie menor a *areaNotIncluded*.
Paso 3. Operación de cierre (*imclose*) con discos de valor indicado por *blobsDisk*.
Paso 4. Llenado de huecos (*imfill*).
Paso 5. Selección de elementos (*bwareafilt*) con superficie en el rango [*minAreaFilt*, *maxAreaFilt*]
Paso 6. Cálculo de redondez de los elementos seleccionados (*getRoundness*)
Paso 7. Selección de elementos con redondez mayor a *thresholdRound*.

2.3 Detección de Pinhole

En comparación con el análisis de blobs, se observa que la naturaleza del problema es similar. Por este motivo el algoritmo desarrollado para la detección de pinholes posee semejanza al presentado anteriormente. Sin embargo, los pinholes poseen una textura menos homogénea que los blobs, por lo que utilizar solo una técnica de umbralado como fase de segmentación no es suficiente. En este caso se incorpora un paso de dilatación, el cual permite unificar las áreas abarcadas por los pinholes. La secuencia de pasos para esta etapa es la siguiente:

Paso 1. Umbralado (*threshold*).
Paso 2. Operación de dilatación (*imdilate*).
Paso 3. Eliminación de elementos (*bwareaopen*) con superficie menor a *areaNotIncluded*.
Paso 4. Operación de cierre (*imclose*) con discos de valor indicado por *pinHoleArea*.
Paso 5. Llenado de huecos (*imfill*).
Paso 6. Selección de elementos (*bwareafilt*) con superficie en el rango [*minAreaFilt*, *maxAreaFilt*]
Paso 7. Cálculo de redondez de los elementos seleccionados (*getRoundness*)
Paso 8. Selección de elementos con redondez mayor a *thresholdRound*.

2.4 Identificación de scratchs

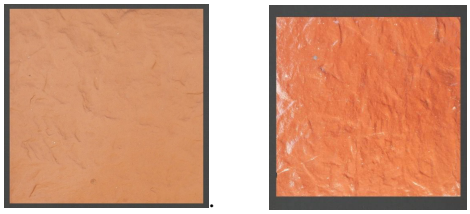
Para la identificación de scratchs se realizaron versiones de algoritmos basadas en detección de bordes (Canny, Sobel, Prewitt), y detección de líneas (Transformada de Hough), en conjunto con umbralados y operaciones morfológicas (Erosión, Dilatación, Apertura y Cierre). Sin embargo, el uso de estas técnicas obligaron a considerar un gran número de parámetros por cada baldosas, y estos parámetros son totalmente dependiente de la baldosa particular a analizar. Además, el número de falsos positivos obtenidos de la ejecución fue alto, por lo que se descartó el uso de estas técnicas.

En consecuencia, se optó por incluir el análisis de texturas utilizando un algoritmo basado en la desviación estándar local (*stdfilt*). El algoritmo desarrollado comprende la siguiente secuencia de pasos presentada a continuación.

Paso 1. Conversión de la imagen a escala de grises (rgb2gray).
Paso 2. Ecualización de histograma (histeq).
Paso 3. Cálculo del desvío estandar local (stdfilt).
Paso 4. Umbralado (thresold) con nivel de iluminación scratchThreshold.
Paso 5. Operación morfológica de cierre (imclose) con un tamaño scratchDisk.
Paso 6. Llenado de los huecos (imfill).
Paso 7. Creación lista de todos los objetos detectados (bwareafilt) con tamaño dentro del rango [minAreaFilt, maxAreaFilt].

3 Resultados experimentales

El conjunto experimental utilizado para realizar la evaluación de los algoritmos está constituido por 32 muestras otorgadas por la empresa Loimar SRL, de sus productos de pisos Fiorentinos con dimensiones de 35cm x35cm, en sus diferentes versiones Laja Vecchia (21) y Curado a fuego (11).



(a) (b)

Fig. 2 Ejemplos de muestras capturadas correspondientes a: a) Laja Vecchia y b) Curado a fuego.

Los dos tipos de productos a analizar (Fig. 2) cuentan con características disímiles, ya que las baldosas del tipo curado a fuego reciben un esmaltado en su superficie. Esto implica que el brillo desprendido de la superficie durante la captura de las imágenes puede afectar a la identificación correcta de los defectos presentes en la superficie (blobs, scratchs y pinholes).

El conjunto de muestras está comprendido por 4 baldosas con defectos en sus bordes, 4 con defectos en las esquinas, 4 presentan fallas de dimensiones, 6 presentan blobs, 8 pinholes y 6 scratchs.

Los algoritmos fueron ejecutados utilizando sobre una PC con un procesador Intel Core i5-4590 CPU @ 3.30 GHZ, 8 Gb de memoria RAM y disco rígido SATA. El sistema operativo utilizado fue Windows 10.

Las imágenes fueron capturadas con una cámara estándar utilizando como sistema de iluminación luz natural, sobre un fondo con capacidades de absorción de luz, de manera de evitar reflexiones que impidan una correcta captura de los defectos del producto. Las resoluciones de las imágenes capturadas fueron de 5152x3864 píxeles.

Tabla 1. Resultados experimentales obtenidos respecto a tasas de acierto y errores, del análisis del lote de baldosas,

Algoritmo	Muestras analizadas	Muestras con fallas	Porcentaje de acierto	Falsos positivos
Bordes	32	4	100%	0
Esquinas	32	4	100%	0
Dimensión	32	4	100%	0
Blobs	17	6	100%	0
Pinhole	17	5	94.7%	0
Scratchs	17	6	94.7%	1

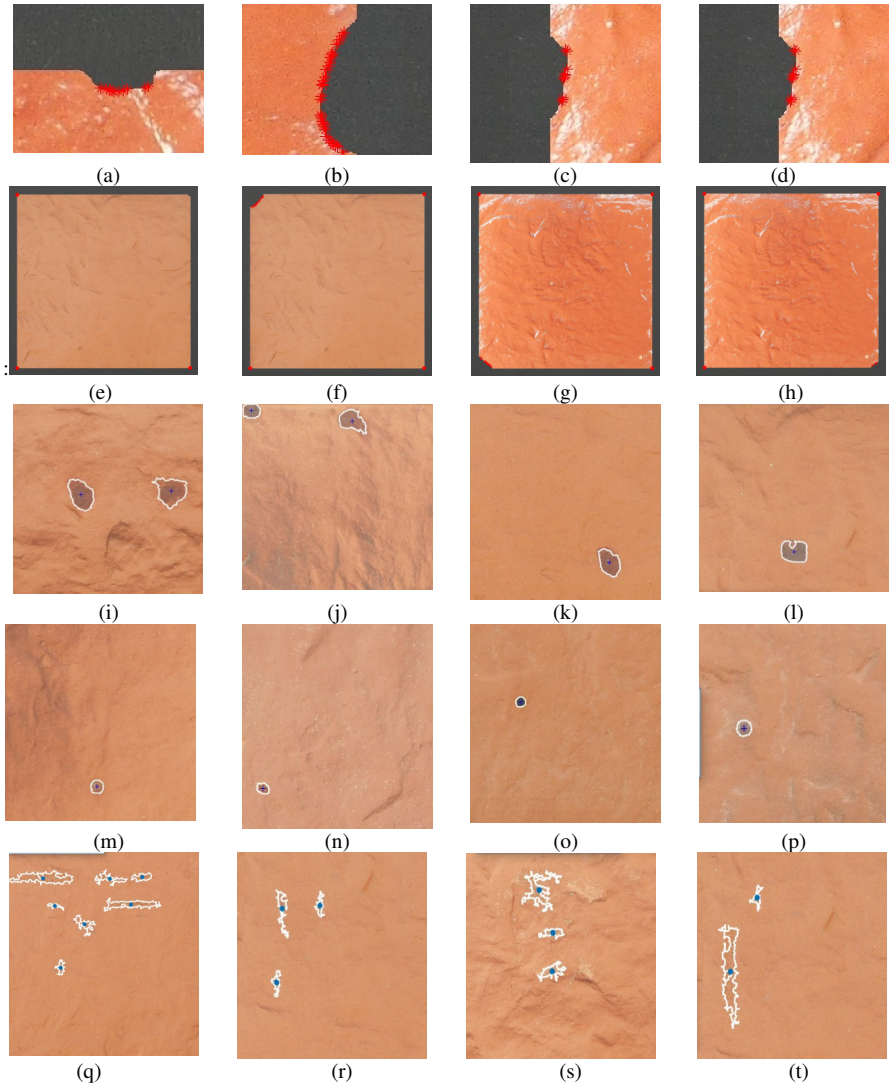
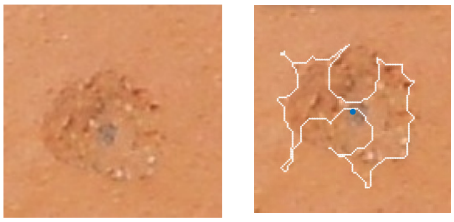


Fig. 3 Resultados obtenidos del proceso para cada uno de los algoritmos: Identificación de defectos en bordes (a, b, c, d); Identificación de defectos en esquinas (e, f, g, h); Identificación de presencia de blobs (i, j, k, l); Identificación de presencia de pinholes (m, n, o, p); Identificación de presencia de scratchs (q, r, s, t).

Del número total de muestras, los resultados relativos al conjunto completo de muestras (Tabla 1) se observa que la primera etapa de clasificación que identifica fallas de bordes, esquinas y dimensión posee una tasa de error nula. Las figuras Fig. 3.(a-h) presentan los resultados obtenidos de las muestras. En este sentido, para la siguiente etapa (identificación de blobs, pinholes y scratchs) se descartan el lote de piezas defectuosas (4 con errores de bordes, 4 con errores de esquina y 4 con errores de dimensiones). Si bien el lote restante se encuentra conformado por un total de 20

baldosas, los algoritmos correspondientes a la segunda etapa (blobs, pinholes y scratchs) presentan dificultades para la identificación de fallas en el subconjunto de baldosas del tipo curado a fuego (con esmalte). Considerando esto, el análisis presentado se realizó sobre un total de 17 piezas, es decir sobre el lote del tipo Laja Vecchia sin defectos de corners, borders ni dimensión.

El algoritmo de identificación de blobs, presentó resultados con la máxima precisión (Fig. 3.i, Fig. 3.j, Fig. 3.k y Fig. 3.l), mientras que el algoritmo de identificación de pinholes, pudo identificar el defecto en 4 de las 5 piezas con presencias de este tipo de errores (Fig. 3.m, Fig. 3.n, Fig. 3.o y Fig. 3.p). El caso particular del falso negativo (Fig. 4.a), se corresponde con un defecto con características muy similar al resto de la superficie, tanto en intensidad de color como de textura. La ausencia de un contorno oscuro característico en este tipo de defectos, provoca que el umbralado, con la parametrización seleccionada, no pueda detectarlo.



(a) (b)

Fig. 4 Resultados en baldosa con pinhole usando: a) algoritmo de detección de pinoles, b) algoritmo de detección de scratchs.

Durante la ejecución del algoritmo de identificación de scratchs, si bien todas las fallas son encontradas (Fig. 3.q, Fig. 3.r, Fig. 3.s y Fig. 3.t), el algoritmo recae en un falso positivo. Analizando la detección errónea, se observa que el defecto encontrado en la superficie no se trata de la presencia de un scratch, sino es el pinhole omitido por el algoritmo de detección de pinholes (Fig 4.b).

De esta manera se puede concluir, que si bien uno de los defectos fue mal clasificado, el total de los errores de superficie del lote fue identificado de manera correcta, excluyendo los productos con esmaltado.

Tabla 2. Promedio de tiempos de cada algoritmo en diferentes computadoras.

PC / Defecto	PC1		PC2	
Resolución	5152x3864	1700x1275	5152x3864	1700x1275
Bordes, Corners y Dimensión	5,33 s	0,6 s	3,26 s	0,34 s
Blobs	0,24 s	0,04 s	0,17 s	0,03 s
Pinholes	0,29 s	0,05 s	0,2 s	0,03 s
Scratchs	0,49 s	0,07 s	0,29 s	0,04 s

Repecto a los tiempos de ejecución, se analizó el tiempo de respuesta utilizando el hardware descripto (PC1), asi como tambien en una PC de mayor rendimiento (PC2), con un procesador Intel Core i7-7700 CPU @ 3.60 GHZ, 16.00 Gb de memoria RAM y disco de estado sólido. Además se utilizaron dos resoluciones de imágenes de

entrada para la evaluación de la performance de los algoritmos. En la tabla 2 se detallan los tiempos de las combinaciones resultantes para cada una de las configuraciones de hardware como resolución.

Los resultados demuestran que si bien el aumento de la potencia del hardware incide sobre los tiempos de respuesta, el aumento de rendimiento máximo alcanzado es inferior a 2X. Sin embargo una reducción del mapeo espacial de la imagen, a través de la utilización de imágenes de menor resolución, presentan una aceleración superior a 5X, e incluso en los algoritmos de bordes, corners y dimensión cercano a 9X. Lo expuesto denota que si se desea un aumento considerable del rendimiento, es recomendable utilizar imágenes de menor resolución, siempre y cuando se considere que la resolución utilizada sea lo suficientemente grande como para permitir identificar los defectos.

Agradecimientos

Los autores del trabajo agradecen a Loimar, por la donación del material utilizado para la realización del mismo. Este trabajo fue parcialmente financiado por la SeCAT de UNICEN (Código de Proyecto 03/C287) y la Secretaría de Investigación de la Facultad de Ingeniería, Universidad FASTA.

Conclusiones

La actividad que involucra la industria de fabricación de baldosas tiene especial interés en realizar adecuados controles de calidad en sus líneas de producción, siendo estas etapas las que definen el impacto en el mercado. En este trabajo se desarrolló un conjunto de algoritmos capaces de detectar defectos de superficie en baldosas cerámicas. El alcance del trabajo comprende la detección de defectos en borders, corners, dimensión, blobs, scratchs y pinholes. Los resultados obtenidos de la ejecución de los algoritmos se encuentran dentro de los requerimientos iniciales de aplicación en cuanto a precisión y demuestran la factibilidad de proveer una solución automática. Se han detectado todos los defectos existentes en las distintas baldosas sin la presencia de falsos positivos. Si bien los algoritmos poseen una latencia considerable, se plantea migrar estas soluciones a sistemas embebidos industriales utilizando hardware especialmente diseñado para procesamiento de imágenes. Los resultados de los algoritmos demuestran además que son aplicables en la detección de defectos sobre superficies de texturas aleatorias.

Los resultados obtenidos demuestran que mayores aceleraciones de rendimiento pueden ser obtenidas utilizando imágenes con resoluciones menores que permitan la identificación de los defectos. En este sentido se recomienda realizar un análisis minucioso respecto del tamaño de los defectos a inspeccionar, y la representación de los mismos en las imágenes de entrada al sistema, durante la especificación del sistema.

Finalmente, se plantea realizar un estudio de modelos de iluminación, que permita una correcta ejecución de los algoritmos sobre el lote de baldosas esmaltadas. Las

características de brillo existente en este tipo de producto, impidió realizar una inspección de superficie satisfactoria.

Referencias

1. D. T. Pham y R. J. Alcock, Smart Inspection Systems: Techniques and applications of intelligent vision, Academic Press, 2003.
2. E. R. Davies, Machine Vision: Theory, algorithms, practicalities, Fourth Edition, Academic Press, 2004.
3. U. S. Khan, «Automatic inspection system using Machine Vision», Proceedings of the 34th Applied Imagery and Pattern Recognition Workshop, pp. 217-222, 2005.
4. H. Golnabi, «Design and application of industrial Machine Vision systems», Robotics and Computer-Integrated Manufacturing, vol. 23, issue 6, pp. 630-637, December 2007.
5. Y. Meena y A. Mittal, «Blobs and Scratchs Detection on Plain Ceramic Tile Surface», International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, issue 7, pp. 647-652, 2013.
6. H. Elbehiery, A. Hefnawy y M. Elewa, «Surface Defects Detection for Ceramic Tiles Using Image Processing and Morphological Techniques», World Academy of Science, Engineering and Technology, pp. 158-162, 2005.
7. S. Singh y M. Kaur, «Machine Vision System for Automated Visual Inspection of Tile's Surface Quality», IOSR Journal of Engineering, vol. 2, issue 3, pp. 429-432, 2012.
8. Ž. Hocenski y T. Keser, «Failure Detection and Isolation in Ceramic Tile Edges Based on Contour Descriptor Analysis», Mediterranean Conference on Control and Automation, 2007.
9. G. M. Atiqur Rahaman, Md. Mobarak Hossain, Automatic Defect Detection and Classification Technique from Image: A Special Case Using Ceramic Tiles, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 1, No. 1, May 2009
10. Loimar, http://www.loimar.com
11. Sacmi Ibérica, http://www.sacmiiberica.com/
12. Karimi, Mohammad H., and Davud Asemani. "Surface defect detection in tiling Industries using digital image processing methods: Analysis and evaluation." ISA transactions 53.3 (2014): 834-844

Virtual and Augmented Reality Therapy Framework for Phobia Treatment

Franco N. Tripi[1], Matías Selzer[1,2,3,4], Juan A. Biondi[1,3,5], Martín Larrea[1,2,3]

[1]Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (DCIC-UNS)
[2]Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET)
[3]Laboratorio de I+D en Visualización y Computación Gráfica, (UNS-CIC Prov. de Buenos Aires)
[4]Comisión de Investigaciones Científicas (CIC Prov. de Buenos Aires)
[5]Laboratorio de Desarrollo en Neurociencia Cognitiva, Instituto de Investigaciones en Ingeniería Eléctrica (IIIE), Departamento de Ingeniería Eléctrica y de Computadoras (DIEC), Universidad Nacional del Sur (UNS - CONICET)
francotripi92@gmail.com, {matias.selzer, jab, mll}@cs.uns.edu.ar

Abstract. Phobias are a common type of anxiety disorder that affects a great number of people. Treating phobias is not always straightforward and require a significant effort from both the patient and the therapist. In recent years, Virtual Reality Exposure Therapy and Augmented Reality Exposure Therapy have emerged to help in phobia treatments by using virtual content. However, most available systems are not free and require expensive hardware. In this paper, we present a free and open-source framework for phobia treatment, designed for both the therapist and the patient. A Virtual Reality scenario for acrophobia (fear of heights) and an Augmented Reality scenario for arachnophobia were developed for validation.

Keywords: Phobia Treatment, Virtual Reality Exposure Therapy, Augmented Reality Exposure Therapy, Android, Unity3d, Human Computer Interaction.

1 Introduction

Phobias are one of the most common types of anxiety disorders, affecting a significant number of people around the world [1]. One of the most commonly used methods for phobia treatment is exposure therapy [2], [3]. The advance of technology contributed to the generation of more Virtual Reality (VR) and Augmented Reality (AR) audiovisual content than ever before. Because of that, Virtual Reality Exposure Therapy (VRET) and Augmented Reality Exposure Therapy (ARET) emerged as novel alternatives for phobia treatment. They are innovative tools for conducting exposure therapy by which the users can be systematically exposed to specific feared stimuli in a safest environment.

Many research studies and clinical experiments have shown that VR and AR can be used as a highly efficient treatment for obsessive-compulsive, anxiety, and phobic disorders [4], [5], [6], [7].

Although these studies presented many different approaches, they have only focused in very specific applications for particular scenarios. For this reason, we consider that a general approach is necessary.

On this paper, we present an open-source and free framework to facilitate the implementation of VRET and ARET scenarios for phobia treatment. This work supports the implementation of affordable alternatives for phobia treatment by using low-cost VR and AR systems.

This paper is organized as follows. The next two sections discuss the efficiency of VR and AR for phobia treatment and the systems and prototypes presented in the literature. Section 4 details the proposed system. In Section 5 a detailed description of the framework's architecture is presented. Then, two study cases developed for this work are presented in Section 6. Finally, the last section outlines the conclusions and future work.

2 Background

VR is a technology that generates a complete Virtual Environment (VE) by replacing the existing physical environment and providing users with new ways of interaction and navigation. In contrast to VR, AR is a technology that uses virtual elements to augment the existing environment. While AR users maintain their sense of presence in the real world, VR users are completely immersed in the virtual world [8].

The literature provides strong support for the effectiveness of exposure therapy for anxiety disorders [9]. VR and AR can provide a high level of presence and immersion within the exposure, which may allow for higher engagement and fewer distractions than other methods. They also provide the ability to stimulate multiple senses during the exposure, through tailoring the visual and auditory stimuli, arousing the patient's sense of touch, or by adding relevant smells. The dynamic interaction between the patient and the VE, and the ability to engage multiple senses, facilitates a sense of truly "being there". Indeed, a previous study discovered that adding additional types of sensory input within VR, such as olfactory and auditory cues, increase the participant's sense of presence and memory [10].

VRET is a well-established technique to treat psychological disorders such as phobias. In a published meta-analysis on the efficacy of VRET on different phobias, the authors concluded that real life interventions were not significantly more effective than VRET [11].

AR also has great potential for phobia treatment because virtual objects are shown as if they belong to the real world. Hence, patients can see their own body and interact with the virtual objects in a more natural fashion [12]. Despite this, AR systems for phobia treatment are not as common as VR systems.

A more recent meta-analysis discussed the effects of VR interventions for anxiety disorders [13]. They considered disorders such as social phobia, panic disorder with or without agoraphobia, and post-traumatic stress disorder. The authors reported no overall difference between VRET and cognitive behavior interventions on behavioral assessments and post-treatment.

Other studies have shown the benefits of ARET for insect phobia treatment. Chicchi et al. [5] provide a review of recent studies on the evaluation and treatment of psychological disorders. It details how this technology can be used in the psychology field as an efficient tool for phobia therapy.

Botella et al. [6] describe their AR system and how it is used to activate anxiety in participants who suffer cockroach or spider phobia. They also discuss the extent to which participants considered the experience as real. The study showed the advantages of using AR in cockroach phobia treatment as all participants improved significantly in all outcome measures after the treatment.

3 Previous Works

The literature presents many VR and AR systems specifically designed for phobia treatment [14], [7], [15], [16], [17]. Most of them usually consist of two main components. First, an application for the patient to experience a particular scenario. Second, an application for the therapist to configure the scenario for the patient. The therapist's application may contain more features such as live video streaming of the patient's view, stimuli tuning, or patient's data storage.

PsiusVR [14] is a paid VR system that offers a real-time streaming of the patients' view, biofeedback sensor for patient monitoring, and automatically-generated reports. It consists of a web application running on the therapist computer, and a mobile VR application that runs on a smartphone. The patients must use a VR headset to run the smartphone application. The system presents many scenarios for different phobia treatments, such as acrophobia, agoraphobia, claustrophobia, amaxophobia, fear of flying, needles, animals, and public speaking.

NeuroVR [7] is a free and open-source VR platform. It consists of two main components, the Editor and the Player. The Editor is used to visually create scenarios without requiring programming skills. The user can choose from a database of 2D and 3D objects, placing them into a pre-designed virtual scenario. The Player is used to navigate and interact with the VEs created. The system can be visualized on both a regular PC and a head-mounted display (HMD). However, the use of a smartphone as HMD is not supported.

PhoVR [18] is a VRET system developed in Unity3D to give patients a presence in a VE to face their fears. This system focuses on arachnophobia, acrophobia and agoraphobia. By using the Microsoft Kinect, the patients are able to move and see their bodies in the visualization. However, that visualization is only previewed on desktop PC, and HMDs are not supported. Furthermore, an internet connection is required to communicate the patient and the clinician.

Virtually Better [15] is a company that develops VR applications for many disorders such as phobias, post-traumatic stress disorder (PTSD), and stress and pain management. They developed applications that consist of two parts, a mobile application which runs on iPhone and Android, and a desktop application for Windows and MacOS. This system offers scenarios to treat the fear of flying, acrophobia (elevator and bridges experiences), fear of public speaking, and fear of storms. In addition, they developed a system to conducting exposure therapy in two war-zone virtual

environments: Iraq and Afghanistan; and a system for treating addictions such as alcohol and tobacco.

Some studies also applied VRETs to other types of phobias. In [16] they use a VRET to treat dental phobia. Their system contains a desktop application for the therapist to control the stimuli, and a VR application for the patient. The system uses an Oculus Rift HMD to present the VE to the patients. To enhance the immersion of the VE, the typical smell of a dental clinic was introduced. Their results suggest that in order to establish the efficacy of VRET in the treatment of dental anxiety, more carefully controlled trials are required.

Other studies take advantage of the fusion between the real and virtual world that AR provides. Fatharany et al. [19] present a design and implementation of an AR application for cockroach phobia therapy. This system was developed using Unity3D and the OpenCVForUnity library. It only runs on desktop PC by using an external camera. Thus, the patients have to look at the computer screen to be immersed in the augmented scenario.

Another AR system for treating spiders and cockroaches phobias is presented in [17]. That system uses ARToolKit and VR Modeling Language (VRML) to incorporate AR virtual objects. The patient must wear a HMD and the therapist has a live video streaming of the patient's view. To make this possible, they used a creative NX-Ultra camera attached to the HMD.

Corbett et al. [12] developed an interactive AR system that contains a 3D representation of the real environment and the patient's body. Hence, a virtual spider that is presented to the patient can move and react to the real environment in a very realistic manner. Furthermore, the spider can walk up, around, or behind real objects and can be touched, carried, and occluded by the patient. This system uses the Microsoft Kinect to track the patient's body and movements.

Most of the mentioned systems are not free or open-source. Whenever a new scenario to the treat a patient is needed, a new system must be implemented from scratch. On the other hand, not many therapy systems support low-cost VR or AR. The therapist should therefore invest a lot of money in a fully equipped system. To the best of our knowledge, there is no open-source and free framework for the implementation of VRET and ARET scenarios for phobia treatment.

4 Proposed System

In this study, we present a VRET and ARET open-source and free framework for phobia treatment. This framework was developed to be easily customized, allowing the creation of VR and AR environments using the cross-platform game engine Unity3D.

The system consists of two main components. The desktop application that is used by the therapist as a control and configuration tool, and the mobile application that is used by the patients to get immersed in the virtual scenarios.

The framework is currently available on GitHub, and the detailed guidelines for the implementation can be found in the repository website. The framework architecture is presented next.

5 Framework's Architecture

In order to create a new scenario for phobia treatment, some components must be implemented. Those components are described next.

5.1 Desktop Application

The desktop application is used by the therapist as a control and configuration tool. By using the desktop application the therapist can establish a connection with more than one patient at the same time. Once the connection is established, the desktop application can communicate to the mobile application to manipulate the scene and to control the exposure levels.

To provide the therapist with the possibility to see what the patient is watching in real-time, a video streaming feature was implemented. Hence, the therapist can see and better understand what the patient is experiencing at any time.

Finally, the desktop application is used to select among the different available scenarios. This provides the possibility to have more than one phobia scene in the same desktop application. Once a scenario is selected, the application loads and shows on the screen the necessary User Interface (UI) elements to manipulate the features of that particular scenario.

5.2 Mobile Application

The mobile application is used by the patients to get immersed in the VR and AR scenarios. The application runs on a mobile phone inside a low-cost VR headset. In a VR scenario, patients feel fully surrounded by the virtual world. They can turn their head and see the virtual world as if they were there. On the other hand, in an AR scenario, patients see the real world that is captured by the mobile phone's camera. In order to present the virtual objects, a regular AR marker is used. When the application detects that marker, a virtual object is displayed on it. Thereby, when patients see the marker through the application, it provides the sensation that the virtual objects belong to the real world. It is important to notice that patients are experiencing the AR scenario through a headset with their full field of view. This improves the immersion level compared to the typical AR applications, where the mobile phone is on the users' hands.

The mobile application contains every possible scenario that can be chosen in the desktop application. A Camera Component is used to detect QR codes for connection, to detect the AR marker, and to record what the patient is watching for real-time video streaming. The mobile application also contains a Network Manager Component to manage the connection, and a Scene Manager Component to control the events execution. These components are explained next in more detail.

5.3 Network Manager Component

A communication between the desktop application and the mobile application must be established. As the system is designed to be used by more than one patient at the same

time, a client-server model was implemented. The desktop application acts as a server and the mobile application must know the desktop PC local IP address. For this reason, the desktop PC and the mobile phone must be connected to the same local network.

Instead of manually typing the IP address, a more practical method was implemented. The local IP address of the desktop application is encoded in a QR code that is presented on the screen. Then, when the mobile application starts, this QR code must be scanned, automatically establishing the connection.

The Network Manager Component manages the connection between the desktop PC and the smartphone. Once the therapist selects a phobia scenario on the Desktop application, the Network Manager Component is instantiated and awaits for the phone connection. On the phone side, the Network Manager Component is already created and connects to the Desktop application as mentioned. Once the connection has been established, the Scene Manager Component is created on both the desktop and mobile application.

5.4 Scene Manager Component

Once the connection has been established, this component handles the messages exchange between the desktop and mobile applications. The system uses Remote Procedure Calls (RPCs) to send and receive specific messages through the network. On the desktop application, the Scene Manager Component contains the method definitions to perform the RPCs that will be executed on the mobile system. When the therapist selects one of the available actions on the UI, a specific RPC is performed.

On the other hand, for the RPCs to work properly, the Scene Manager Component implemented on the mobile application must have the same method definitions as the desktop application. These methods implement actions to modify the scene according to the selected event on the desktop application.

6 Case Study

In order to test the framework, two phobia treatment scenarios were developed. First, we developed a VR scenario for acrophobia (fear of heights), a disorder that affects approximately 5% of the general population [20] and is one of the most common phobias treated by using VR [21, 22]. Treating this phobia without VR can be a very complicated and dangerous process since it requires the patient to be exposed to different heights.

Then, we developed an AR scenario for arachnophobia (fear of spiders), one of the most common animal phobias, that affects approximately 3.5-6.1% of the general population [23]. The use of AR for animal phobia treatment has become very popular as it creates the sensation of being in front of the animal. Also, the exposure to real animals can be very complicated, expensive, and mostly, very dangerous.

This implementation was tested on a desktop computer running Windows 10 and a smartphone running Android 6.1. The implemented scenarios are described next, and the source code and the executable versions are available on GitHub.

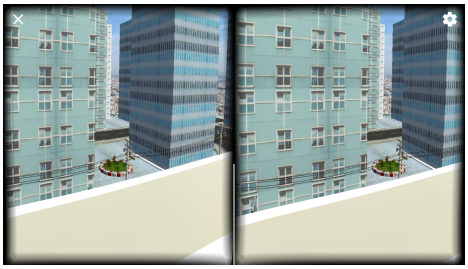


Fig. 1. Android Application for acrophobia treatment. The patient is approaching the balcony's edge, with walls activated.



Fig. 2. Desktop Application. UI Control Scene for acrophobia treatment.

6.1 Acrophobia Scenario

In the proposed VR scenario for acrophobia treatment, the patient is located in a room with a balcony inside a building in the middle of a city. Patients are able to walk inside the VE by using a wireless joystick. Thus, they can approach the edge of the balcony at their own pace, as depicted in Figure 1.

By using the desktop application, the therapist can modify the height of the building, i.e. the floor where the patient is located. Also, the therapist can hide or show the floor and walls of the balcony. Figure 2 shows the UI in which the therapist can progressively increase the exposure levels, avoiding abrupt jumps.

6.2 Arachnophobia Scenario

An AR scenario was developed for arachnophobia treatment. It is important to remember that the patient is wearing a low-cost VR headset but, unlike in VR, in this AR scenario the patients see the real world that is captured by the mobile phone's camera. As shown in Figure 3, when the patient looks at a specific AR marker that is over the table, a 3D virtual spider appears. The spider provides the illusion of being walking over the table.

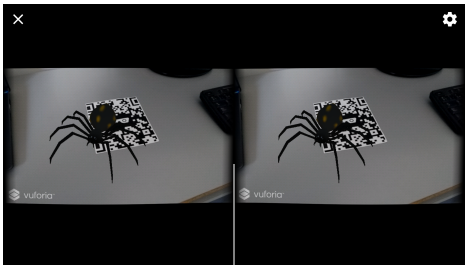


Fig. 3. Android Application for arachnophobia treatment. The patient can see a virtual spider walking over the table.

By using the desktop application, the therapist can change the spider's size, make it walk, and create additional spiders. As shown in Figure 4, the therapist can progressively increase the exposure levels.

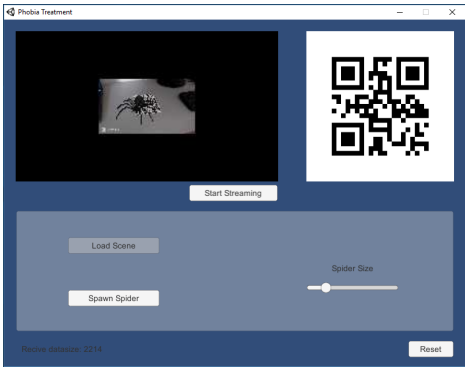


Fig. 4. Desktop Application. UI Control Scene for arachnophobia treatment.

7 Conclusions

VR and AR technology provides the opportunity to include aspects of feared stimuli that could otherwise be too expensive or impractical. Additionally, it allows the therapist to control every aspect of the exposure. This level of control also allows for repetition exposures which may not be possible in real life.

Even though VRET and ARET have existed for many years, the literature presents different systems and applications that were designed only for one specific purpose. On the other hand, the available commercial software is not free and may require expensive hardware.

In this paper, we have presented a VR and AR framework for the creation and customization of VEs for phobia treatment. This framework is free and open-source

and is designed to be used with low-cost VR and AR devices. By using this framework, the applications for both the therapist and the patients are created.

In order to show the potential of the framework, two phobia scenarios were developed addressing acrophobia and arachnophobia treatments. These prototype scenarios showed the framework's flexibility in terms of the immersion and extensibility. Informal tests were performed with volunteers that collaborated by giving feedback about the immersion sensation they felt. Their comments helped adjusting the scenarios' features as well as setting the exposure levels behaviors. Even though the visual quality produced by a smartphone and a low-cost VR headset is low compared to high-end VR devices, all the participants indicated an appropriate sensation of immersion. Also, most of the volunteers considered that the implemented behaviors were suitable to increase the exposure levels.

The implementation of the case studies and the test suggested the framework's efficiency. However, some limitations are worth noting. Although the system was tested by participants, formal tests were not performed. Future work should therefore test the implemented scenarios in real situations with real patients, and with the collaboration of experienced therapists or psychologists. This would corroborate the efficacy of the developed system and would provide new ways of improvement. Also, more studies are needed in order to test the performance of VRET and ARET compared to standard exposure therapy. Being an open-source system, we expect the VR research community to collaborate and improve it by offering new alternatives, not only for phobia treatments, but also for others psychological disorders that can be faced with this novel technology.

Acknowledgment

This work was partially supported by the following research projects: PGI 24/N037 and PGI 24/ZN29 from the Secretaría General de Ciencia y Tecnología, Universidad Nacional del Sur, Argentina.

References

1. Eni S Becker, Mike Rinck, Veneta Türke, Petra Kause, Renee Goodwin, Simon Neumer, and Jürgen Margraf. Epidemiology of specific phobia subtypes: findings from the dresden mental health study. *European Psychiatry*, 22(2):69–74, 2007.
2. Barbara Olasov Rothbaum and Ann C Schwartz. Exposure therapy for posttraumatic stress disorder. *American journal of psychotherapy*, 56(1):59, 2002.
3. Edna B Foa and Michael J Kozak. Emotional processing of fear: exposure to corrective information. *Psychological bulletin*, 99(1):20, 1986.
4. Azucena Garcia-Palacios, Cristina Botella, H Hoffman, and Sonia Fabregat. Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. *Cyberpsychology & behavior*, 10(5):722–724, 2007.
5. Irene Alice Chicchi Giglioli, Federica Pallavicini, Elisa Pedroli, Silvia Serino, and Giuseppe Riva. Augmented reality: a brand new challenge for the assessment and treatment of psychological disorders. *Computational and mathematical methods in medicine*, 2015, 2015.

6. Cristina Botella, Juani Bretón-López, Soledad Quero, Rosa Baños, and Azucena García-Palacios. Treating cockroach phobia with augmented reality. *Behavior Therapy*, 41(3):401–413, 2010.
7. Giuseppe Riva, Andrea Gaggioli, Alessandra Grassi, Simona Raspelli, Pietro Cipresso, Federica Pallavicini, Cinzia Vigna, Andrea Gagliati, Stefano Gasco, and Giuseppe Donvito. Neurovr 2-a free virtual reality platform for the assessment and treatment in behavioral health care. In *MMVR*, pages 493–495, 2011.
8. Oliver Baus and Stéphane Bouchard. Moving from virtual reality exposure-based therapy to augmented reality exposure-based therapy: a review. *Frontiers in human neuroscience*, 8:112, 2014.
9. Antonia N Kaczkurkin and Edna B Foa. Cognitive-behavioral therapy for anxiety disorders: an update on the empirical evidence. *Dialogues in clinical neuroscience*, 17(3):337, 2015.
10. Huong Q Dinh, Neff Walker, Larry F Hodges, Chang Song, and Akira Kobayashi. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Virtual Reality, 1999. Proceedings., IEEE*, pages 222–228. IEEE, 1999.
11. Mark B Powers, Snorri R Sigmarsson, and Paul MG Emmelkamp. A meta–analytic review of psychological treatments for social anxiety disorder. *International Journal of Cognitive Therapy*, 1(2):94–113, 2008.
12. Sam Corbett-Davies, Andreas Dünser, and Adrian Clark. An interactive augmented reality system for exposure treatment. In *Mixed and Augmented Reality (ISMAR-AMH), 2012 IEEE International Symposium on*, pages 95–96. IEEE, 2012.
13. David Opriş, Sebastian Pintea, Azucena García-Palacios, Cristina Botella, Ştefan Szamosközi, and Daniel David. Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. *Depression and anxiety*, 29(2):85–93, 2012.
14. Website. Psious vr therapy. `https://www.psious.com/`. Accessed: 2018-04-12.
15. Website. Virtually better. `http://www.virtuallybetter.com/`. Accessed: 2018-04-12.
16. Kumar Raghav Gujjar, Ratika Sharma, and Ad De Jongh. Virtual reality exposure therapy for treatment of dental phobia. *Dental update*, 44(5):423–435, 2017.
17. M Carmen Juan, Mariano Alcañiz, Jérôme Calatrava, Irene Zaragozá, Rosa Baños, and Cristina Botella. An optical see-through augmented reality system for the treatment of phobia to small animals. In *International Conference on Virtual Reality*, pages 651–659. Springer, 2007.
18. M Brandon Haworth, Melanie Baljko, and Petros Faloutsos. Phovr: a virtual reality system to treat phobias. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*, pages 171–174. ACM, 2012.
19. Fiandra Fatharany, Anny Yuniarti, and Ridho Rahman Hariadi. Design and implementation of markerless augmented reality application for cockroach phobia therapy using adaptive threshold. *Jurnal Teknik ITS*, 5(2):A317–A320, 2016.
20. Carlos M Coelho and Guy Wallis. Deconstructing acrophobia: physiological and psychological precursors to developing a fear of heights. *Depression and anxiety*, 27(9):864–870, 2010.
21. Carlos M Coelho, Allison M Waters, Trevor J Hine, and Guy Wallis. The use of virtual reality in acrophobia research and treatment. *Journal of Anxiety disorders*, 23(5):563–574, 2009.
22. Paul MG Emmelkamp, Mary Bruynzeel, Leonie Drost, and Charles AP G van der Mast. Virtual reality treatment in acrophobia: a comparison with exposure in vivo. *CyberPsychology & Behavior*, 4(3):335–339, 2001.
23. Christiane Eichenberg. Application of'virtual realities' in psychotherapy: Possibilities, limitations and effectiveness. In *Virtual reality*. InTech, 2010.

Método de estimación de *head pose* para la obtención del punto de atención en base a una cámara web estándar

Camila García[1, 2] and Juan P. D'Amato[1, 3]

[1]PLADEMA, UNICEN, Tandil, 7000, Argentina
[2]Comisión de Investigaciones Científicas, Provincia de Buenos Aires, Argentina
[3]Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

Abstract. En toda interfaz de Interacción Humano-Computadora (IHC), se requiere un mecanismo específico que permita conocer el lugar donde el usuario tiene centrada su atención. Ciertos enfoques de la IHC se apoyan en la Visión Computacional para obtener esta información a partir del análisis del rostro. En los últimos años, han surgido muchos trabajos relacionados, en los cuales se propone estimar el punto de atención en una escena a partir de la orientación del rostro (o *head pose*). En los mismos, se emplea mayoritariamente hardware de captura específico (como cámaras infrarrojas, cámaras en *stereo* o cámaras 3D), que suele ser muy costoso e inacccesible. Por otro lado, existen diferentes bibliotecas de análisis facial que, tal como se encuentran actualmente, no pueden ser utilizadas para la tarea perseguida. En este trabajo se presenta una versión inicial de un método de estimación de la *head pose* para IHC, basado en reconstrucción 3D y seguimiento de características faciales. Uno de los aportes presentados en este paper es la definición de un pipeline de procesamiento unificado que pueda funcionar en tiempo real. Adicionalmente, se incluye el estudio de factibilidad del uso cámaras web estándar, incluídas en las notebooks actuales. Para validar el esquema propuesto, se comparó la precisión de la estimación en dos implementaciones basadas en diferentes bibliotecas de análisis facial. Se realizó un estudio preliminar con 3 usuarios en diferentes escenarios.

Keywords: head pose, computer vision, HCI, visual tracking

1 Introducción

Para construir interfaces inteligentes de computadoras, es necesario que el sistema conozca la intención del usuario y pueda inferir la zona o el punto en donde el mismo tiene depositada su atención. Dado que el movimiento de la posición de la cabeza de una persona (*head pose*, en inglés) y la dirección de su mirada (*gaze*) están profundamente relacionadas con su atención (e intención), la detección de dicha información puede ser utilizada para respaldar interacciones que no requieran el contacto con dispositivos periféricos [1], [2], [3].

La entrada de datos controlada por la posición del usuario o su mirada, es especialmente atractiva para la selección de elementos y tareas de seguimiento

de blancos porque utiliza los movimientos inherentes a estas tareas como forma de control. Esta clase de interfaces resulta ideal, sobre todo, en aquellos casos en los que la motricidad del usuario se halla comprometida [4], [5].

Los elementos de captura disponibles en la actualidad, han impulsado el avance de las capacidades de procesamiento computacional de las imágenes digitales. Una gran área del campo de procesamiento de imágenes digitales está relacionada con la detección facial, tarea que ha dado lugar a una amplia variedad de análisis y aplicaciones con los más diversos objetivos. Entre las posibilidades que se presentan ante el avance del análisis facial en imágenes se encuentra, precisamente, la capacidad de estudiar la posición de la cabeza de un usuario.

En visión computacional, la pose de un objeto se refiere a su orientación relativa y posición respecto a una cámara. La pose puede variar tanto moviendo el objeto con respecto a la cámara, o la cámara con respecto al objeto. Para el análisis de la posición de la cabeza se ha utilizado, en gran parte, el reconocimiento de puntos de referencia faciales como base [6], [7].

Resulta evidente, entonces, que el procesamiento digital de imágenes puede ser utilizado como cimiento hacia una Interacción Humano-Computadora diferente. Ello permite pensar en el movimiento cefálico del usuario como una variable de control para su comunicación con la tecnología.

Actualmente, el mercado cuenta con la presencia de diversos dispositivos que realizan el análisis de la mirada o de la posición de la cabeza. El desafío actual en este campo se encuentra en el empleo de algoritmos que no requieran hardware específico para la captura de imágenes, sino que puedan realizar un procesamiento robusto y confiable en base a cámaras web estándar y sin costos adicionales.

En los últimos veinte años se ha enfatizado el impacto que tendrían las cámaras embebidas en el avance de los métodos de determinación de la mirada [8], permitiendo una mayor disponibilidad de herramientas asistenciales. Como prueba de ello, es posible ver diversos trabajos en los que se ha abordado la problemática desde entonces, empleando mecanismos que, en su procesamiento, logran compensar las dificultades de utilizar esta clase de dispositivos [9], [10], [12].

2 Método propuesto

Idealmente, tras el procesamiento de una imagen conteniendo una cara, podría obtenerse información sobre la posición de la cabeza respecto a la pantalla y utilizar los ángulos para proyectar un punto a la pantalla. Se presenta en esta sección una descripción general del método seguido para la estimación del punto de atención de un usuario, partiendo desde la captura de la imagen del mismo frente a la cámara de una computadora, con su posterior procesamiento y análisis. La implementación del método se realizó en una aplicación en C++ y utilizando OpenCV [13].

El pipeline implementado puede exponerse de manera general como se indica en la Figura 1. Se ha pretendido desarrollar un método de procesamiento unifi-

cado de forma que, definiendo el conjunto de datos de entrada requerido para su funcionamiento, pueda ser utilizado en conjunto con cualquier herramienta de análisis facial que genere tal información.

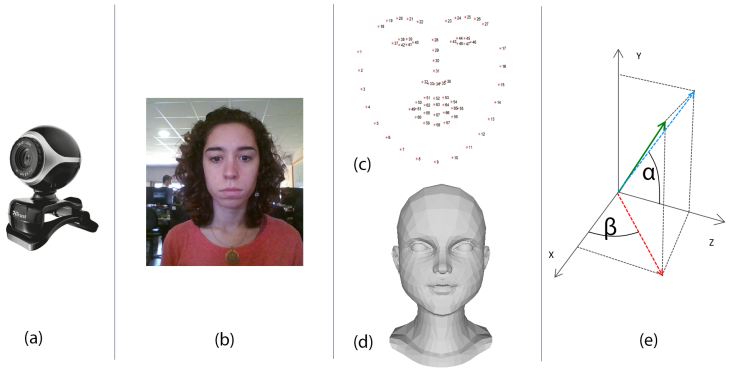


Fig. 1. Pipeline general con las fases del método seguido. (a) Dispositivo de captura, cámara web estándar; (b) Captura de *frames*; (c) Procesamiento 2D y obtención del conjunto de *landmarks* faciales; (d) Malla 3D obtenida mediante el análisis de imágenes o provista de forma estática; (e) Estimación de la transformación y cálculo de los ángulos de rotación de la cabeza.

2.1 Captura de frames

Como ha sido mencionado, para efectuar la captura de imágenes fue propuesta la utilización de un dispositivo disponible al público general por un bajo costo. En la mayoría de los casos, las computadoras portátiles modernas cuentan con una cámara web estándar embebida. Esta clase de periféricos no provee una alta definición en las imágenes capturadas, característica que afecta directamente la precisión y fiabilidad de los algoritmos de procesamiento utilizados en base a ellas. Sin embargo, teniendo en cuenta que el objetivo del análisis realizado en este trabajo se encuentra directamente influenciado por la accesibilidad, se consideró preferible la disminución de la precisión en pos de la posibilidad de brindar capacidad de acceso a un mayor número de personas.

2.2 Análisis 2D de la imagen

El proceso comienza con la extracción de información a partir de una imagen capturada con una cámara web. Sobre ésta, pueden ser aplicados diversos algoritmos de procesamiento que permiten detectar la presencia o ausencia de rostros y, ante la presencia de uno o más, realizar la extracción de puntos característicos faciales o *landmarks*. Esta etapa es determinada como la entrada al

método implementado y se espera, cualquiera sea la biblioteca empleada para su realización, la obtención de un conjunto mínimo de 9 landmarks faciales.

2.3 Modelo 3D facial

Para poder estimar la posición de un objeto respecto a una cámara, se requiere un modelo en tres dimensiones del mismo, que establezca el punto de partida desde el cual analizar sus movimientos. En el caso de estudio, el escenario ideal implica una malla 3D que replique el rostro de cada usuario particular, reduciendo las posibilidades de introducción de errores en los cálculos al proveer las proporciones particulares a cada muestra facial. Sin embargo, es posible reducir la exigencia de los requerimientos utilizando únicamente un conjunto reducido de puntos 3D, correspondientes a una selección de *landmarks* específicos para la detección en el análisis 2D. En los casos en los que no resulta factible obtener un modelo facial del usuario específico que utilizará el sistema, es una posibilidad emplear uno genérico, aunque la precisión de los resultados obtenidos en base al mismo podría verse comprometida.

2.4 Obtención de una matriz de transformación

Contando con los puntos fiduciales 2D y el modelo 3D de la geometría facial, es posible llevar a cabo la identificación de la transformación sufrida por el rostro en un frame determinado respecto a su posición neutral. Dicha transformación es representada por la traslación y las rotaciones en los tres ejes del modelo. La realización de tal tarea fue basada en el algoritmo *solvePnP* provisto por *OpenCV*, que se apoya en el modelo de cámara estenopeica para definir la vista de una escena determinada como la proyección de puntos 3D en el plano de imagen usando una transformación de perspectiva:

$$m' = A[R|t]M', \tag{1}$$

donde: A es la matriz de parámetros intrínsecos de la cámara; [R|t] la matriz de rotación-traslación; t el vector de traslación y M la matriz de los puntos 3D del modelo. La matriz A es independiente de la escena, por lo que puede ser establecida al comienzo del procesamiento y re-usada, utilizando aproximaciones relativas a las características de las imágenes capturadas cuando no se cuenta con información específica sobre las distancias focales o el punto principal de la cámara en uso. Considerando este modelo y la información disponible, el algoritmo previamente mencionado *solvePnP* es capaz de encontrar [R|t], la matriz de rotación-traslación. La misma, expresa los ángulos de rotación percibidos en el rostro con tres grados de libertad, además del desplazamiento del mismo respecto a su posición original.

2.5 Proyección del puntero a la pantalla

Obtenida la matriz de rotaciones, resulta posible extraer los ángulos de rotación correspondientes a los ejes de interés para la estimación de un punto de atención

en pantalla. Las rotaciones relevantes a esta tarea son *yaw* (β) (o guiñada, el movimiento con centro en el eje vertical, que permite desplazar la mirada a lo largo del eje horizontal) y *pitch* (α) (o cabeceo, el movimiento con centro en el eje horizontal, con el cual se puede variar el punto de atención a través del eje vertical). La combinación de estos dos grados de libertad son los que permiten recorrer el plano de la pantalla, pudiendo descartar el movimiento alrededor del eje z (*roll* o alabeo).

$$m_x = \left(\frac{\beta}{\beta_{max}}\right) \times \left(\frac{w}{2}\right) + \left(\frac{w}{2}\right),\ m_y = \left(\frac{\alpha}{\alpha_{max}}\right) \times \left(\frac{h}{2}\right) + \left(\frac{h}{2}\right). \qquad (2)$$

El puntero proyectado a la pantalla es obtenido con las ecuaciones en (2), donde m_x y m_y son las coordenadas de desplazamiento, w y h los datos de ancho (*width*) y alto (*height*) de la pantalla, y α_{max} y β_{max} los ángulos máximos considerados para pitch y yaw respectivamente, correspondientes a movimientos máximos establecidos para alcanzar los extremos de la pantalla [14].

2.6 Estabilización de puntero

La salida del pipeline implementado está condicionada por los landmarks 2D resultantes de la utilización de una biblioteca externa de análisis facial y recibidos como entrada para el procesamiento. Considerando esto, la proyección del punto de atención a la pantalla entre un frame y el siguiente puede sufrir fluctuaciones que deberán ser corregidas. Landmarks erróneamente detectados o una detección poco estable generan un efecto de arrastre del error que debe ser corregido de acuerdo al contexto. Para ello, se añadió a la etapa final una implementación del filtro de Kalman [15], un estimador que puede inferir parámetros de interés a partir de observaciones poco precisas.

La técnica fue aplicada sobre los puntos de atención obtenidos, buscando lograr la reducción del ruido percibido. Adicionalmente, permitió eliminar un gran porcentaje de los casos de error que, introducidos en alguna de las etapas del método, repercutían en el resultado final con el puntero proyectado.

3 Resultados

Para todos los experimentos, se utilizó el mismo equipamiento: una notebook con un procesador Intel i7-6500 2.50GHz, 8GB de RAM, cámara web integrada con resolución máxima de 1280x720px, pantalla de 15.6" con una resolución de 1366x768 y bajo el sistema operativo Windows 10 de 64bits.

Las características faciales fueron detectadas con dos herramientas gratuitas que realizan análisis facial 2D a partir de imágenes capturadas con cámaras web estándar: *OpenFace* [10] y *RealSense* de Intel [11]. En el caso de *OpenFace*, el procesamiento efectuado provee, de forma adicional a los landmarks faciales en 2D, un modelo 3D ajustado a la geometría del usuario. Por el contrario, *RealSense* sólo realiza el reconocimiento de puntos característicos en 2D. Para

el empleo del método implementado con tal herramienta, se obtuvo una malla 3D de cada uno de los usuarios que participaron del experimento mediante la funcionalidad provista por la demo de Jackson et al. [16].

3.1 Experimentos realizados

Los experimentos realizados responden a dos características consideradas relevantes a la hora de evaluar la factibilidad del método a ser usado como base para la Interacción Humano-Computadora.

Pruebas de seguimiento con una trayectoria rectangular. Resulta necesario, para considerar la posibilidad de utilizar este método como la base de una interfaz de IHC, asegurar la capacidad de realizar movimientos fluidos con el cursor en pantalla. Para ilustrar tal requerimiento, puede analizarse el empleo de un teclado virtual en pantalla que utilice escritura basada en gestos [17]. Esta clase de entrada se centra en el trazado de diversos patrones sobre los caracteres. En el trabajo publicado por [18] se plantean los conceptos de *pose tracking* y *gaze tracking*. Allí, se propone la exposición de una trayectoria rectangular mediante el desplazamiento a velocidad constante de un punto en pantalla. Para este experimento, dicha prueba fue replicada y realizada por tres usuarios, estableciendo el objetivo de seguir de cerca el recorrido trazado por el punto con el puntero proyectado por la orientación de la cabeza.

Las Figuras 2 y 3 muestran la comparación de los resultados de esta trayectoria para los tres usuarios. Se representa en violeta al Usuario A, en naranja al Usuario B y en verde al Usuario C. El trazado de los puntos proyectados incrementa la intensidad con el transcurso de la exploración.

Tras la realización de estas exploraciones, se pudo plantear un análisis respecto al uso de las rotaciones cefálicas como modo de controlar un puntero en la pantalla. En primer lugar, de forma opuesta a lo planteado por [4] para el movimiento ocular, la cabeza permite la realización de desplazamientos fluídos. Sin embargo, se requiere de un período de adaptación para lograr movimientos precisos. En las Figuras 2 y 3 esto se hace notorio, principalmente, en las esquinas de la trayectoria trazada, donde debe pausarse la guiñada (*yaw*) de la cabeza y comenzarse el cabeceo (*pitch*). Además, no resulta una tarea sencilla mantener una velocidad constante en las rotaciones realizadas. Esto resulta visible en las figuras, ya que las secciones trazadas de forma punteada representan un movimiento veloz o repentino, y aquellas con mayor aglomeración de puntos se corresponden con movimientos lentos.

Pruebas de estabilidad basadas en cuadrícula. Al contemplar la interacción, por ejemplo, con el escritorio de una computadora, se hace evidente la necesidad de poder seleccionar diversos íconos para abrir programas o archivos. Alternativamente, en el caso de un teclado virtual en pantalla, debe ser posible elegir entre los distintos símbolos presentados. Para realizar tal interacción, es

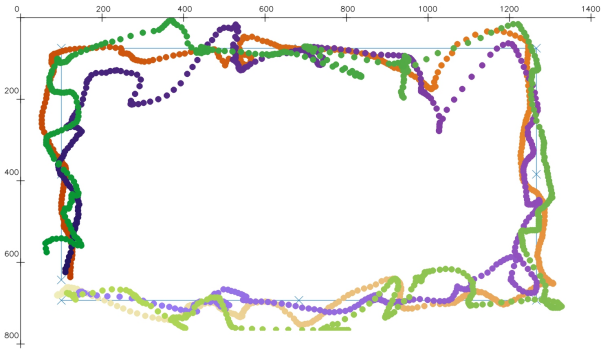


Fig. 2. Experimento de seguimiento de trayectoria con tres usuarios realizado de forma individual con *OpenFace*.

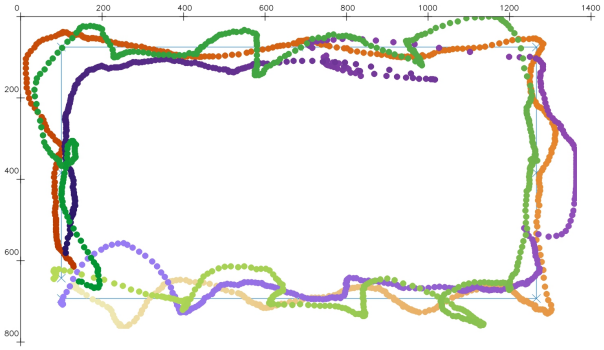


Fig. 3. Experimento de seguimiento de trayectoria con tres usuarios realizado de forma individual con *RealSense*.

requerimiento medir el nivel de precisión con el que se cuenta cuando el usuario pretende mantener su atención en un punto de manera estable.

Para llevar a cabo este estudio se desarrolló un prototipo de aplicación de estilo lúdico. En la dinámica definida, la pantalla es seccionada visualmente con una grilla de cuadrados del mismo tamaño, como muestra la Figura 4. De forma secuencial y aleatoria, son expuestas en la pantalla pistas visuales en distintas regiones de la pantalla, con el objetivo de indicar al usuario hacia dónde debe orientar su cabeza.

La exploración implementada busca estimar el tamaño mínimo que deben tener los elementos expuestos en pantalla para permitir su selección de manera cómoda. Así, la pantalla es inicialmente dividida en regiones de 96x96 pixeles, posteriormente de 48x48 y, finalmente, de 36x36 pixeles.

En el Cuadro 1 se exponen los resultados numéricos correspondientes a las pruebas realizadas por tres individuos con ambas herramientas. Los Usuarios E y D experimentaron la dinámica de interacción por primera vez inmediatamente

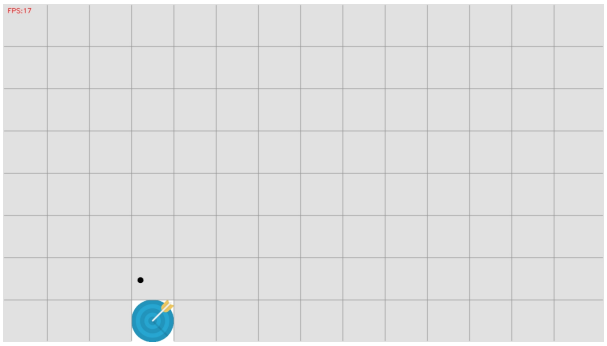


Fig. 4. Captura de pantalla de un experimento basado en cuadrícula.

antes de participar de la actividad explorativa, a diferencia del Usuario B que ya había realizado diversas pruebas con anterioridad. Puede verse en los valores expuestos que, debido a la experiencia adquirida por el Usuario B respecto al modo de funcionamiento y la respuesta ofrecida por *OpenFace* y *RealSense* a sus movimientos, logró un desempeño casi perfecto. Debido al modelo 3D adaptativo con el que cuenta *OpenFace*, los dos sujetos inexperimentados pudieron llevar a cabo la actividad sin mayores problemas. En cambio, dado el modelo 3D estático utilizado para las pruebas con *RealSense*, se pudo observar una alta susceptibilidad a cambios de iluminación y posición del individuo respecto a la cámara, provocando un desempeño pobre, de ambos usuarios, al incrementar las restricciones.

Table 1. Resultados de los experimentos realizados con tres usuarios de las pruebas basadas en regiones.

Usuario B	96x96	48x48	36x36
Tiempo Promedio con *OpenFace*	5.576s	6.697s	6.931s
Tiempo Promedio con *RealSense*	6.434s	7.063s	7.171s
Regiones no logradas con *OpenFace*	0	0	0
Regiones no logradas con *RealSense*	0	0	1
Usuario D	96x96	48x48	36x36
Tiempo Promedio con *OpenFace*	5.841s	5.879s	6.834s
Tiempo Promedio con *RealSense*	7.172s	6.679s	7.104s
Regiones no logradas con *OpenFace*	0	0	2
Regiones no logradas con *RealSense*	1	1	6
Usuario E	96x96	48x48	36x36
Tiempo Promedio con *OpenFace*	5.814s	6.119s	7.551s
Tiempo Promedio con *RealSense*	5.396s	7.437s	7.623s
Regiones no logradas con *OpenFace*	0	0	0
Regiones no logradas con *RealSense*	0	2	6

Respecto a los tiempos promedio calculados para la selección de regiones, en la mayoría de los casos se pudo observar que la estabilidad brindada por *OpenFace* permitía alcanzarla con mayor facilidad y, como se pudo prever, la disminución del tamaño del área restrictiva introdujo demoras en el tiempo requerido para seleccionar una región en todos los casos.

4 Conclusiones

Se estudió la factibilidad de la realización de una herramienta asistencial para la interacción humano-computadora basada en los movimientos de la cabeza de los usuarios. El enfoque se basó en el modelo de *camera mouse*, presentado por Nabati y Behrad [9], para la implementación de un elemento de control que no requiera contacto físico y sea soportado por hardware ampliamente disponible y de bajo costo. Para llevarlo a cabo, se analizaron y utilizaron dos herramientas gratuitas particulares: una de ellas de código abierto, y otra comercial.

Al desarrollar el estudio planteado, se utilizó un *pipeline* unificado como estructura inicial para conectar las diversas tareas implicadas en la proyección de un puntero en la pantalla, a partir de la captura de imágenes a través de una cámara web estándar.

La tasa de procesamiento de imágenes del sistema implementado es altamente dependiente de la biblioteca base empleada para el análisis facial en 2D. Al utilizar *RealSense* se llegó a procesar a 30fps, dado que la misma cuenta con optimizaciones en su implementación para procesadores Intel. Con *OpenFace* la tasa máxima alcanzada fue de 20fps; no obstante, en el tiempo de cálculo se incluye la obtención de la malla 3D de la geometría facial del usuario.

Los resultados cuantitativos y cualitativos de los experimentos realizados, si bien han sido obtenidos con una cantidad de participantes reducida, permiten hipotetizar que el método resultaría prometedor para condiciones normales de captura con cámaras de bajo costo, aunque no se equipararía con el desempeño publicitado por los dispositivos comerciales existentes.

El trabajo futuro incluye la prueba del mismo con una muestra mayor de usuarios y la adición al sistema de un mecanismo de interacción basado en acciones. Gestos como el parpadeo, guiño o movimiento de las comisuras de la boca, podrían ser detectados en forma de acciones faciales para proveer un mecanismo de activación. Es posible, también, continuar hacia una implementación con mayor robustez, contemplando la variabilidad de iluminación y efectuando adaptaciones respecto a los ángulos máximos de movimiento utilizados para la proyección del puntero.

References

1. Böhme, M., Meyer, A., Martinetz, T., Barth, E.: Remote eye tracking: State of the art and directions for future development. In: Proc. of the 2006 Conference on Communication by Gaze Interaction (COGAIN), pp. 12–17 (2006)

2. Valenti, R., Sebe, N., Gevers, T.: Combining head pose and eye location information for gaze estimation. IEEE Transactions on Image Processing, 21(2):802–815 (2012)
3. Chandra, S., Sharma, G., Malhotra, S., Jha, D., Mittal, A. P.: Eye tracking based human computer interaction: Applications and their uses. In: Man and Machine Interfacing (MAMI), 2015 International Conference on, pp. 1–5. IEEE (2015)
4. Jacob, R. J.: What you look at is what you get: eye movement based interaction techniques. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 11–18. ACM (1990)
5. Mange, A. A., Choudhari, A. V., Prasad, S.: Gaze and blinking base human machine interaction system. In: Computational Intelligence and Computing Research (ICCIC), 2015 IEEE International Conference on, pp. 1–4. IEEE (2015)
6. Chauhan, V., Morris, T.: Face and feature tracking for cursor control. In: Proceedings of the Scandinavian Conference on Image Analysis, pp. 356–362 (2001)
7. Murphy-Chutorian, E., Trivedi, M. M.: Head pose estimation in computer vision: A survey. IEEE transactions on pattern analysis and machine intelligence, 31(4):607–626 (2009)
8. Betke, M., Gips, J., Fleming, P.: The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities. IEEE Transactions on neural systems and Rehabilitation Engineering, 10(1):1–10. (2002)
9. Nabati, M., Behrad, A.: Camera mouse implementation using 3d head pose estimation by monocular video camera and 2d to 3d point and line correspondences. In Telecommunications (IST), 2010 5th International Symposium on, pp. 825–830. IEEE (2010)
10. Baltrušaitis, T., Robinson, P., Morency, L.-P.: OpenFace: an open source facial behavior analysis toolkit. En Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on, pp. 1–10. IEEE (2016)
11. Intel Corporation. Intel® RealSense™ SDK 2016 R2 Documentation. `https://software.intel.com/sites/landingpage/realsense/camera-sdk/v1.1/documentation/html/index.html?doc_devguide_introduction.html`. (Online; accedido el 12 de Marzo del 2018)
12. Cheung, Y. M., Peng, Q.: Eye gaze tracking with a web camera in a desktop environment. IEEE Transactions on Human-Machine Systems, 45(4), 419-430 (2015)
13. Bradski, G., Kaehler, A.: Learning OpenCV: Computer vision withthe OpenCV library, O'Reilly Media, Inc. (2008)
14. Nabati, M., Behrad, A.: 3D head pose estimation and camera mouse implementation using a monocular video camera. Signal, Image and Video Processing, 9(1):39–44 (2015)
15. Kalman, R. E.: A new approach to linear filtering and prediction problems. Journal of basic Engineering, 82(1):35–45 (1960)
16. Jackson, A. S., Bulat, A., Argyriou, V., Tzimiropoulos, G.: Large pose 3D face reconstruction from a single image via direct volumetric CNN regression. International Conference on Computer Vision (2017)
17. Anson, D., Brandon, C., Davis, A., Hill, M., Michalik, B., Sennett, C.: Swype vrs. conventional on-screen keyboards: Efficacy compared. In RESNA Annual Conference (2012)
18. Cazzato, D., Dominio, F., Manduchi, R., Castro, S. M.: Real-time gaze estimation via pupil center tracking. Paladyn, Journal of Behavioral Robotics, 9(1):6–18 (2018)

Simulador de tiro de tanques: NeoNahuel II

Lucas Enrique Guaycochea[1,2], Joel Kaltman[1], María Victoria Galán[1], and Javier Edgardo Luiso[1]

[1] Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF)
[2] Universidad Nacional de General Sarmiento (UNGS)

Resumen El presente artículo describe el trabajo realizado en el desarrollo de un simulador de entrenamiento para el Ejército Argentino. El simulador NeoNahuel II permite el entrenamiento en el uso del Tanque Argentino Mediano (TAM). Se muestran los análisis y las soluciones a los distintos desafíos tecnológicos que plantearon los requerimientos del sistema solicitado. Se integran conocimientos en computación gráfica, simulación, inteligencia artificial y sistemas distribuidos. Se presentan decisiones de arquitectura y despliegue, y la implementación de algoritmos de inteligencia artificial de los agentes autónomos del simulador. Este simulador constituye una segunda versión desarrollada totalmente in-house por el equipo de desarrollo, a diferencia de su primera versión que utilizaba el juego comercial Steel Beasts [1].

Keywords: simuladores de entrenamiento, computación gráfica, inteligencia artificial, sistemas distribuidos.

1. Introducción

El uso de simuladores para el entrenamiento de procedimientos y operatoria de equipos o máquinas se ha ido expandiendo ampliamente. El concepto de simulación ha ido evolucionando en el tiempo gracias al avance tecnológico logrando simuladores que, a partir de una representación virtual, logran una inmersión tal que el usuario siente estar operando el equipo real en una situación real. La simulación virtual involucra a gente real operando sistemas simulados [2]. Los simuladores logran un beneficio tanto en el ahorro de recursos económicos, como en permitir ejercitar y repetir situaciones, que incluso serían riesgosas de replicar en la realidad. La finalidad última de los simuladores es desarrollar competencias en sus educandos, donde una competencia se define como: "una combinación observable y medible de conocimiento, destrezas, habilidades y atributos personales que contribuyen a la mejora del desempeño del empleado y en el éxito organizacional" [3].

El área de Defensa de varios países ha impulsado el desarrollo en el campo de la simulación al requerir simuladores de entrenamiento para sus fuerzas armadas, buscando el crecimiento de las competencias del personal militar de una manera vasta, económica y medible. La representación virtual de escenarios

logrando imágenes y animaciones (secuencia de imágenes) fotorrealistas contribuyen ampliamente a la inmersión del usuario en la situación de entrenamiento. El avance tecnológico en la computación, más precisamente en la computación gráfica, tanto en el hardware como en los algoritmos de renderizado[3] en tiempo real, viabilizan estos resultados. Otro aspecto fundamental en la fidelidad del entrenamiento es que el usuario interactúe con los mismos elementos físicos que interactuará en la realidad. Para esto, muchos simuladores desarrollan además soluciones de hardware que emulan los dispositivos físicos reales (Fig. 1). Por último, el entrenamiento puede involucrar la interacción del usuario con otros agentes inteligentes en la situación simulada, que pueden ser otros usuarios o agentes simulados. La primera alternativa suele involucrar el desarrollo de un simulador distribuido, donde diferentes usuarios utilizan distintos sistemas interconectados para trabajar sobre un mismo escenario. La segunda requiere el desarrollo de comportamientos realistas para los agentes virtuales que son simulados, este tipo de simulación se la llama constructiva [2] e involucra la implementación de algoritmos de inteligencia artificial.



Figura 1: Izq: Interior de torre del tanque TAM. Der: Cabina del simulador.

El simulador de tiro del Tanque Argentino Mediano, NeoNahuel II, solicitado por el Ejército Argentino, es un producto encuadrado en todo lo descripto en el párrafo anterior. Sus requerimientos han planteado la necesidad de soluciones dentro de las áreas descriptas que fueron totalmente desarrolladas por este equipo de trabajo. Este artículo presenta a continuación una descripción del simulador requerido y llevado a cabo a partir de las soluciones implementadas. Las mismas surgieron luego de un análisis de las necesidades y el estudio de las alternativas vigentes en el cuerpo de conocimientos de las ciencias de la computación y la ingeniería de software. De esta manera primero se describirá la solución a nivel arquitectura y despliegue y luego, a nivel algorítmico y de modelado, se detalla el comportamiento o inteligencia de los enemigos.

[3] proceso por el cual a partir de un modelo virtual de una escena 3D se genera una imagen desde un punto de vista dado

2. Descripción del simulador

El Tanque Argentino Mediano (TAM) es el principal tanque de combate utilizado por el Ejército Argentino. En el ámbito de combate terrestre, el significado que tiene el manejo de estas armas es de gran importancia por su capacidad de influir decisivamente en el desarrollo de las acciones bélicas. La efectividad obtenida no depende de otra cosa más que de la aptitud e idoneidad de su tripulación ante cualquier situación que pueda presentarse. Dentro de este marco, adquiere singular importancia la variedad de entrenamientos que puedan proporcionarse al personal a cargo de su uso, permitiendo un aprendizaje progresivo.

El simulador NeoNahuel II, aquí presentado, es un sistema distribuido que consta de tres programas o aplicaciones. La aplicación Instructor es utilizada para la elección de los diferentes ejercicios, escenarios y cabinas que participarán del mismo. La aplicación Cabina se ejecuta en una cabina física que simula el interior del TAM con los instrumentos de la torre del tanque. Los mandos y controles de la cabina están conectados a una PC que además permite a cada tripulante visualizar el terreno desde su punto de vista mediante los respectivos aparatos de puntería, visores del Jefe de Tanque y Apuntador, respectivamente (Fig. 2). Por último, la aplicación Editor, que permite la creación de ejercicios de forma personalizada para luego poder ser utilizados dentro de la simulación.

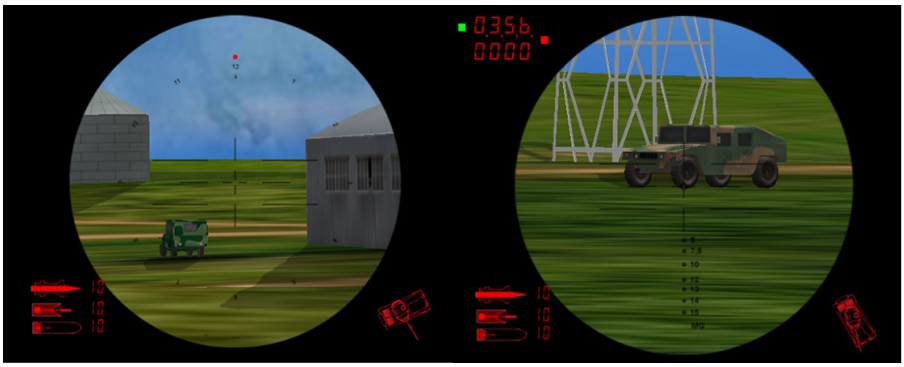


Figura 2: Capturas del sistema. Visualizaciones desde los aparatos de punterías del Jefe de Tanque (izq.) y del Apuntador (der.)

Finalmente, el simulador es una herramienta flexible y económica que complementa el adiestramiento en el terreno del Jefe de Tanque y Apuntador del TAM. Se recrean situaciones de combate que permiten a sus usuarios realizar apreciaciones y mediciones, para resolver en tiempo real situaciones ante un enemigo virtual al que se le podrá disparar y este podrá responder al fuego. Además, permite el entrenamiento a partir situaciones virtuales que combinan todo tipo de dificultades que pueden presentarse en un campo de batalla, desde

inconvenientes técnicos o mecánicos del vehículo hasta la interacción con distintos tipos de enemigos de diferentes embergaduras. Este simulador permite realizar el entrenamiento en forma individual de cada tripulante o de la tripulación completa de un vehículo. También permite realizar ejercicios de duelo de tanques con dos cabinas y adiestrar a una sección completa (3 tanques) en el marco de un ejercicio conjunto.

De esta breve descripción del simulador y los objetivos de instrucción perseguidos, surgen las características y requerimientos técnicos que fueron necesarios implementar para obtener el sistema desarrollado. A continuación se exponen las soluciones desarrolladas que implicaron un mayor trabajo de investigación, estudio, ingeniería e implementación.

3. Despliegue y Arquitectura del simulador

El simulador requiere la participación de varios usuarios con distintos roles de manera simultánea. Esto implica diseñar la solución para ejecutarse en distintas PCs. Se decidió entonces la realización de un despliegue tal que el instructor opere una PC y cada una de las cabinas que representarán un tanque TAM en la simulación utilizarán también una PC. Para cada cabina se desarrollaron los paneles de hardware y la electrónica necesaria para la comunicación de estos con la PC a través de un puerto serie. Se desarrollaron los mandos de ambos tripulantes que luego son vistos por la aplicación como unos joystick estándar. Por último, se conectaron dos monitores que son ubicados en la cabina, tras unas miras con el adecuado lente óptico, donde se muestra la escena virtual 3D en la cual se lleva a cabo el ejercicio.

Las distintas PCs, y las aplicaciones que ejecutan, necesitan estar interconectadas en una red de área local (LAN) para poder comunicarse entre sí. Es una característica del simulador que más de una cabina, y también el instructor, puedan interactuar en una misma realidad simulada, de manera que el modelo de la simulación y la apreciación de la misma a través de su visualización evolucione en forma idéntica para todos los participantes. Para resolver este requerimiento el simulador se construyó sobre la Plataforma de Desarrollos 3D Distribuidos (P3D) desarrollado de manera concurrente por el mismo equipo de trabajo. La P3D es equivalente a un motor de juegos. Brinda particularmente la funcionalidad de un motor gráfico para la generación de las imágenes en tiempo real de la escena virtual e implementa un middleware distribuido que abstrae los aspectos de sincronización de la escena distribuida [5] [6]. En [7] se describen las características de la plataforma P3D y el plan de trabajo actual sobre la misma.

Finalmente, las aplicaciones desarrolladas para este simulador, utilizan como una capa inferior de servicios a la P3D y además crean un protocolo de comunicación entre ellas, implementado mediante TCP/IP. Este protocolo permite administrar la carga de recursos, inicializar y finalizar una escena y un ejercicio, y compartir datos y/o eventos de interés que forman parte de los reportes de cada ejercicio (Fig. 3).

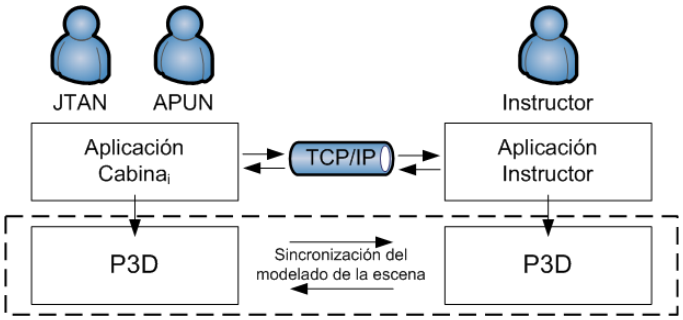


Figura 3: Despliegue y arquitectura del simulador.

4. Comportamiento de Enemigos

Para obtener un simulador con un alto grado de realismo no sólo es importante la calidad de las imágenes fotorrealistas generadas por la plataforma gráfica 3D, sino también la simulación de los comportamientos de los enemigos. La manera en la cual un enemigo se desplaza por el terreno, reacciona ante el contacto con el tanque TAM y procede a atacar, son aspectos que deben ser resueltos y presentarse tan reales como sea posible, siendo esto muy significativo en la experiencia de usuario final. Los algoritmos que son utilizados para resolver estos aspectos pertenecen al área de la inteligencia artificial. En este campo, a las entidades que van a tomar sus propias decisiones se las conoce como agentes autónomos y, en lo que refiere a juegos, como personaje no-jugador o NPC por sus siglas en inglés, *non-playable character*.

El simulador precisa que los vehículos enemigos tomen determinaciones como si estuvieran comandados por humanos. En los ejercicios militares, el terreno donde se desarrolla adquiere un rol fundamental en la toma de decisiones [8]. A nivel requerimientos, el comportamiento esperado de los enemigos es descripto en términos de: "el enemigo se pone a cubierto", "las tripulaciones del TAM deben reaccionar ante una emboscada" o "un vehículo de menor potencia de fuego debe huir a una zona de escondite", entre otros. Además, como se discute en Smelik et al. [8] y en Wells [9], las características de un terreno geo-típico[4] no sólo son esenciales a un nivel táctico y estratégico en un ejercicio, sino que también son fundamentales para los jugadores como claves para orientarse y recorrer el escenario.

La capacidad de un agente autónomo de utilizar las características del terreno como parte de la toma de decisiones o, en definitiva, la relación entre la inteligencia artificial y el terreno, se lo denomina *"razonamiento espacial o del terreno"*[5] [10]. Aparte del modelo o mallado 3d y la textura que son utilizados

[4] En contraposición a un terreno aleatorio cualquiera, es un terreno con características geográficas similares a una zona conocida del planeta

[5] en inglés, spatial/terrain reasoning.

por la plataforma gráfica para la representación visual de la topografía, el simulador precisa disponer de los datos de la superficie y el escenario en un formato conocido y útil a la hora de obtener la información necesaria para los algoritmos de inteligencia artificial. Esto es entonces representado por 3 capas de datos superpuestas de manera tal que puedan ser accedidos eficientemente. Cada capa es una matriz bidimensional que discretiza los valores de la superficie del terreno. La primera capa contiene el valor de elevación, la segunda un valor que identifica el tipo de suelo (pasto, barro, camino, agua, no transitable) y por último la tercera capa que sirve para identificar si esa zona posee algún objeto natural o edificado y la altura del mismo por sobre el nivel de la superficie. Estos datos serán requeridos frecuentemente por los algoritmos descritos a continuación.

4.1. Path Finding

Path finding consiste en encontrar el camino óptimo para llegar a un punto de destino desde una ubicación de origen. A partir de la representación discreta del terreno, se utiliza la variante del algoritmo de camino mínimo de Dijkstra llamado A* *(A-estrella)*, popularmente usado para resolver este problema [11]. Cada celda representa un nodo que esta conectado a sus ocho celdas vecinas. El peso asignado al paso por un nodo esta directamente asociado al valor de tipo de suelo que está en la segunda capa de la representación del terreno; tal que, por ejemplo, el asfalto es asociado con un peso bajo, el barro con uno más alto, y se penaliza las zonas "no transitable" con un peso infinito. Existe un preprocesamiento donde se asigna un valor de "no transitable" en la capa 2 coincidiendo con la descripción de los objetos sobre el terreno de la capa 3. Además, también se tienen en cuenta para la ponderación la pendiente del terreno (para lo cual se consulta la primera capa) y los cambios de dirección.

4.2. Line of sight

Por otro lado se implementó el algoritmo conocido como "línea de vista" (en inglés, *line-of-sight*) que resuelve si hay una visión directa entre dos puntos sin ningún objeto o irregularidad del terreno que se interponga entre ellos. Este algoritmo es utilizado para identificar si en un momento dado un enemigo tiene visualización directa sobre el o los tanques TAM y para encontrar un sector del terreno donde un enemigo puede ocultarse respecto del jugador, esto es, estar a cubierto por alguna elevación o depresión del terreno o detrás de algún objeto.

Para realizar este cómputo el algoritmo hace uso de la información de la primera y tercera capa de los datos del terreno, es decir, la elevación y la existencia de objetos sobre la superficie. Cabe destacar que sólo los objetos de un volumen y opacidad tal que pueden ocultar un vehículo son considerados en esta capa. Por ejemplo, edificaciones como hangares, silos, arboledas son incluidos, y se excluyen objetos tales como rocas, alambrados, arboles poco frondosos y arbustos, entre otros. Similar a lo propuesto en [12], se identifican los dos participantes de esta consulta, el vehículo "observador" y aquel que se quiere definir si es visible por el primero, se toman ambas posiciones sobre la superficie del terreno (junto

con la altura del aparato de visión del primero y la altura media del segundo) y se traza un segmento 3D entre ellos (Fig. 4). Luego, aplicando el algoritmo de Bresenham [13], se discretiza el segmento y se calcula si la altura en cada paso es menor que la suma de las alturas de las capas 1 y capa 3; de ser así, hay un obstáculo entre ellos, y se concluye que la primera posición no puede visualizar a la segunda.

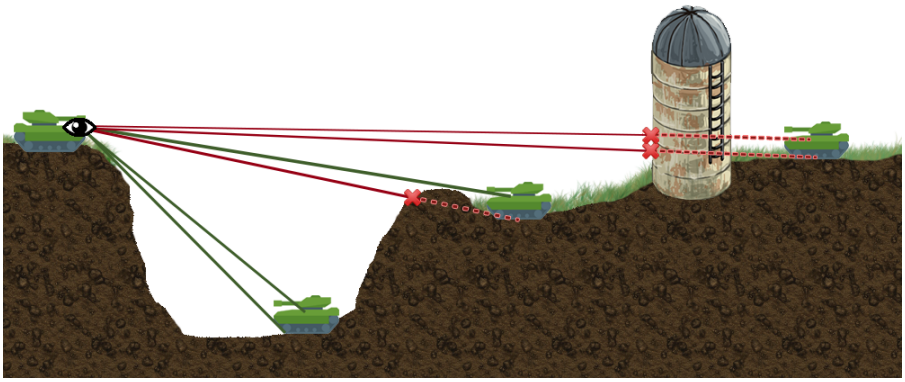


Figura 4: Esquema de las situaciones que resuelve el algoritmo de line-of-sight. Se diferencian las visualizaciones total, parcial y nula de enemigos (de izq. a der.).

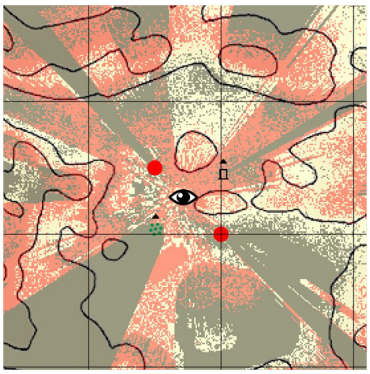
Por último, además de ponerse en práctica el algoritmo durante el transcurso de un ejercicio, se utiliza dentro del editor de situaciones, en donde se permite generar un mapa de visión a partir de una ubicación particular, diferenciando entre posiciones visibles, no visibles y de visibilidad parcial. Las mismas se muestran sombreadas en rosa, en un color claro y oscurecidas, respectivamente. (Fig. 5a).

4.3. Búsqueda de una posición de ocultamiento

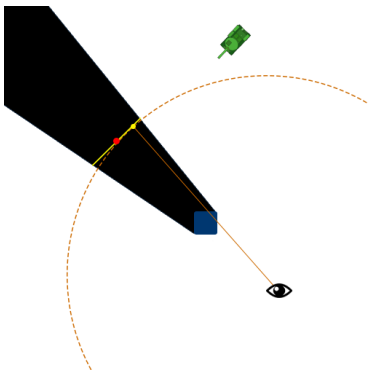
Utilizando el cómputo de la "linea de vista" se desarrolló un algoritmo de búsqueda de posiciones que permite encontrar el lugar óptimo de cobertura en el terreno. Este algoritmo está conformado por una serie de etapas en las cuales se considera si una ubicación y sus alrededores son visibles. Utilizando como base la discretización antes mencionada del terreno, se buscan las posiciones candidatas a ser elegidas como escondite dentro de una circunferencia que tiene como centro la posición del NPC. En caso de no encontrar una opción viable en su interior, se irá ampliando su radio sucesivamente hasta encontrarla. Esto tiene como objetivo evitar excesivas cantidades de procesamiento innecesario en escenarios muy grandes. Entonces se recorren todas las posibles ubicaciones dentro de la

circunferencia, descartando aquellos puntos que estén descubiertos según su linea de visión. Una vez separados los lugares ocultos de los expuestos, se valorizan de acuerdo a dos criterios: distancia al NPC (a menor distancia mejor valor) y visibilidad de posiciones vecinas (si gran cantidad de lugares adyacentes son ocultos, también mejorará su valor). De todos los posibles candidatos se elije el de mejor valoración resultante.

Luego de la implementación del algoritmo descripto, en sucesivas pruebas se encontró que el agente se ubicaba en posiciones donde podía verse parcialmente desde el jugador. Se analizó que este resultado imperfecto es producto de la discretización, ya que el agente, al elegir el punto más cercano de una determinada zona, se ubicaba en un borde de esta y por lo tanto, en ocasiones, no cumplía su objetivo. Es por esto entonces que se debe considerar la totalidad de la zona de ocultamiento que se elige, con el objetivo de acomodarse lo mejor posible dentro de la misma. Para lo cual se plantea una circunferencia con centro en la posición del observador y su radio será la distancia al punto elegido, dejando a dicho punto en el borde de la circunferencia. Se traza una linea tangencial que pase por ese punto y atraviese a todas las posiciones ocultas que rodeen a la elegida. El trazado será detenido en cuanto encuentre puntos visibles para el observador. Finalmente se elige el punto mas próximo al centro de la linea, modificando el punto elegido en caso de ser necesario para "pulir" los resultados imperfectos (Fig. 5b).



(a) Mapa de visión.



(b) Resolución para los casos de ocultamiento parcial.

Figura 5: Resolución del ocultamiento

4.4. Máquina de Estados Finita

Finalmente, queda la tarea de implementar el comportamiento en si mismo, es decir, las distintas actitudes en la que se encuentra un enemigo y la toma de

decisiones que deben realizar. Se definieron 5 comportamientos distintos posibles para estos agentes, estos son, ataque, defensa, huida, marcha pasiva y marcha activa. Por ejemplo, el comportamiento de ataque se describió como: "Cuando el agente *ve*[6] a un tanque TAM decide tomar como objetivo de sus ataques el tanque TAM más próximo a su posición. Tras un tiempo de retardo de inicio de fuego, se fija si se encuentra en rango para realizar un disparo. Si no lo está, decide aproximarse en dirección a su objetivo una determinada distancia y vuelve a fijarse si está en rango. Si se encuentra en rango, abre fuego dos veces y si vence a su objetivo retorna al estado pasivo inicial, sino, decide aproximarse hacia su objetivo para volver a abrir fuego".

Existen distintas alternativas para modelar e implementar un comportamiento de este tipo en la industria, las más conocidas, simples y utilizadas, son a través de un *árbol de decisiones*, *sistemas basados en reglas* y *máquinas de estados finitas* [14]. La primera corresponde a una implementación hardcodeada de cláusulas *if-else*. La segunda consiste en entender el comportamiento como un conjunto de reglas junto con una memoria de datos. Las reglas son representadas por uno o más antecedentes y uno o más consecuentes. Un antecedente es una condición y un consecuente incluye acciones a ejecutar. Existen trabajos que implementan un motor de reglas genérico [15] que permiten definir las reglas a un nivel de abstracción superior que el lenguaje de programación, en un archivo de configuración de texto (esta opción se eligió para la implementación del controlador de cada ejercicio que maneja la instalación de nuevos enemigos, el tiempo de exposición de cada uno y las condiciones de finalización de un ejercicio, entre otras). Por último, un modelo basado en una máquina de estados finitos consiste en identificar los distintos estados en los que se puede encontrar un agente, las acciones que debe realizar en cada uno de ellos y las acciones, condiciones o eventos externos que implicarán un cambio de estado [16].

En este trabajo se eligió utilizar una máquina de estados finita para modelar cada uno de los comportamientos posibles de los enemigos. Esta elección se debió a que presentan una técnica de modelado intuitiva y simple, y requieren una implementación que es sencilla y eficiente.

5. Conclusión

El simulador NeoNahuel II se encuentra desarrollado y está siendo utilizado por el líder de usuario o asesor técnico y sus colegas del arma de caballería del Ejército Argentino en su etapa final de depuración y validación de su funcionamiento. El desarrollo de un producto de estas dimensiones, tras un arduo trabajo de relevamiento y análisis, ha requerido la integración de soluciones a diversos desafíos tanto de hardware como de software, y en distintas disciplinas como computación gráfica, simulación, inteligencia artificial y sistemas distribuidos.

Finalmente se obtuvo un simulador que fue realizado totalmente por este equipo de trabajo, adquiriendo el conocimiento, y logrando un producto com-

[6] Se define que un enemigo *ve a un tanque TAM*, cuando el mismo entra en lo visualizado por alguno de los tripulantes del TAM.

pletamente a medida de lo solicitado por los clientes, pensado desde el inicio para que acompañe el entrenamiento y la formación de los miembros de la caballería, incluso desde antes de su primer contacto con una unidad real del Tanque Argentino Mediano.

Referencias

1. eSimGames: Steel Beasts. 2009, http://www.steelbeasts.com.
2. *Department of Defense Modeling and Simulation (M&S) Glossary*, July 1, 2013. Available at http://www.msco.mil/MSGlossary.html
3. University of Nebraska–Lincoln: The Definition of Competencies and Their Application at NU. University of Nebraska–Lincoln, http://hr.unl.edu/compensation/nuvalues/corecompetencies.shtml, accedido Abril 2017.
4. Ejército Argentino: Tiro con Tanque VC TAM. RFP-02-02, Ejército Argentino, República Argentina, 1995.
5. Guaycochea, L., Luiso, J. y Del Rio Garcia, F.: Middleware P2P para la Sincronización de Eventos Discretos en una Simulación Distribuida de Sistemas que evolucionan en el tiempo. XVIII Congreso Argentino de Ciencias de la Computación, CACIC 2012. Bahía Blanca, Argentina, 2012.
6. Luiso, J., Guaycochea, L. y Abbate, H.: Simuladores de entrenamiento distribuidos: Plataforma de desarrollo para ocultar los aspectos de la distribución. 1er Workshop Argentino sobre VideoJuegos, WAVI 2010, 2010.
7. Guaycochea, L., Luiso, J., Galán, María V. y Abbate, H.: Plataforma de desarrollos 3D Distribuidos (P3D). X Simposio de Informática en el Estado (SIE 2016) - XLV Jornadas Argentinas de Informática e Investigación Operativa (45º JAIIO). Tres de Febrero, Argentina, 2016.
8. Smelik, R.M., Tutenel, T., de Kraker, K.J. y Bidarra, R.: Declarative Terrain Modeling for Military Training Games. International Journal of Computer Games Technology, Hindawi Publishing Corporation, Volume 2010.
9. Wells, W.D.: Generating enhanced natural environments and terrain for interactive combat simulations (GENETICS). en Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '05), pp. 187-191, ACM Press, New York, NY, USA, 2005.
10. van der Sterren, W.: Terrain Reasoning for 3d Action Games. En Game Programming Gems 2, Ed. Cengage Learning, 2001.
11. Cui, X. y Shi, H.: A*-based Pathfinding in Modern Computer Games. IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.1, Enero 2011.
12. Seixas, R.B., Mediano, M., Gattass, M.: Efficient Line-of-Sight Algorithms for Real Terrain Data. III Operational Research Symposium and IV Logistic Navy Symposium - SPOLM'99, 1999.
13. Bresenham, J.E.: Algorithm for computer control of a digital plotter. IBM Systems Journal (Volume: 4, Issue: 1), 1965.
14. Millington, I. y Funge, J.: Artificial Intelligence for Games. 2da edición, Ed. Morgan Kaufmann Publishers, 2009.
15. Wright, I.P. y Marshall, J.A.R.: RC++: a rule-based language for game AI. First International Conference on Intelligent Games and Simulation (GAME-ON), 2000.
16. Buckland, M.: Programming Game AI by Example. Ed. Wordware, 2004

Reality Computing: An end-to-end process for Herpetological Heritage

Graciela Rodríguez, Nicolás Jofré, Yoselie Alvarado,
Jacqueline Fernandez, and Roberto Guerrero

Laboratorio de Computación Gráfica (LCG)
Universidad Nacional de San Luis,
Ejército de los Andes 950
Tel: 02664 420823, San Luis, Argentina
{npasinetti,gbrodriguez,ymalvarado,jmfer,rag}@unsl.edu.ar

Abstract. Documentation of institutional biological collections are essential for scientific studies and conservation of the biodiversity of a region. In particular, preserved specimens require the development of a short- and long-term plan to prevent damage.
In this context, the 3D digitisation of this type of documentation provides innovative mechanisms to safeguard the valuable information provided by the collections and at the same time prevent any possible loss of information. At the moment, the potential of laser scanning in model reconstruction is well-known, but developed works using this method for 3D construction reveal a lack of reliable, precise and flexible solutions. Furthermore, visualisation of results is often very useless and does not go beyond web-based applications.
This work presents an analysis of 3D modelling using two digitisation techniques: laser scanning and photogrammetry; combined with real time VR and AR visualizations and 3D printing.

Keywords: Reality Computing, Digital Photogrammetry, 3D Scanning, Virtual Reality (VR), Augmented Reality(AR), 3D Printing.

1 Introduction

Generally, digitisation has been identified as one of the major trends changing society and business in the near and long term future, breaking down industry barriers and creating opportunities for new applications [1, 2].

Particularly, three-dimensional digital models have become relevant in many applications such as inspection, navigation, object identification, visualisation and animation [3–5]. It can be said that the most important use it has been given in the last years is cultural heritage digital archiving thorough different motivations: documentation in case of loss or damage, virtual tourism and museum, educational resources, interaction without risk of damage, and so forth [6].

In this context, nowadays there are institutional biological collections which are specialised repositories of strong scientific matrix, being essential for scientific studies on the biodiversity of the region and its conservation [7, 8]. These collections are very useful for students, professors, researchers and other professionals, turning them into centres of scientific and social studies. Consequently, these collections are usually digitised. Nevertheless, similarity among digital collections with real collections will depend essentially on the used technique [9, 10].

At present, in computer graphics, capturing geometric models is a result of the use of *laser triangulation* and/or *image processing*[11, 12]. Particularly, 3D reconstruction from images has undergone a revolution in the last few years [13, 14, 12]. Computer vision techniques use photographs from data set collection to rapidly build detailed 3D models. The results are promising because the obtained models are beginning to challenge the precision of laser-based reconstructions [15].

Furthermore, the fast popularisation of sophisticated human-computer interaction devices has brought unrivalled convenience and entertainment experience in human life [16]. Thanks to technologies such as *Virtual Reality* (VR) and *Augmented Reality* (AR), the user has the ability to manipulate the perception to such an extent that can enter to another type of reality [17]. VR allows the creation of interaction's environments that facilitate new contexts of exchange and communication of information. Currently, VR applications consist in user immersion into a computer-generated environment, resulting in a natural idea to improve the impression of living in a simulated reality. On the other hand, an AR system can overlay computer-generated contents on views of the physical scene, augmenting a users perception and cognition of the world. Development of AR technology with precise information augmentation in real-time is a foreseeable reality that can be used in almost any domain. Over the past decade, AR has undergone a transition from desktop to mobile computing [18].

3D digitisation have promoted not only the mentioned technologies but also 3D printing technology. 3D printing technology was a change in the industrial/manufacturing field and in the way of human life. Basically, the printing process is defined as the process of joining materials to make objects from 3D model data. In recent times, many educational research have begun to use this technology like a learning tool [19].

This work explains the set of processes developed for the preservation and use of a representative set of type herpetological specimens through digital techniques. The specimens were processed into a three-dimensional textured model that can be used in VR and AR environments and 3D printing, allowing for multiple educational and research purposes.

2 Context

This effort aims to preserve and enrich the collections of the *Unidad de Herpetología* at the *Universidad Nacional de San Luis*, due to the fact that it is

a space that concentrates specimens of the type series and typical sites, bibliographic material of difficult achievement, as well as a vast photographic record, which represents and protects a large part of the herpetological heritage of the Argentinian Central and Patagonian Regions.

It is common for samples of preserved specimens to suffer deterioration that can sometimes affect their functionality, the most common wear and tear is due to the passage of time and the handling of users.

With the aim of protecting and conserving the heritage of the *Unidad de Herpetología*, this work consisted of generating digital replicas of a representative set of type specimens, thus achieving the corresponding 3D models of the real samples. These 3D models are intended to be used for the creation of a structured and organised digital collection of the preserved specimens, developed according to a predefined conceptual scheme, in order to expand the didactic-scientific potential of these resources. In addition, digitisation will improve the professional and educational transfer capacity of the *Unidad de Herpetología*, given the valuable and specific educational contribution that the collections can offer [20].

The virtual interaction with the documented specimens on the computer (e.g. deliberate cross sections; changing perspectives) allows for a more detailed understanding from it. These computer-based 3D representations of reality also offer various possibilities for reproducing models for publication and presentation purposes. A comprehensive 3D process chain can also result in animations, games and elaborate computer applications [21, 22].

3 Digitising

Today a large number of remote sensing sensors and data are available for mapping purposes and digital recording of visual Cultural Heritage. Generally optical recording sensors are divided in passive and active systems. Passive sensors deliver image data which are then processed with some mathematical formulations to infer 3D information from the 2D image measurements. On the other hand, active sensors can provide data directly from 3D information or ranges [23].

This work was developed by researchers from the *Laboratorio de Computación Gráfica* at the *Universidad de San Luis*, which has the necessary equipment to work with both sensor systems.

At first, to obtain the digital models corresponding to the herpetological specimens, the process chain was started using optical recording sensors based on active systems because this method was recommended for beginners in digitisation.

3.1 Active sensors systems: Laser scanning

These systems use laser scanning technology that analyses and captures the geometry and colors of physical objects to transform them into digital 3D models [24]. In this instance, the *NextEngine* 3D laser scanner was used and the following steps were established:

1. **Object positioning:** In this case, *ScanStudio*, the *NextEngine* scanner software, was employed. The scanner has a rotating platform dedicated to holding the objects to be scanned. Different orientations of the specimens were analyzed, obtaining the best result by vertically aligning the spine.
2. **Object and Turntable Setup:** Before starting the scanning process, the area to be scanned is selected to limit the amount of foreign information registered by the scanner. Then, it is rotated 360 degrees to the object, stopping momentarily to analyze the surface of the object using triangulation techniques by infrared light. As a result of this process, the quality of a model is directly related to the amount of analysis (stops; 32 in this case) carried on the object.
3. **Model and Texture Scanning:** Each surface analysis is accompanied by an object's snapshot, which is associated with the digitized model. Different resolutions were considered and the highest resolution of those offered by the tool was selected (2.048×1.536 pixels).
4. **Model Verification and Correction:** Once the digitized model was obtained, it is necessary to check it. Digitisation failure are displayed by incomplete or incorrect parts, and manual corrections had to be made. For this issue, the *Blender 3D* modelling tool was used. In reptile specimens the areas most prone to problems are claws and tail.

The 3D models resulting from this digitisation methodology were analysed by the *Unidad de Herpetología* people without obtaining their approval. The reason was the poor detail obtained from both the 3D model and the texture, specifically, the appreciation of each specimen scale at the lumbar and ventral area, the reproductive region and the head area. Therefore, a more professional method of obtaining realistic digital models was needed.

3.2 Passive sensors systems: Photogrammetry

There is a technique based on passive sensor systems known as *Digital Photogrammetry*. It is based on image data processing and can deliver accurate, metric and detailed 3D information at any scale of application with estimates of accuracy and reliability of unknown parameters from a set of images [25].

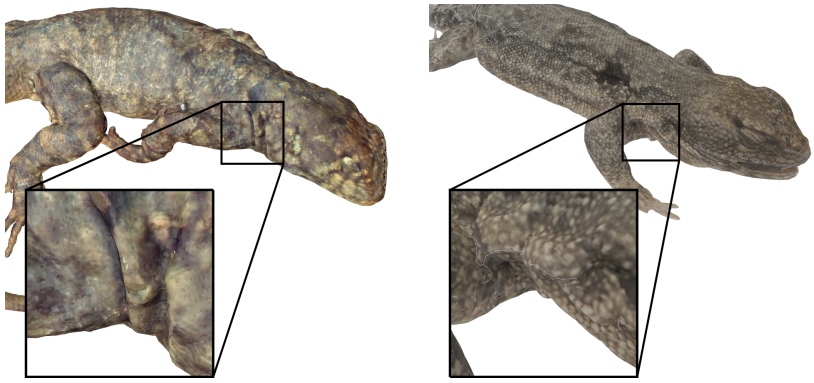
Objects well suited for this technique should have amorphous geometries, structured surfaces, many edges, many corresponding image points and an inhomogeneous colouring, characteristics of typical objects in herpetology and preservation of specimens.

Currently, there are tools dedicated to digital photogrammetry such as *ReCap*, *Agisoft* and *Regard 3D*, which require high performance computers for the large amount of calculations that this methodology requires. According with these tools, the following stages of the digitization process can be set:

1. **Data acquisition:** Ideally, images of the specimen should cover each region in order to achieve a complete 3D reconstruction. In this work, a *Nikon D3400* camera with a resolution of 3.872×2.592 pixels was used to capture

specimen images (over 100 captures per specimen). Climate is a key factor, because it directly influences the light received by the objects and the texture of the digitised model. Therefore, a cloudy weather is recommended where light rays are distributed evenly. For this reason, the main setting on the camera is light sensitivity (ISO).

2. **Data processing:** The processing of image data consist of image matching algorithms which identify homologous points between images by correlating grey level variations and contrast within a template window. This step achieves the alignment of the photos and the generation of a cloud of dense points representing the outer skin of the object. In this stage it must be adjusted, among other variables, the number of points that will be searched in other captures (set 60.000 points) and the number of points of links between different captures (set 4.000 points).
3. **3D model generation:** This stage reconstructs a 3D polygonal mesh representing the object surface based on the dense or sparse point cloud. Generally there are two algorithmic methods available that can be applied to 3D mesh generation: for planar type surfaces or for any kind of object. Due to the specimens are complex objects and it is desired to obtain high quality meshes, it was specified that the final models are conformed by at least 50.000 polygons.
4. **Texturing:** After geometry is reconstructed, it can be textured. Different texturing modes determines how the object texture will be packed in the texture atlas. Because it is important to record small details, such as specimen scales, it was necessary to obtain high resolution texture images (4.096×4.096 pixels).
5. **Model Verification and Correction:** In this case, it was necessary to reconstruct some small parts of the specimens (claws) using *Blender 3D*.



(a) Laser scanning (b) Digital Photogrammetry

Fig. 1: Texture detail.

Figure 1 shows the texture returned by the digitization methodologies tested. Once again, the new resulting 3D models were analysed by the herpetology unit people and in this time, with a positive feedback. The main reason for the improvement of the 3D models is the possibility of selecting certain areas of the specimens to maximize the detail of the resulting model.

On the other way, the *Unidad de Herpetología* people expressed the need to show each 3D specimen model to users for analysis and study.

4 3D models implementations approaches

Once an object has been digitised, in this case a specimen, the computer uses are almost endless. It can be displayed as a two-dimensional image through a process called *3D rendering* or used in a computer simulation. Accordingly, display methods are classified into those that perform real-time processing and those that do not.

Consequently, since the 3D modelling process was performed from real objects which is usually called *Reality Capture*, this work use this in combination with 3D printing to create an end-to-end process known as *Reality Computing*. In this process the virtual model allows the physical recreation of the real object using 3D printing devices [26].

In accordance with the above, the following subsections describe the developments in both visualisation and 3D printing in which the 3D models obtained were included.

4.1 Visualization

The *Laboratorio de Computación Gráfica* works with different real-time visualization paradigms such as virtual reality and augmented reality, among others. These paradigms were proposed as a solution for the appreciation of digitized specimens to users.

Immersive Virtual Reality: It is known Virtual-reality-enhanced interactive learning environments are increasingly common. VR brings together a mixture of virtual and real-life scenarios for a wide range of potential possibilities in teaching and learning. Particularly, the *Laboratorio de Computación Gráfica* of *Universidad Nacional de San Luis* owns a Cave-like multi-VRmedia System which comprises the hardware and software necessary to gather the information obtained during the interaction between the user and the system: via a motion sensing device, sound system, microphone, screen/projection surfaces and projectors, among others.

This system provides the necessary structure for attributes definition, rendering and collaborative multi-visualizations, as well as the needed interactive resources. Therefore, as an innovative and interesting way where users can interact and observe the different digitised specimens, a scenario, set up with ambient

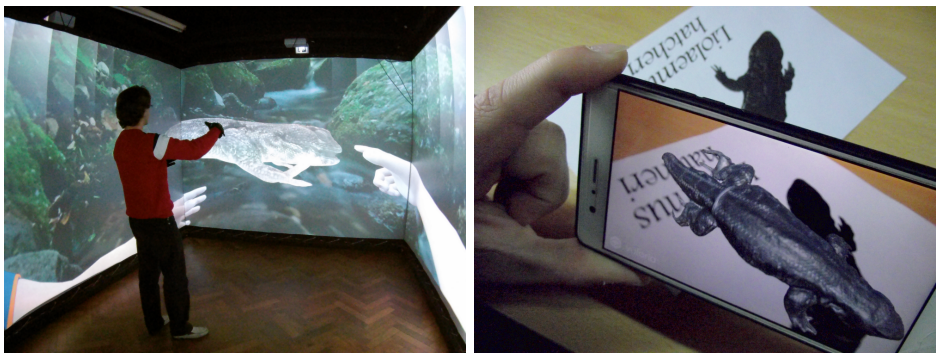
sound, inanimated objects (trees, plants, ground) and animated objects (specimens) was modelled and rendered. The specimens on stage are dynamic objects which are controlled by an avatar (representing the user). The user can control (e.g. rotate) the virtual replica of the specimen by using a keyboard, mouse and motion detection devices. In this way, the user will be able to observe it in more detail and from various angles (See Fig. 2a).

Augmented Reality: Many students and general audience of San Luis province don't know about biodiversity of the region and its conservation. Thus, a interesting way of diffusion of specimens' biodiversity located in San Luis is trough mobile applications, which are highly popularised.

As was mentioned, an AR system allows overlay computer-generated contents on views of the physical scene. In this case, an AR mobile application was implemented where image-based tracking uses 2D targets denominated *markers*. The goal is to stimulate learning interests in a less complex way when compared to other traditional teaching methods (books, web sites, among others).

The application consisted of 3D visualisation of native reptiles and amphibians so that people can understand and observe the skin texture, the traits and the structure. The user must use the camera of mobile device to identify visual specific markers (one marker per specimen), to showcase an overlay only when a marker is sensed by the device (See Fig. 2b).

Once the marker is recognised, a 3D virtual version of the corresponding specimen is displayed on the screen. In such a way, the user will observe it in more detail and from various angles and by rotating the marker would rotate the virtual replication as well.



(a) Immersive Virtual Reality System (b) Augmented Reality System

Fig. 2: Visualization Systems

4.2 3D printing

The end-to-end process from *Reality Computing* ends with 3D printing. Once a suitable 3D model is created, it can quickly be produced by a 3D print.

3D printing is a form of additive manufacturing technology where a three dimensional object is created by laying down or build from successive layers of material. 3D printing is a great way to create objects because you can create objects that you couldn't make otherwise without having complex expensive molds created or by having the objects made with multiple parts.

A 3D printed herpetological specimen can be edited by simply editing the 3D model; in such manner, for example, it is possible to scale a reptile or modify its original pose.

A type of filament called Acrylonitrile Butadiene Styrene (ABS), was used for printing reptile models. According to the characteristics of each model, printing tests and modifications were made to the models in order to achieve prints without the need to use supports on the parts. It is because the supports must be cut increasing the chances of damaging the printed specimen.

Figure 3 shows a real specimen and its printing.



(a) Real Specimen (b) Printed Specimen

Fig. 3: Reality Computing

5 Conclusions and Future Works

This paper describes an approach to digitising herpetological specimens. The digitising process allows to built a 3D model from a real object. Because different ways lead to a 3D model for the development of this work, more than one technique was tested.

At the beginning, this work consisted in the survey of a representative set of type specimens, which were digitised through the *Laser Scanning Method* in

order to achieve a three-dimensional textured model of each one. According with the *Unidad de Herpetología* feedback these models and theirs characteristics were not suited to the needs, interests and preferences, herpetologically speaking. Then, the *Photogrammetry Method* was carried out which allowed obtaining models with a higher resolution; which in turn were considered as more accurate replicas and according to the needs of the *Unidad de Herpetología*.

In order to the digitised models to be visualised, two applications were developed: an *Augmented Reality* application that allows the visualisation of the specimens in an attractive way and from any mobile device; and an *Immersive Virtual Reality* application that achieves a more personalised and detailed use of the digitised samples. In addition, the *3D prints* allowed to obtain tangible replicas of the herpetological specimens in which it is possible to appreciate the non-textured characteristics of a specimen.

Given the interdisciplinarity of the work carried out, the technologies used for its implementation and the unique characteristics of the collections, the process generated is, in itself, unique and innovative, and its result is a trigger for a wide range of scientific and educational activities. The obtained 3D photorealistic models offer new possibilities for the daily practice of herpetology, protection and conservation of specimens. As well as being objective 3D documentations, they achieve a reliable conservation of the herpetological heritage.

Future works will include: reproduce the process with all the specimens of the *Unidad de Herpetología*; build a 3D digital library of the *Unidad de Herpetología*, which can be accessed online; incorporate animations to give more realism to the models; automate the process related to the capture of photographic samples, with the necessary equipment, as a rotating base, light-box photo studio, among others.

References

1. Christophe Degryse. Digitalisation of the economy and its impact on labour markets. Working Paper 2016.02, European Trade Union Institute, Brussels, 2016.
2. Harry Bouwman, Mark de Reuver, and Shahrokh Nikou. The impact of Digitalization on Business Models: How IT Artefacts, Social Media, and Big Data Force Firms to Innovate Their Business Model. Technical report, 2017.
3. Nady Hoyek, Christian Collet, Franck Di Rienzo, Mickael De Almeida, and Aymeric Guillot. Effectiveness of three-dimensional digital animation in teaching human anatomy in an authentic classroom context. 7, 11 2014.
4. Hammad H. Malik, Alastair R.J. Darwood, Shalin Shaunak, Priyantha Kulatilake, Abdulrahman A. El-Hilly, Omar Mulki, and Aroon Baskaradas. Three-dimensional printing in surgery: a review of current surgical applications. *Journal of Surgical Research*, 199(2):512 – 522, 2015.
5. P. Santos, M. Ritz, C. Fuhrmann, and D. Fellner. 3d mass digitization: a milestone for archeological documentation. *Virtual Archaeology Review*, 8(16):1–11, 2017.
6. Nicola Lercari, Jurgen Shulze, Willeke Wendrich, Benjamin Porter, Margie Burton, and Thomas E. Levy. 3-D Digital Preservation of At-Risk Global Cultural Heritage. In C. Catalano and L. De Luca, editors, *Eurographics Workshop on Graphics and Cultural Heritage*. The Eurographics Association, 2016.

7. Claude Lavoie. Biological collections in an ever changing world: Herbaria as tools for biogeographical and environmental studies. *Perspectives in Plant Ecology, Evolution and Systematics*, 15(1):68 – 76, 2013.
8. D. Bakk and R. Urban. Scientific and social value of biological collections. pages 6–9, 08 2014.
9. Reed S. Beaman and Nico Cellinese. Mass digitization of scientific collections: New opportunities to transform the use of biological specimens and underwrite biodiversity science. *Biological Reviews*, 209:717, 2012.
10. S. Keaveney, C. Keogh, L. Gutierrez-Heredia, and E. G. Reynaud. Applications for advanced 3d imaging, modelling, and printing techniques for the biological sciences. In *2016 22nd International Conference on Virtual System Multimedia (VSMM)*, pages 1–8, Oct 2016.
11. Francesco Nex and Fabio Remondino. Uav for 3d mapping applications: A review. 6, 03 2014.
12. Wassim Moussa. Integration of digital photogrammetry and terrestrial laser scanning for cultural heritage data recording, 01 2014.
13. Mesapam Shashi and Kamal Jain. Use of photogrammetry in 3d modeling and visualization of buildings. 2, 01 2007.
14. Mozhdeh Shahbazi, Gunho Sohn, Jrme Thau, and Patrick Menard. Development and evaluation of a uav-photogrammetry system for precise 3d environmental modeling. 15:27493–27524, 10 2015.
15. Alfonso Ippolito and Michela Cigola. *Handbook of Research on Emerging Technologies for Digital Preservation and Information Modeling*. 01 2017.
16. D. Vajak and . Livada. Combining photogrammetry, 3d modeling and real time information gathering for highly immersive vr experience. In *2017 Zooming Innovation in Consumer Electronics International Conference (ZINC)*, pages 82–85, May 2017.
17. S. Aukstakalnis. *Practical Augmented Reality: A Guide to the Technologies, Applications, and Human Factors for AR and VR*. Usability Series. Addison-Wesley, 2016.
18. Peng Chen, Xiaolin Liu, Wei Cheng, and Ronghuai Huang. A review of using augmented reality in education from 2011 to 2016, 09 2017.
19. Simon Ford and Tim Minshall. 3d printing in teaching and education: A review of where and how it is used, 10 2017.
20. B. Espeche, Y. Alvarado, G. Rodríguez, N. Jofré, M. Jofré, and R. Guerrero. Repositorio virtual de ejemplares - Unidad de Herpetología - UNSL, Argentina. In *XVIII Congreso Argentino de Herpetología*, 2017.
21. Naci Yastikli. Documentation of cultural heritage using digital photogrammetry and laser scanning. *Journal of Cultural Heritage*, 8(4):423 – 427, 2007.
22. Jocelyn Behm, Brenna R. Waite, S. Tonia Hsieh, and Matthew R. Helmus. Benefits and limitations of three-dimensional printing technology for ecological research. *bioRxiv*, 2018.
23. Giovanna Sansoni, Marco Trebeschi, and Franco Docchio. State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, 9(1):568–601, 2009.
24. Susan C Kuzminsky and Megan S Gardiner. Three-dimensional laser scanning: potential uses for museum conservation and scientific research. *Journal of Archaeological Science*, 39(8):2744–2751, 2012.
25. Toni F Schenk. *Fotogrametría digital*. Number Sirsi) i9788426713315. 2002.
26. R. Deutsch. *Convergence: The Redesign of Design*. AD Smart. Wiley, 2017.

Segmentación multi-atlas de imágenes médicas con selección de atlas inteligente y control de calidad automático

Lucas Mansilla[1], Enzo Ferrante[1]

[1] *Sinc(i)*, Instituto de Investigaciones en Señales, Sistemas e Inteligencia Computacional, Universidad Nacional del Litoral / CONICET

Ciudad Universitaria UNL, Ruta Nacional Nº 168, km 472.4,
FICH, 4to Piso (3000) Santa Fe – Argentina
{lmansilla, eferrante}@sinc.unl.edu.ar

Abstract. La segmentación multi-atlas (MAS por sus siglas en inglés) es una de las técnicas más utilizadas para la segmentación anatómica de imágenes médicas. Esta técnica utiliza un conjunto de atlas (imágenes previamente etiquetadas por expertos) para estimar la segmentación correspondiente a una nueva imagen, por medio de la registración y fusión de dichos atlas. Dado que los atlas de una población pueden variar signi cativamente entre sí, la selección de aquellos atlas más apropiados contribuye a mejorar su desempeño. En este trabajo, se presenta un estudio comparativo de segmentación basada en multi-atlas que utiliza distintos métodos de selección del atlas aleatorios y basados en similaridad de imagen en el contexto de imágenes de radiografía torácica, para medir el impacto dicha selección en la precisión del algoritmo. Adicionalmente, para evaluar el desempeño del modelo en contextos en donde no se dispone de segmentaciones de referencia (también conocidas como *ground-truth*), se propone el uso del índice de Reverse Classi cation Accuracy (RCA) como estimador automático que posibilita el control de calidad de las segmentaciones en ausencia de segmentaciones de referencia. Los resultados obtenidos en dos grandes bases de datos demuestran que (i) la selección de atlas basada en similitud mejora signi cativamente los resultados al segmentar imágenes radiográ cas y (ii) por medio de RCA es posible determinar un valor umbral de referencia que permite descartar automáticamente las segmentaciones de mala calidad en ausencia de *ground-truth,* tarea fundamental si se pretende avanzar en la utilización de este tipo de estrategias en el ámbito clínico real.

Keywords: segmentación basada en multi-atlas, selección del atlas, reverse classi cation accuracy, segmentación, imágenes médicas

1 Introducción

La segmentación de imágenes es un problema de gran interés para la comunidad médica dado que constituye un paso fundamental en la extracción de información útil para el diagnóstico, seguimiento y tratamiento de diversas patologías. En particular, la segmentación automática de estructuras anatómicas en imágenes de radiografía de tórax (tales como pulmones y corazón) reviste gran importancia para el diagnóstico asistido por computadora por medio de estas imágenes dado que facilita, entre otras cosas, la cuanti cación de diversas propiedades morfológicas en dichas estructuras [11]. En la actualidad, una amplia gama de enfoques ha sido propuesta para resolver el problema de segmentación anatómica, siendo la segmentación multi-atlas [1] uno de los más adoptados debido a su robustez.

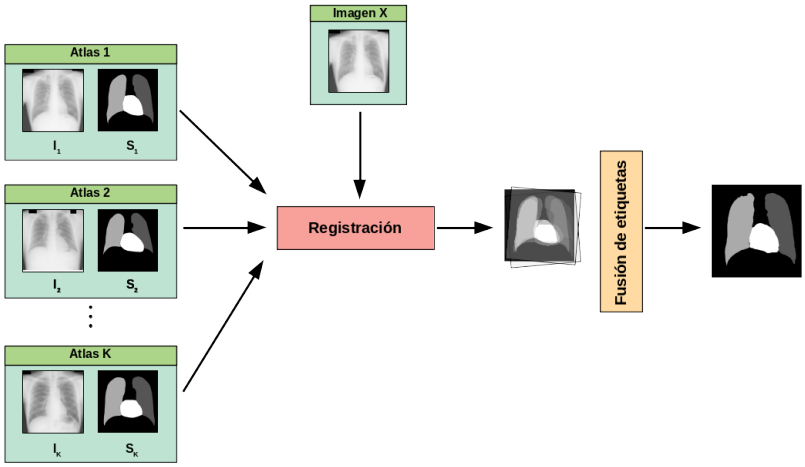


Fig. 1. Esquema del método de segmetación multi-atlas (MAS). La imagen de intensidad I_i de cada atlas se transforma para que resulte similar a la imagen a segmentar X. Esta transformación se aplica a la máscara de segmentación S_i de cada atlas. La segmentación nal se obtiene mediante la fusión de las máscaras deformadas.

La segmentación multi-atlas (MAS) [1] utiliza un conjunto de atlas, esto es, pares formados por una imagen y su correspondiente máscara de segmentación, para obtener nuevas segmentaciones sobre imágenes que aún no han sido etiquetadas. Para lograrlo, la imagen de intensidad de cada atlas se transforma para que resulte similar a la imagen a segmentar. Esta transformación se aplica a la máscara de segmentación de cada atlas, para luego obtener la segmentación nal mediante la combinación de las máscaras deformadas. Este último paso es conocido como fusión de etiquetas.

El proceso de deformar una imagen para que resulte similar a otra se conoce como registración deformable de imágenes [10]. En este contexto, la imagen de intensidad del atlas se denomina imagen móvil *M* y la imagen a segmentar se denomina imagen ja *F*. El resultado de registrar la imagen móvil *M* en la imagen ja *F* es una transformación *T** de *M* a *F*. La transformación óptima *T** es obtenida minimizando la siguiente función de energía:

$$T^* = \arg\min_T \Phi(M \circ T, F) + R(T), \tag{1}$$

donde Φ corresponde al criterio de (dis)similitud entre las dos imágenes *M* y *F*, la función *R* corresponde al término de regularización que impone restricciones geométricas a la solución, y *T* parametriza una transformación que mapea cada punto de la imagen *M* con *F*. Resolviendo este problema de minimización, es posible computar la transformación óptima *T** alineando dichas imágenes. Nótese que la transformación *T** puede ser lineal o no lineal. En este trabajo utilizaremos transformaciones no lineales deformables (representadas por un campo vectorial que indica las deformaciones a nivel pixel) dado que los métodos multi-atlas alcanzan mayor precisión cuando las imágenes son deformadas localmente antes de fusionar las etiquetas. En segmentación multi-atlas, esta transformación *T** es aplicada a la segmentación *S* de *M* en un proceso conocido como propagación de etiquetas, para obtener una segmentación deformada que representa una segmentación candidata para la imagen ja *F* [2].

La selección del atlas es un punto importante de todo algoritmo MAS. Inicialmente, se puede utilizar el total de los atlas disponibles en el dataset. Sin embargo, también es posible considerar un subconjunto del total disponible, dado que el proceso de registración suele ser costo en términos computacionales. De esta forma, es posible mejorar la e ciencia computacional del algoritmo al tener una mejor cantidad de atlas que registrar. Otro punto a tomar en cuenta es que la segmentación nal no depende sólo de una única registración sino de cada una de las registraciones de los atlas seleccionados, por lo que un proceso de selección de los atlas puede ayudar a mejorar la precisión de los resultados nales obtenidos, por ejemplo, descartando aquellos atlas que son muy diferentes a la imagen a segmentar y considerando sólo los que resulten mas similares a esta [1, 2].

El nivel de precisión de un método de segmentación se suele medir evaluando la calidad de las segmentaciones resultantes. Tradicionalmente, esto es llevado a cabo usando diferentes métricas [3], en donde la segmentación predicha es comparada con una segmentación de referencia, también llamada *ground-truth* (GT), generada manualmente por un experto. La medida más popular en el contexto de la segmentación de imágenes médicas es el coe ciente de Dice-Sorensen (DSC) que

mide el solapamiento entre la segmentación estimada Y y la segmentación de referencia GT como:

$$DSC(Y, GT) = \frac{2|Y \cap GT|}{|Y| + |GT|} \quad (2)$$

Sin embargo, cuando se desea evaluar la performance de un algoritmo de segmentación sobre bases de datos sin ground-truth, no es posible aplicar estos métodos debido a la ausencia de máscaras de referencia. Este escenario es en realidad el más usual, dado que en la práctica cotidiana los métodos son desarrollados para segmentar imágenes que aún no han sido etiquetadas, y disminuir el trabajo de los expertos médicos evitando que realicen anotaciones manualmente. En este trabajo, adoptaremos la técnica Reverse Classi cation Accuracy (RCA) [4] para evaluar la calidad de la segmentación de una imagen individual en ausencia de GT, incorporando de esta forma control de calidad automático al proceso de segmentación. En este método, la segmentación predicha sirve de "ground truth" para entrenar un clasi cador RCA que luego es evaluado con un conjunto de imágenes que sí disponen de GT. El mejor de los resultados de esta evaluación es el que se utiliza como medida de referencia para evaluar la calidad del método de segmentación.

Contribuciones de este trabajo: Este trabajo presenta dos contribuciones mayores: (i) en primer lugar, un estudio comparativo de la in uencia de la técnica de selección del atlas en la calidad de las segmentaciones obtenidas en el contexto de imágenes de radiografía torácicas, evaluado en la base de datos pública JSRT [6] y alcanzando resultados del estado del arte; y (ii) la generación de máscaras de segmentación para un subconjunto del dataset público X-Chest [9] publicado recientemente (el cual no cuenta con segmentaciones manuales) incluyendo, para cada una, un estimador de la calidad de la segmentación obtenido mediante RCA que permite descartar segmentaciones erroneas.

2 Método

El método implementado para la segmentación y control de calidad automático de los resultados se compone de cuatro etapas que se describen detalladamente a continuación.

1. Selección inteligente de los atlas: El modelo propuesto de segmentación basada en multi-atlas incorpora un componente de selección de atlas para mejorar la calidad de las segmentaciones, y un componente RCA para el control automático de las mismas. Supongamos que se dispone de un conjunto de N atlas $\{(I_1, S_1), (I_2, S_2), \ldots, (I_N, S_N)\}$ donde I_i corresponde a una imagen en escala de grises y S_i a su

correspondiente segmentación. De este conjunto, se seleccionan los K atlas $\{(I_1, S_1), (I_2, S_2), \ldots, (I_K, S_K)\}$ más similares a la imagen que desea segmentar X, según una medida de similaridad de imagen de nida. Existen diversas medidas de similitud entre imágenes, algunas de ellas basadas en principios estadísticos como la información mutua (MI) [8] (considerando las distribuciones de intensidad en ambas imágenes como si fueran distribuciones de probabilidad y estimando la distribución conjunta) o algunas más simples como la distancia Euclidiana (D2) entre las imágenes I_i y X vectorizadas. En este trabajo, evaluaremos ambas medidas como método de selección de atlas y las compararemos con un método de selección aleatorio.

2. Registración de las imágenes: Cada una de las K imágenes I_i (seleccionadas en el paso anterior) es registrada con la imagen X de forma independiente. En dichos procesos de registración, las imágenes I_i son consideradas móviles mientras que la imagen X es considerada la imagen ja (ver Fig. 1).

Dicha registración es realizada minimizando la ecuación (1) de nida en la sección anterior. En este trabajo se utilizó el paquete SimpleElastix [7] que implementa algoritmos modulares de registración deformable. Como función de disimilitud Φ se utilizó una medida basada en la información mutua entre las imágenes. Adicionalmente, como función de regularización R se utilizó un término de penalización basado en la exión del campo de deformación [12]. Al nalizar el proceso de registración de dichas imágenes, se obtienen K segmentaciones deformadas $\{D_1, D_2, \ldots, D_K\}$, todas ellas alineadas al espacio de coordenadas de la imagen X.

3. Fusión de etiquetas: Las segmentaciones obtenidas son combinadas utilizando un mecanismo de fusión de etiquetas resultando en una segmentación Y nal correspondiente a la imagen X. Uno de los mecanismos más simples utilizados para realizar fusión de etiquetas es el voto por mayoría. En ese caso, para cada pixel x de la imagen $X(x)$ se analiza la etiqueta otorgada por cada una de las K segmentaciones $\{D_1(x), D_2(x), \ldots, D_K(x)\}$ y se asigna aquella que reciba la mayor cantidad de votos. Este mecanismo es sumamente simple, pero no incorpora al proceso de fusión de etiquetas información proveniente de los píxeles vecinos, resultando en ocasiones en segmentaciones espúreas alejadas de la realidad. Por esta razón, en este trabajo utilizaremos un método avanzado de fusión de etiquetas conocido *simultaneous truth and performance level estimation (STAPLE)* [5] que intenta maximizar el consenso entre las etiquetas imponiendo a la vez restricciones de homogeneidad entre píxeles vecinos.

4. Estimación de la calidad de la segmentación por RCA: La idea detrás del concepto de Reverse Classi cation Accuracy [4] es simple: en escenarios donde no contamos con segmentaciones de referencia en la imagen X para evaluar la calidad de la segmentación obtenida Y, utilizaremos el par (X, Y) de forma inversa, es decir, entrenaremos un algoritmo de segmentación usando sólo el par (X, Y) para segmentar

imágenes de las cuales sí tenemos segmentaciones de referencia. En el contexto de este trabajo, utilizaremos un algoritmo de segmentación basado en un único atlas como estimador RCA en lugar de múltiples atlas, es decir, que no poseerá etapa de fusión de etiquetas dado que contaremos con una única segmentación candidata. El par (X, Y) será utilizado como atlas único para segmentar las imágenes de las cuales poseemos segmentaciones de referencia. Luego, se evaluará la calidad de los resultados obtenidos, y tomaremos el máximo de dichas calidades como un estimador de la calidad real de la segmentación Y. La hipótesis sobre la que se sustenta este método es que si la imagen X fue segmentada correctamente con la máscara Y, entonces el par (X, Y) servirá como atlas para segmentar correctamente al menos una imagen del dataset con referencias (aquella que sea más parecida a la imagen X). La métrica a utilizar para cuanti car la calidad de las segmentación será el coe ciente DSC de nido en la ecuación 2. Para una discusión completa sobre RCA ver [4].

3 Experimentos

Datasets. El modelo propuesto se evaluó en la tarea de segmentación de estructuras anatómicas en imágenes de radiografía de tórax. En primera instancia, se evalúo el desempeño del modelo bajo distintos métodos de selección del atlas, utilizando para ello la base de datos JSRT [6]. Esta es una base de datos de acceso público, compuesta por 247 imágenes de radiografías de tórax AP de 1024x1024 pixeles. Para cada imagen, se proveen segmentaciones manuales de pulmones y corazón, que serán utilizadas en este trabajo para evaluar el impacto de la estrategia de selección del atlas en los resultados obtenidos. Notar que en este caso se cuenta con segmentaciones de referencia para todas las imágenes, razón por la cual el coe ciente DSC puede ser calculado directamente sin necesidad de estimarlo por medio de RCA.

En segunda instancia, se evaluó el desempeño del modelo con RCA utilizando una muestra aleatoria de 500 imágenes de radiografías de tórax AP sin segmentaciones de referencia, extraída del dataset X-Chest [9] recientemente publicado. Dicho dataset no cuenta con segmentaciones anatómicas a nivel pixel, por lo que se utilizó el método RCA para estimar el coe ciente DSC y poder realizar un control automático de calidad que permita descartar segmentaciones defectuosas. En ambos casos, todas las imágenes utilizadas fueron reducidas a un tamaño de 256x256 pixeles para disminuir el costo computacional del método, y las segmentaciones obtenidas fueron posteriormente resampleadas a la resolución original de las imágenes.

Estudio comparativo de métodos para selección de atlas. Del conjunto de 247 imágenes con segmentaciones de referencia que componen la base de datos JSRT, 50 fueron utilizadas para evaluar el modelo bajo distintos métodos de selección del atlas, utilizando las restantes 197 como base de datos a partir de las cuales se seleccionaron

los atlas utilizados para segmentar cada imagen. De esta manera, se corrieron 3 instancias del algoritmo, cada una con un método de selección diferente, utilizando *K=5* atlas para segmentar cada imagen. La Tabla 1 muestra el DSC promedio en cada estructura anatómica para selección por información mutua (MI), selección por mínima distancia Euclidiana al cuadrado (D2) y selección aleatoria (Random), alcanzando resultados comparables al estado del arte para JSRT reportados en [11].

Tabla 1. Valores de DSC promedio por etiqueta utilizando información mutua (MI), distancia Euclidiana al cuadrado (D2) y selección aleatoria (Random) de atlas. En cada estructura, el DSC nal se obtuvo promediando entre los DSC calculados para esa estructura en particular en las 50 imágenes de test.

Método	Pulmón derecho	Pulmón izquierdo	Corazón
MI	0.940 (0.020)	0.950 (0.021)	0.882 (0.050)
D2	0.932 (0.034)	0.944 (0.030)	0.862 (0.065)
Random	0.920 (0.031)	0.929 (0.029)	0.826 (0.109)

Para una mejor visualización, en la Fig. 2 se muestra la distribución de valores de DSC en cada estructura utilizando boxplots. El modelo con información mutua es el que logró mejores medidas de DSC para todas las estructuras anatómicas.

Control de calidad automático con RCA. Dado que el modelo de segmentación MAS con selección de atlas candidatos por información mutua fue el que presentó mejor rendimiento, éste fue utilizado para segmentar las 500 imágenes del dataset X-Chest, utilizando RCA para estimar los valores de DSC, considerando igualmente *K=5* atlas candidatos por cada imagen a segmentar. Cabe destacar que los atlas candidatos utilizados para segmentar las imágenes de X-Chest fueron seleccionados a partir de las 197 imágenes del dataset JSRT con anotaciones de referencia. En la Fig. 3 se muestra la distribución de valores de DSC estimados por medio de RCA para el dataset sin anotaciones de referencia. Para incorporar el control de calidad automático, se calculó el valor de DSC mínimo (DSC_{min}) necesario que debe tener una imagen segmentada para no ser descartada. Siendo 0.88 y 0.04, la media y la desviación estándar de la distribución de valores de DSC estimados con RCA, se estableció DSC_{min}= 0.88 – 0.04 = 0.84. Así, una imagen segmentada con un valor de DSC inferior a 0.84 será descartada automáticamente.

En la Fig. 4 se muestran algunas segmentaciones obtenidas junto con el DSC calculado por RCA. Se puede apreciar visualmente que el modelo es capaz de producir resultados de segmentación muy precisos, como es el caso de los dos ejemplos de la parte superior. Los dos restantes de la parte inferior son ejemplos en donde el modelo no se comporta adecuadamente, en este caso, estas segmentaciones resultarían descartadas automáticamente por tener un valor DSC menor a DSC_{min}.

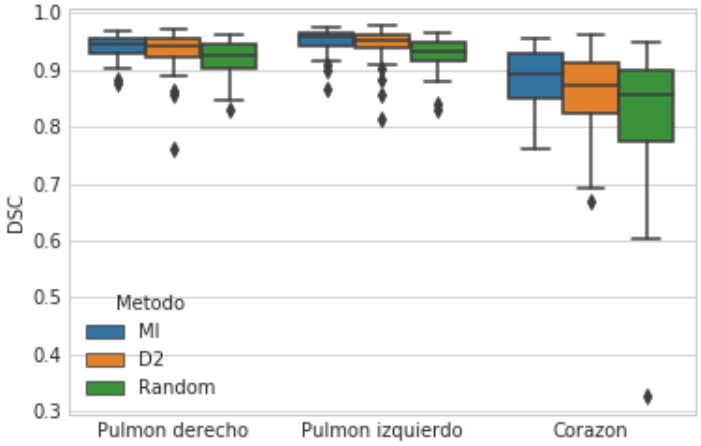


Fig. 2. Valores DSC por etiqueta utilizando información mutua (MI), distancia Euclidiana al cuadrado (D2) y selección aleatoria de atlas. Para medir la signi cancia estadística de las diferencias entre las distribuciones se utilizó el test de Wilcoxon por cada posible combinación de métodos, que arrojó p-value < 0.05 para todos los casos, siendo la única excepción el par MI-D2 para la etiqueta corazón.

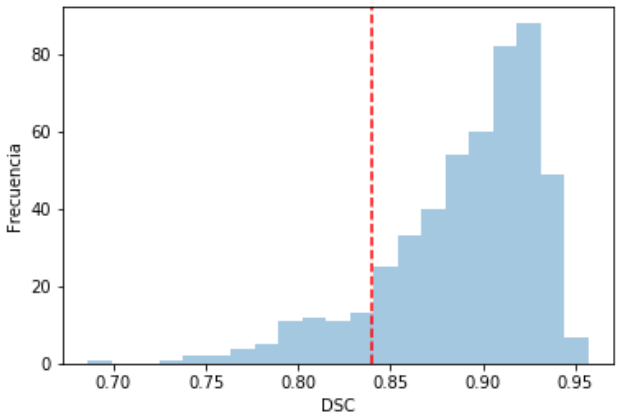


Fig. 3. Distribución de valores de DSC estimados con RCA. La línea vertical roja vertical representa el mínimo valor de DSC (DSC_{min}=0.84) necesario que debe tener una imagen segmentada para no ser descartada.

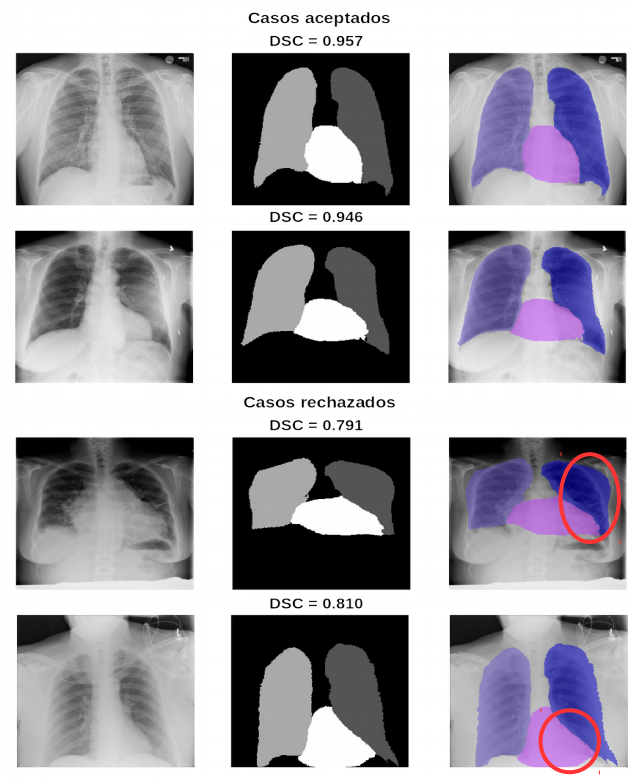


Fig. 4. Ejemplos de segmentaciones obtenidas junto con su correspondiente DSC estimado por RCA. (Izquierda) Ejemplos de imágenes a segmentar. (Centro) Segmentaciones producidas por multi-atlas con selección por MI. (Derecha) Se muestran los pares alineados. Puede observarse que en los casos rechazados por el control automático de calidad, las segmentaciones resultantes poseen serios errores que han sido resaltados en rojo.

4 Conclusiones y trabajos futuros

En este trabajo se presentó un método de segmentación basado en multi-atlas, que incorpora un componente de selección de atlas candidatos basado en similitud y utiliza RCA para medir la calidad de la segmentación en contextos donde no se dispone de segmentaciones de referencia. Los resultados obtenidos muestran que, cuando se utilizan métodos basados en similaridad de imagen para seleccionar atlas candidatos en el contexto de imágenes radiográ cas de torax, la precisión del modelo aumenta signi cativamente. Para evaluar el desempeño del mejor modelo de segmentación obtenido cuando no se dispone de información de referencia, se incorporó la técnica RCA para estimar la calidad de la segmentación en ausencia de

GT. A partir de los valores de DSC estimados por RCA, se pudo de nir un valor mínimo (DSC_{min}) para efectuar un control de calidad de las segmentaciones de forma automática y detectar segmentaciones defectuosas alejadas de la realidad.

El uso combinado de RCA y segmentación multi-atlas abre las puertas al desarrollo de métodos que utilicen dichas segmentaciones (junto con las estimaciones de calidad provistas) como insumo para, por ejemplo, clasi car dichas imágenes como sanas o patológicas. El dataset utilizado X-Chest provee, para cada imagen, el diagnóstico médico indicando si la misma corresponde a un paciente sano o posee alguna patología como cardiomegalia o pneumo-torax, entre otras. En el futuro, se pretende utilizar la información provista por las segmentaciones obtenidas en conjunto con su indicador de calidad y las imágenes originales para intentar mejorar la tasa de acierto de los métodos existentes para el diagnóstico automático de estas enfermedades [9].

Referencias

1. Iglesias, J.E., Sabuncu, M.R.: Multi-atlas segmentation of biomedical images: a survey. Medical Image Analysis, 24(1), 205-219 (2015)
2. Agarwal, M.: Multi-Atlas Based Image Segmentation Applied on Pulmonary CT Scan Data (2018)
3. Crum, W. R., Camara, O., Hill, D. L.: Generalized overlap measures for evaluation and validation in medical image analysis. IEEE TMI, 25(11), 1451 -1461 (2006)
4. Valindria, V. et al.: Reverse Classi cation Accuracy: Predicting Segmentation Performance in the Absence of Ground Truth. IEEE TMI, 36(8), 1597-1606 (2017)
5. War eld, S. et al.: Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation. IEEE TMI, 23(7), 903-921 (2004)
6. Shiraishi, J. et al.: Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. AJR 174; 71-74 (2000)
7. Marstal, K., Berendsen, F., Staring, M., Klein, S.: SimpleElastix: a user-friendly, multi-lingual library for medical image registration. In: IEEE CVPR (2016)
8. Viola P., Wells, W.M.: Alignment by Maximization of Mutual Information International Journal of Computer Vision, 24(2) 137-154 (1997)
9. Wang, X. et al.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classi cation and localization of common thorax diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
10. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable Medical Image Registration: A Survey. IEEE Transactions on Medical Imaging, 32(7), 1153–1190 (2013)
11. Van Ginneken, B. et al.: Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. MedIA, 10(1), 19-40 (2006)
12. Rueckert, D. et al.: Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Trans. Med. Imaging 18(8), 712-721 (1999)

Mejora de imágenes a color utilizando un enfoque morfológico multiescala

Rodolfo Cardozo, Angel Méndez,
Horacio Legal Ayala, and José Luis Vázquez Noguera

Universidad Nacional de Asunción, Facultad Politécnica
San Lorenzo, Paraguay
{rkrdozo,aramfara}@gmail.com
{hlegal,jlvazquez}@pol.una.py

Abstract. La mejora de imágenes tiene aplicaciones prácticas en distintas áreas científicas. Una estrategia para la mejora de imágenes es la morfología matemática. Uno de los desafíos de la morfología matemática en imágenes a color es determinar un orden entre las componentes de la imagen. Las imágenes a color se representan mediante estructuras de datos multidimensionales lo que implica que no existe un orden natural entre sus componentes. En este trabajo se propone la mejora de imágenes utilizando la extensión de la morfología matemática multiescala para imágenes a color con diferentes espacios de colores y métodos de ordenamiento. Los experimentos realizados muestran que el método de mejora de la imagen propuesto, genera resultados auspiciosos utilizando distintos métodos de ordenamiento en términos de contraste tanto local como global, así como la calidad del color la imagen.

Keywords: Imágenes, Color, Mejora de Imagen, Morfología Matemática, Espacio de color, Ordenamiento

1 Introducción.

El procesamiento y análisis de imágenes digitales consiste en un conjunto de métodos y técnicas aplicables a las imágenes digitales con el objetivo de obtener información relevante de las mismas. Las imágenes a color permiten obtener mayor cantidad y calidad de información que pueden ser aprovechadas en distintas aplicaciones prácticas para diversas áreas de investigación [10].

Las imágenes a color nos lleva a definir el concepto de espacio de color. Los espacios de color son modelos de referencia donde se describe la forma en que se organizan y se representan los colores[8]. Algunos espacios de color que podemos citar son RGB, L*a*b*, HLS, HSI, HSV, CMYK.

Las imágenes a color requieren de estrategias para que puedan ser procesadas, la morfología matemática es una estrategia de procesamiento de imágenes digitales que consiste en una serie de definiciones teóricas y técnicas basadas en la teoría de conjuntos [10]. Los operadores definidos para la morfología matemática permiten realizar operaciones sobre las imágenes. Los operadores básicos de la

morfología matemática son la erosión y la dilatación, a partir de las operaciones básicas de erosión y dilatación se extienden otras operaciones como la apertura y la clausura morfológica, entre otros.

Las operaciones de la morfología matemática requieren la comparación de los píxeles de las imágenes. En imágenes de escala de grises realizar la comparación entre píxeles es un proceso casi trivial, ya que se tiene un orden natural con los valores de los píxeles, por otra parte, comparar los píxeles de las imágenes a color no es un proceso trivial y requiere un esfuerzo extra para realizar la comparación de píxeles.

Las imágenes a color, requieren la comparación de estructuras de datos multidimensionales de las que se conforman los píxeles de una imagen a color, algunas estrategias de ordenamientos que podemos citar son, el ordenamiento lexicográfico [1] con sus variantes alfa lexicográfico y modulo lexicográfico [10], ordenamiento por distancia euclidiana en los espacios RGB y L*a*b* [10], ordenamiento de entrelazo de bits [4], ordenamiento utilizando información información del histograma de la imagen [15], entre otros.

La propuesta del presente trabajo es la mejora de la imagen utilizando la morfología matemática multiescala con distintos métodos de ordenamientos y espacios de color. Los objetivos del mismo son extender la morfología matemática multiescala de imágenes en escala de grises a imágenes a color, analizar y evaluar la mejora del contraste local, global y la mejora del color de la imagen utilizando distintos métodos de ordenamiento.

La comparación del método de mejora de la imagen multiescala utilizando la transformada Top-hat multiescala [12] con otros métodos del estado del arte queda cómo propuesta de trabajo futuro. En éste sentido se podrían extender otros métodos multiescala como por ejemplo [3],[11].

La distribución de las secciones del artículo se realiza de la siguiente manera. En la sección 2 se presenta el Marco Teórico. En la sección 3 se presenta la Propuesta del trabajo. En la sección 4 se muestran los Resultados Experimentales. En la sección 5 se presentan las Conclusiones del trabajo.

2 Marco Teórico.

Una imagen digital RGB puede ser representada como una función $f: \mathbb{Z}^2 \Rightarrow \mathbb{Z}^n$, en donde cada píxel en la posición $(r, c) \in \mathbb{Z}^2$ se representa mediante una estructura de datos multidimensional. En el espacio de color RGB los colores se forman de tres planos de imágenes independientes $C = \{R, G, B\}$, cada plano puede tratarse cómo una imagen en escala de grises que corresponde a un color primario rojo, verde y azul.

El espacio de color es una notación que permite especificar, organizar y combinar los colores a ser utilizados para el procesamiento de las imágenes digitales [8]. Existen una gran variedad de espacios de color entre los que podemos citar el espacio de color HSI [14], el espacio de color L*a*b* [6] y el espacio de color RGB (Red, Gren, Blue) representa un color mediante cantidades positivas de rojo, verde y azul [10].

2.1 Morfología matemática.

La morfología matemática es un área del procesamiento digital de imágenes y se basa en la teoría de conjuntos, principios algebraicos y geométricos [10]. Las operaciones básicas de la morfología matemática son la erosión y la dilatación [10]. La operación de erosión $\varepsilon_B(f)$ se obtiene cómo resultado de obtener el menor valor dentro de una ventana B llamada elemento estructurante .

$$\varepsilon_B(f)(r,c) = \min\{f(r+d, c+l) | (d,l) \in B\}. \quad (1)$$

La operación de dilatación δ_B se obtiene con el elemento de mayor valor dentro del dominio del elemento estructurante B.

$$\delta_B(f)(r,c) = \max\{f(r-d, c-l) | (d,l) \in B\}. \quad (2)$$

La morfología matemática tiene otras operaciones como la apertura y clausura que se obtienen realizando combinaciones de las operaciones básicas. La apertura morfológica de una imagen f por un elemento estructurante B se denota por $f \circ B$ y se define como la erosión de f por B, seguida de la dilatación por el mismo elemento estructurante.

$$f \circ B = \delta(\varepsilon(f, B), B). \quad (3)$$

La clausura de una imagen f por un elemento estructurante B se denota por $f \bullet B$ y se define como la dilatación de f por B, seguida de la erosión por el mismo elemento estructurante.

$$f \bullet B = \varepsilon(\delta(f, B), B). \quad (4)$$

La transformada Top-hat es una técnica de mejora del contraste que utiliza las operaciones la morfología matemática combinando las operaciones de apertura y clausura [12]. La transformada Top-hat puede ser del tipo White Top-hat (WTH) o Black Top-hat. La transformada WTH es utilizada para resaltar regiones brillantes de la imagen que son más pequeñas que el elemento estructurante. La transformada WTH se obtiene realizando una diferencia entre la imagen original f y la apertura $(f \circ B)$ de la misma imagen es decir:

$$WTH(f) = f - (f \circ B). \quad (5)$$

La transformada Black Top-hat (BTH) es utilizada para extraer regiones oscuras de una imagen y que son más pequeñas que el elemento estructurante. La transformada BTH se obtiene realizando una diferencia entre la operación de clausura $(f \bullet B)$ de una imagen y la imagen original f, es decir:

$$BTH(f) = (f \bullet B) - f. \quad (6)$$

2.2 Estrategias de ordenamiento.

En esta sección se describen distintas estrategias de ordenamientos, entre las que podemos mencionar y clasificar como dependientes e independientes de la información del histograma de la imagen.

Ordenamiento independientes del histograma.

- **Lexicográfico (LEX):** Se basa en la atribución de prioridades a los componentes del vector para que unos tengan mayor importancia que otros en el momento de comparar y definir el orden [5]. Existen variantes, cómo por ejemplo el **Alfa Lexicográfico (ALEX)**, donde se reduce la cantidad de veces en que el orden se decide por el primer componente del vector sumándole un valor α al primer componente del vector [10]. La variante **Módulo Lexicográfico (AMLEX)** reduce la cantidad de veces que el orden se decide por el primer componente del vector es realizando una división entera entre el primer componente del vector y un valor constante α [10]. El ordenamiento lexicográfico tiene variantes con distintos espacios de color, por ejemplo, **HSI Lexicográfico** donde los vectores se representan como coordenadas del espacio HSI [10].
- **Distancia euclidiana(ED):** Esta estrategia de ordenamiento utiliza el cálculo de la distancia a un píxel de referencia, que por lo general es el origen, cuando es así la distancia es equivalente a la norma del vector $\|v\|$. Esta estrategia de ordenamiento tiene variantes con los espacios de color, cómo por ejemplo el espacio RGB y L*a*b* cómo se menciona en [10].
- **Bit mixing(BM):** Esta estrategia utiliza la técnica de reducción donde se define la transformación de cada vector como el valor entero obtenido de entrelazar los bits que representan a las intensidades de cada componente del vector [4].

Ordenamiento dependiente del histograma.

- **Ordenamiento utilizando información del histograma:** En el trabajo [15] se propone una estrategia de ordenamiento donde se realiza la reducción del color C mediante la función de transformación T donde se aplica el producto interno entre el color y un vector de pesos $w = (w_1, w_2, w_3)$ de la siguiente manera:

$$T(C) = \sum_{k=1}^{3} (w_k \cdot C_k) \tag{7}$$

  donde k es el índice del componente y $w_k \in \mathbb{R}$ es un peso extraído del componente k de la imagen. Los colores C_x y C_y, con $C_x \neq C_y$, pueden tener la misma transformación, es decir $T(C_x) = T(C_y)$, es por eso que la transformación $T(C)$ se utiliza como primer componente del orden lexicográfico.
  El método de ordenamiento se basa en el cálculo de pesos w el cual se realiza mediante una función ϕ aplicada sobre el histograma de la imagen en donde también se realizan subdivisiones de la imagen en regiones denominadas dominios o ventanas.

Calculo de Pesos: La función ϕ aplicada al histograma de cada componente k de la imagen f_v para el cálculo de los pesos w_k. Los pesos utilizados pueden ser, el promedio ($MEAN$) de niveles de intensidades de grises i, el mínimo (MIN)

que es el menor nivel de intensidad i, el máximo (MAX) que representa el mayor nivel de intensidad i, la moda mínima ($MO1$) que es el menor nivel de intensidad i que más se repite, la moda máxima ($MO2$) que es el mayor nivel de intensidad i que más se repite, la varianza (VAR) que representa la varianza de los niveles de intensidad i, todos estos para un dominio D determinado.

Los pesos obtenidos de las métricas mencionadas anteriormente se obtienen dividiendo la imagen en ventanas o dominios, de tal forma a obtener información local de una región determinada de la imagen.

División de la Imagen en Ventanas: La imagen f se divide en sub-regiones $W_1, W_2, \ldots, W_s$ como se menciona en [15]. Las sub-regiones W_s permiten definir áreas para obtener características locales de imagen. Sea B un elemento estructurante, el dominio D correspondiente al elemento estructurante B centrada en (r, c) se define cómo el conjunto de sub-regiones $W_{\{1,2,\ldots,s\}}$, que intersectan algún píxel de B. Cada dominio D se procesa de forma independiente para obtener las características particulares o locales asociada a cada dominio.

3 Propuesta.

3.1 Morfología matemática multiescala.

La morfología matemática multiescala es una variante de la morfología matemática en donde se utiliza un elemento estructurante escalable. Un elemento estructurante escalable [9] se obtiene mediante un elemento estructurante B y un valor entero n que representa el factor de escala del elemento estructurante. El valor nB se obtiene dilatando B recursivamente sobre si mismo $n-1$ veces, siempre y cuando B sea convexo.

Operaciones de la morfología matemática multiescala: Las operaciones morfológicas de erosión, dilatación y sus combinaciones apertura, clausura se utilizan con el elemento estructurante escalable para imágenes en escala de grises f. La transformada Top-hat multiescala [9] es una estrategia de mejora del contraste utilizando las operaciones de erosión, dilatación, apertura y clausura multiescala.

3.2 Transformada Top-hat multiescala para imágenes a color.

La estrategia del mejoramiento de imágenes a color propuesta incluye la utilización de la Transformada Top-hat multiescala para imágenes en escala de grises mencionada en [9] cuya aplicación práctica se extiende para las imágenes digitales a color. La estrategia de mejora del contraste para imágenes a color propuesta se representa de la siguiente manera:

$$\tilde{f}_v(r,c) = f_v(r,c) + 0.5 \sum_{n=1}^{m} F^{\circ}_{nB}(r,c) - 0.5 \sum_{n=1}^{m} F^{c}_{nB}(r,c), \tag{8}$$

donde, $\tilde{f}_v$ representa el resultado de mejora aplicada a la imagen a color f_v. Las características brillantes de la imagen se obtienen mediante la sumatoria

de la transformada white Top-hat $\sum_{n=1}^{m} F^{\circ}_{nB}(r,c)$ donde F°_{nB} representan las operaciones de apertura morfológica multiescala. Las características oscuras de la imagen se obtienen mediante la sumatoria de la transformada black Top-hat $\sum_{n=1}^{m} F^{c}_{nB}(r,c)$, donde F^{c}_{nB} representan las operaciones de clausura morfológica multiescala donde las operaciones se realizan n veces con un elemento estructurante B escalable. La constante 0.5 evita que se desborden los niveles de grises de los pixeles de la imagen.

4 Resultados Experimentales.

4.1 Métricas de evaluación.

Las métricas utilizadas para validar el método de mejora de la imagen utilizando morfología matemática multiescala para imágenes a color son la Varianza (VAR), la relación de mejora del contraste (CIR) que es una media de mejora de contraste para una región determinada de la imagen [13] y la métrica CEF Color Enhancement Factor que mide la percepción de la calidad del color de una imagen [7].

4.2 Resultados

Como referencia, para el ordenamiento dependiente de la información del histograma se agrega un sufijo WX, donde X representa el número de ventanas en la que la imagen fue divida. Si $X = 0$, el dominio es el elemento estructurante B, para $X = 1$, el dominio es toda la imagen, $X = 9$ la imagen fue divida en 9 sub-regiones de tamaño 3×3 y así sucesivamente.

Experimento I - Mejora del contraste global (VAR): En el primer experimento se evalúa el estiramiento de contraste global de la imagen a color y se muestran los mejores resultados, en promedio, en cada una de las iteraciones y para todos los ordenamientos mencionados anteriormente.

En la Tabla 1 se muestran los 10 mejores resultados en promedio correspondientes a las 100 imágenes de la métrica varianza, en donde se observa que el ordenamiento **BM** tiene mejor varianza en todas las iteraciones m, luego siguen los ordenamientos **MIN**, **MO1** y **MO2**.

Métodos	m= 2	m= 3	m= 4	m= 5	m= 6	m= 7
BM	**3316,576**	**3864,611**	**4665,250**	**5697,019**	**6806,257**	**7904,776**
MINW49	3258,434	3796,188	4578,353	5587,750	6678,743	**7759,869**
MINW25	3258,442	3796,102	4578,500	5587,919	**6678,947**	**7759,401**
MINW0	**3258,952**	**3797,176**	**4579,518**	**5588,219**	6678,795	7757,443
MINW1	**3258,627**	**3797,238**	**4579,773**	**5588,741**	**6678,826**	7755,024
MINW9	3258,531	3796,315	4578,349	5587,758	6675,713	7754,924
MO1W9	3259,042	3797,097	4577,706	5583,205	6667,705	7738,447
MO2W9	3259,058	3797,106	4577,740	5583,222	6667,388	7738,182
MO2W1	3258,906	3796,771	4577,421	5581,397	6667,439	7737,405
MO1W1	3258,901	3796,770	4577,417	5581,367	6667,399	7737,398

Tabla. 1. Top 10 Promedio de mejores resultados. Métrica **VAR**.

En general todos los ordenamientos generan buenos resultados en promedio utilizando la métrica **VAR**. El ordenamiento es importante en todos los casos, así como el tipo de configuración de ventanas que se utilizan para los métodos de ordenamiento dependientes del histograma. En este experimento el método **BM** genera mejores resultados en promedio para todas las iteraciones, posteriormente le siguen los ordenamientos que dependen del histograma, donde el dominio del elemento estructurante resultó mejor que la división por ventanas. En la Figura 1 se observa la imagen original de la base de datos [2].



Figura. 1. Imagen original tomada de la base [2]

En la Figura 2 se observa el histograma de la imagen original y el histograma en la iteración m igual a 5.

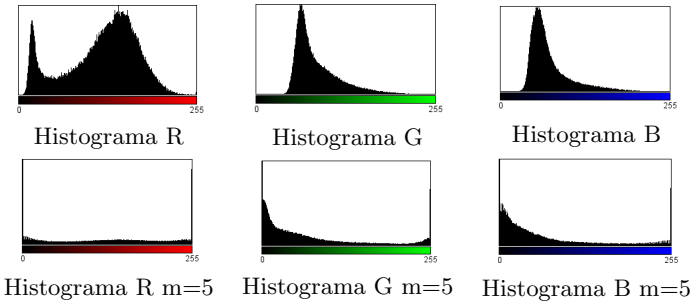


Figura. 2. Imágenes RGB para cada valor de m

Experimento II - Mejora del contraste local (CIR): En este experimento se evalúa el estiramiento de contraste local de la imagen donde se muestran los mejores resultados en promedio para cada una de las iteraciones y para cada tipo de ordenamiento.

En la Tabla 2 se puede observar que el ordenamiento dependiente del histograma aparecen entre los diez primeros resultados, los pesos son MEAN, MO1, MO2 y MAX.

Experimento III - Mejora del color (CEF): En el último experimento se evalúa la mejora del color de las imágenes donde se muestran los mejores resultados para cada iteración y para cada ordenamiento.

En la Tabla 3 se muestran los diez mejores resultados de la métrica **CEF**, en donde se observa que los métodos de ordenamiento dependientes de la infor-

Métodos	*m=2*	*m=3*	*m=4*	*m=5*	*m=6*	*m=7*
MEANW0	0,6148	3,2564	8,9259	17,7897	28,2444	**37,8354**
MO1W0	**0,6171**	**3,2668**	**8,9516**	**17,8376**	**28,2935**	**37,8182**
MEANW49	0,6146	3,2559	8,9183	17,7590	28,1968	**37,8176**
MO2W0	**0,6173**	**3,2665**	8,9496	**17,8276**	**28,2870**	37,8156
MEANW9	0,6146	3,2566	8,9204	17,7630	28,1920	37,8087
MEANW25	0,6147	3,2565	8,9207	17,7612	28,1963	37,8080
MAXW0	0,6157	3,2579	8,9221	17,7478	28,1703	37,8031
MEANW1	0,6146	3,2582	8,9264	17,7728	28,2002	37,7993
MO1W25	**0,6169**	**3,2692**	**8,9566**	17,8201	**28,2535**	37,7957
MO1W49	0,6168	3,2680	**8,9524**	**17,8211**	28,2504	37,7956

Tabla. 2. Top 10 mejores resultados CIR.

mación del histograma aparecen entre los diez mejores resultados en todas las iteraciones en las distintas configuraciones.

Métodos	*m=2*	*m=3*	*m=4*	*m=5*	*m=6*	*m=7*
MINW1	**1,013045**	**1,051822**	**1,124872**	**1,231890**	**1,368238**	**1,529108**
MO1W1	**1,009614**	**1,040297**	**1,105122**	**1,206343**	**1,341457**	**1,504726**
MO2W1	**1,009619**	**1,040295**	**1,105109**	**1,206350**	**1,341461**	**1,504693**
MO1W9	1,009181	1,039318	1,102214	1,202283	1,334507	1,499447
MO2W9	1,009194	1,039327	1,102255	1,202257	1,334250	1,499198
MINW9	1,009470	1,040141	1,104012	1,203391	1,334754	1,496082
MO1W25	1,009264	1,039142	1,101910	1,201898	1,333040	1,494616
MO2W25	1,009269	1,039118	1,101847	1,201633	1,332722	1,494025
MINW25	1,008607	1,037187	1,099015	1,196997	1,328103	1,490025
MINW49	1,008336	1,036289	1,096778	1,194335	1,325822	1,489206

Tabla. 3. Top 10 Mejores Resultados CEF.

En este experimento donde utilizamos la métrica **CEF**, que determina la mejora en la percepción de la calidad del color, se verifica como el ordenamiento dependiente del histograma con los distintos pesos tienen mejores resultados que los independientes de histogramas en todas las iteraciones. La mejora del colorido de la imagen se da a partir de la iteración $m \geq 3$ y el dominio de ventanas así como los ordenamientos son importantes para obtener buenos resultados a nivel del colorido de la imagen.

En la Figura 3 se observa una imagen con las mejoras a medida que crece m para el ordenamiento BM, donde se ve que para $m = \{2, 3, 4\}$ las mejoras son mínimas pero a partir de la iteración $m \geq 5$ las mejoras son mas notorias.

5 Conclusiones.

El trabajo presenta una extensión del método de estiramiento de contraste multi-escala en escala de grises propuesto por [9] a imagen a color, en donde utilizamos distintos tipos de ordenamientos para las operaciones morfológicas propuestas.

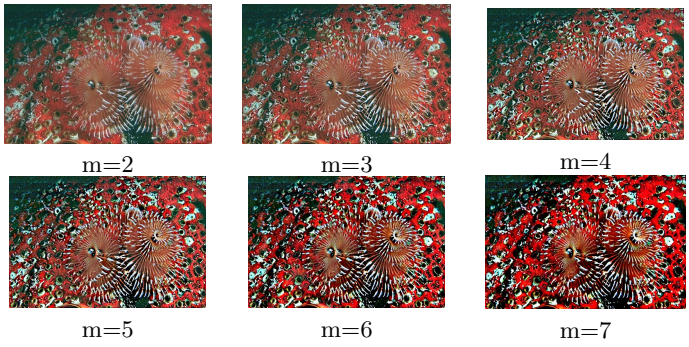


Figura. 3. Imágenes RGB para cada valor de m

En el análisis de mejora del contraste global todos los métodos implementados dieron buenos resultados en promedio, específicamente el ordenamiento **BM** dio mejores resultados en cuestiones de mejora de contraste global medidos por la varianza.

En el análisis del estiramiento del contraste local de la imagen se verifica, que en promedio el ordenamiento dependiente de la información del histograma logra mejores resultados que los que no dependen de la información del histograma. El peso MO1W0 es el que presenta mejores resultados en promedio en todas las iteraciones. Otro punto importante es que la información estadística del histograma de la imagen es un factor importante en el análisis, ya que para cada iteración existe una mejora local del contraste diferente en cada uno de los pesos.

La mejora del color de la imagen, utilizando ordenamientos dependientes del histograma de la imagen con los distintos pesos, presentan mejores resultados en comparación con los ordenamientos independientes al histograma. Verificamos que para cada iteración las mejoras se hacen cada vez que mas notorias, donde los distintos pesos y el dominio juegan un papel importante para la mejora.

Referencias.

[1] Erchan Aptoula and Sébastien Lefevre. “A comparative study on multivariate mathematical morphology”. In: *Pattern Recognition* 40.11 (2007), pp. 2914–2929.

[2] Pablo Arbelaez, Charless Fowlkes, and David Martin. “The berkeley segmentation dataset and benchmark”. In: *see http://www. eecs. berkeley. edu/Research/Projects/CS/vision/bsds* (2007).

[3] Xiangzhi Bai, Fugen Zhou, and Bindang Xue. “Image enhancement using multi scale image features extracted by top-hat transform”. In: *Optics & Laser Technology* 44.2 (2012), pp. 328–336. DOI: 10.1016/j.optlastec.2011.07.009.

[4] J Chanussot and P Lambert. "Bit mixing paradigm for multivalued morphological filters". In: *IEE conference publication*. Vol. 2. Institution of Electrical Engineers. 1997, pp. 804–808.

[5] Jocelyn Chanussot and Patrick Lambert. "Total ordering based on space filling curves for multivalued morphology". In: *COMPUTATIONAL IMAGING AND VISION* 12 (1998), pp. 51–58.

[6] Allan Hanbury, Umasankar Kandaswamy, and Donald A Adjeroh. "Illumination-invariant morphological texture classification". In: *Mathematical Morphology: 40 Years On*. Springer, 2005, pp. 377–386.

[7] David Hasler and Sabine Süsstrunk. "Measuring colourfulness in natural images". In: *Proc. IST/SPIE Electronic Imaging 2003: Human Vision and Electronic Imaging VIII*. Vol. 5007. LCAV-CONF-2003-019. 2003, pp. 87–95.

[8] George H Joblove and Donald Greenberg. "Color spaces for computer graphics". In: *ACM siggraph computer graphics*. Vol. 12. 3. ACM. 1978, pp. 20–25.

[9] S Mukhopadhyay and B Chanda. "Local Contrast Enhancement of Grayscale Images using Multiscale Morphology". In: *Proc. ICVGIP-2000* (2000), pp. 17–24.

[10] Francisco Gabriel Ortiz Zamora. *Procesamiento morfológico de imágenes en color: aplicación a la reconstrucción geodésica*. 2002.

[11] Julio César Mello Román, Horacio Legal Ayala, and José Luis Vázquez Noguera. "Image color contrast enhancement using multiscale morphology". In: *4th Conference of Computational Interdisciplinary Science*. (2016). URL: http://www.epacis.net/ccis2016/papers/paper_101.pdf.

[12] Tomislav Stojić, Irini Reljin, and Branimir Reljin. "Local contrast enhancement in digital mammography by using mathematical morphology". In: *International symposium on signals, circuits and systems (ISSCS)*. Vol. 2. 2005, pp. 609–612.

[13] Sabine Süsstrunk and Stefan Winkler. "Color image quality on the internet". In: *Proc. IST/SPIE Electronic Imaging 2004: Internet Imaging V*. Vol. 5304. LCAV-CONF-2004-026. 2004, pp. 118–131.

[14] Maria C Tobar et al. "Mathematical morphology in the hsi colour space". In: *Iberian Conference on Pattern Recognition and Image Analysis*. Springer. 2007, pp. 467–474.

[15] Jose Luis Vazquez Noguera et al. "A color morphological ordering method based on additive and subtractive spaces". In: *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE. 2014, pp. 674–678.

Calibración de un Algoritmo de Detección de Anomalías Marítimas basado en la Fusión de Datos Satelitales

Mirko Panozzo Zénere, Leandro Ramos, Bruno Marengo, Ricardo Medel, Franco Riberi

Ascentio Technologies S.A.,
Av. Rafael Núñez 5220, Córdoba, Argentina
{mpanozzo,lramos,bmarengo,rmedel}@ascentio.com.ar
fgriberi@gmail.com
https://www.ascentio.com.ar/

Abstract. La fusión de diferentes fuentes de datos aporta una ayuda significativa en el proceso de toma de decisiones. El presente artículo describe el desarrollo de una plataforma que permite detectar anomalías marítimas por medio de la fusión de datos del Sistema de Información Automática (AIS) para seguimiento de buques y de imágenes satelitales de Radares de Apertura Sintética (SAR). Estas anomalías son presentadas al operador como un conjunto de detecciones que requieren ser monitoreadas para descubrir su naturaleza. El proceso de detección se lleva adelante primero identificando objetos dentro de las imágenes SAR a través de la aplicación de algoritmos CFAR, y luego correlacionando los objetos detectados con los datos reportados mediante el sistema AIS. En este trabajo reportamos las pruebas realizadas con diferentes configuraciones de los parámetros para los algoritmos de detección y asociación, analizamos la respuesta de la plataforma y reportamos la combinación de parámetros que reporta mejores resultados para las imágenes utilizadas. Este es un primer paso en nuestro objetivo futuro de desarrollar un sistema que ajuste los parámetros en forma dinámica dependiendo de las imágenes disponibles.

Keywords: Automatic Identification System (AIS), Synthetic Aperture Radar (SAR), Análisis de imágenes, Teledetección

1 Introducción

El transporte marítimo representa aproximadamente el 90% del comercio mundial [1], por lo que comprender el tráfico marítimo es un aspecto clave para países con extensas costas, recursos naturales marítimos o cerca de rutas marítimas muy transitadas. Para estos países, entre los cuales se encuentra Argentina, es de particular importancia monitorear ciertas actividades, tales como la pesca ilegal y la navegación en zonas restringidas.

En la actualidad, el Sistema de Identificación Automática (AIS, del inglés *Automatic Identification System*) [2, 3] facilita la recolección de información sobre

el movimiento de buques, tanto en las zonas costeras como en aguas abiertas. El sistema AIS es obligatorio para buques de viajes internacionales con más de 300 toneladas brutas, buques que no se encuentran en aguas internacionales con 500 o más toneladas y buques de pasajeros [4]. Sin embargo, su correcto funcionamiento está sujeto al comportamiento de la tripulación de los buques monitoreados, ya que estos podrían manipular o apagar sus dispositivos. Por otro lado, debido al aumento del transporte marítimo, la cantidad de información suministrada por el sistema AIS es cada vez más abrumadora para los operadores. Es por esto que, a fin de sacar mejor provecho de los datos provistos por este sistema, es necesario comenzar a automatizar el procesamiento de dichos datos, obteniendo información valiosa a fin de identificar comportamientos de interés y proveer a los responsables de la toma de decisiones con información fiable y a tiempo.

Por otra parte, a diferencia de las imágenes satelitales ópticas, las imágenes bidimensionales de Radares de Apertura Sintética (SAR, del inglés *Synthetic Aperture Radar*) [5, 6] permiten observar todo el espectro marino independientemente de las condiciones meteorológicas imperantes en el momento y captar tanto pequeños como grandes buques. La alta resolución y amplia cobertura de las imágenes SAR se convierten en una gran ventaja, en particular considerando la amplia extensión de la costa Argentina.

Combinar datos AIS con detecciones SAR permite asociar imágenes de buques con su información AIS y detectar naves que no estén transmitiendo dicha información. La fusión de datos SAR-AIS se basa principalmente en tres etapas: (1) detección de probables buques en una imagen SAR, (2) proyección de los puntos AIS (tomando los más cercanos al tiempo de adquisición de la imagen SAR) sobre dicha imagen y (3) asociación entre las detecciones realizadas y los puntos proyectados. En caso de que una detección no corresponda a los puntos AIS proyectados, se está frente a una *anomalía marítima*, la cual deberá ser reportada al operador a fin de que sea investigada para detectar su naturaleza y medir el riesgo que representa.

En la primera etapa de este proceso, es decir, en la detección de probables buques sobre imágenes SAR, se aplica el método de Tasa Constante de Falsas Alarmas (CFAR, del inglés *Constant False Alarm Rate*) [7, 8]. Este es un método adaptativo que permite detectar blancos en presencia de ruido e interferencias. Para ello se desliza una ventana configurable sobre la imagen píxel por píxel y, analizando la intensidad de los píxeles vecinos, se identifica si el píxel en cuestión es un posible candidato.

En la segunda etapa se realizan estimaciones lineales sobre los puntos AIS que han sido adquiridos casi al mismo tiempo que las imágenes SAR (denominados *puntos AIS a tiempo SAR*). Dado que es difícil que la adquisición de la imagen SAR corresponda temporalmente con los puntos AIS adquiridos, lo que se hace es tomar los puntos AIS inmediatamente anteriores e inmediatamente posteriores al momento de adquisición de la imagen SAR. De esta forma se tiene un punto AIS de inicio y un punto AIS de fin para cada buque emitiendo señal AIS. Es sobre este segmento que se debe correlacionar los buques detectados en las imágenes SAR, lo cual corresponde a la tercera etapa [9].

En el presente trabajo presentamos un sistema automático para detección de anomalías marítimas sobre un área de interés (ROI, del inglés *Region of Interest*) empleando imágenes SAR y datos AIS. Se empleó el método CFAR en dos dimensiones, determinando a través de varias pruebas manuales la mejor configuración de parámetros para una ROI específica. Además de las proyecciones de los puntos AIS a tiempo SAR, se realizó un filtrado por distancia, para luego, a través de un algoritmo de vecinos más cercanos (NN, del inglés *Nearest Neighbour*) con validación cruzada (SAR-AIS, AIS-SAR), establecer las asociaciones correspondientes. Esto permite la detección de posibles anomalías, como por ejemplo la detección de un buque que no es informado por AIS pero es detectado usando SAR. La Sección 2 provee una descripción del funcionamiento del sistema desarrollado. Los resultados obtenidos en primera instancia se muestran en la Sección 3. Los experimentos realizados para determinar la mejor configuración y el análisis de sus resultados se presentan en la Sección 4. Finalmente, planteamos las conclusiones y trabajos futuros en la Sección 5.

2 Metodología

El sistema implementado responde a una arquitectura cliente-servidor, por medio de la cual el usuario puede monitorear zonas marítimas mediante una interfaz web propia o a través de la integración con sus propios sistemas de información geográfica (GIS, del inglés *Geographical Information System*) [10]. Los principales módulos que componen el sistema son los siguientes.

1. Detección de buques sobre imágenes SAR.
2. Administración y preprocesamiento de los datos provenientes de AIS.
3. Reproyección de los datos AIS y su asociación con buques detectados.
4. Identificación y notificación de las anomalías encontradas.
5. Administración y orquestación del sistema.

Además de la obtención automática de imágenes SAR, la detección de buques requiere el encadenamiento de las siguientes etapas: lectura e interpretación, proyección, máscara de tierra, detección de objetivos (CFAR), filtrado de objetivos por tamaño y generación de salida (archivo en formato *shapefile*).

El algoritmo CFAR permite la detección de objetivos comparando la intensidad de la señal SAR con cierto umbral. Es por esto que la eficacia de la detección estará sujeta a la elección del umbral adecuado. Si dicho umbral es bajo, entonces más objetivos serán detectados, incrementando la posibilidad de falsos positivos. Por el contrario, si el umbral es alto habrá pocos falsos positivos pero se pone en riesgo la detección de objetivos reales. En síntesis, el umbral debe fijarse de modo de permitir cierto porcentaje aceptable de falsas alarmas, sin perder demasiada precisión en la detección de verdaderos objetivos.

Si el entorno que rodea a un objetivo es constante en función tiempo-espacio, entonces es posible elegir un umbral fijo para alcanzar la probabilidad de falsas alarmas deseada. Sin embargo, las diferentes fuentes de interferencias ocasionan cambios en el nivel de ruido, tanto temporal como espacialmente. Por lo cual,

para poder utilizar el sistema de monitoreo en múltiples contextos (diferentes ventanas de tiempo y ROIs), es necesario elevar o bajar el umbral de forma dinámica, a fin de mantener la probabilidad de falsas alarmas constante.

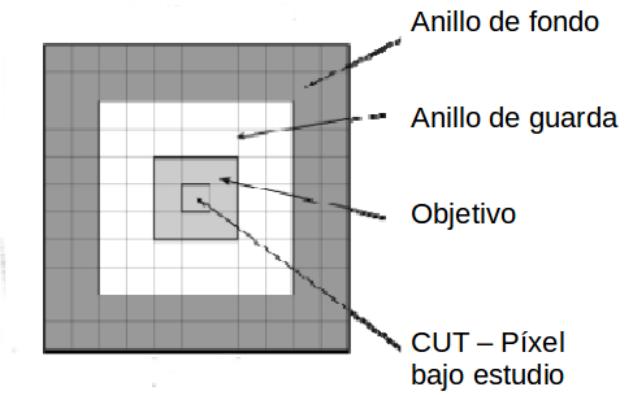


Fig. 1. Ventana deslizante usada en 2D CA-CFAR.

Para la detección de buques en imágenes SAR se aplicó el algoritmo CA-CFAR (del inglés *Cell-Averaging Constant False Alarm Rate*) [11, 12] en dos dimensiones, el cual consiste en deslizar una ventana alrededor de cada píxel bajo estudio o CUT (del inglés *Cell Under Test*), tal como se observa en la Figura 1. Este algoritmo permite adaptar el umbral de detección de un CUT mediante la estimación media de una ventana de N píxeles, teniendo en cuenta su entorno y sus píxeles inmediatamente adyacentes.

Los píxeles de la ventana referenciados como Anillo de Fondo (o *background ring*) en la Figura 1 permiten estimar un valor estadístico $\hat{I}$, que es multiplicado por una constante de ganancia G a fin de obtener el umbral de detección adaptativo $\lambda_{CA} = G * \hat{I}$. Por su parte, los píxeles referenciados como Anillo de Guarda (o *guard ring*) son descartados para evitar que objetivos extendidos interfieran con el valor estadístico $\hat{I}$. Así, si el valor de intensidad del CUT es mayor al umbral λ_{CA}, el mismo se considera un posible objetivo detectado o *target*.

Table 1. Configuración de Parámetros

Parámetros CFAR				
Tamaño Objetivo	**Tamaño Guarda**	**Tamaño Fondo**	**Tamaño Nave**	**Pfa**
3	15	31	15	6

Además de los parámetros mencionados, el algoritmo requiere también de un tamaño aproximado (en píxeles) de los buques a detectar y la probabilidad de fal-

sas alarmas (*Pfa*). Los valores propuestos en el trabajo de Ranjan y Moorth [11] son mostrados en la Tabla 1. En la Sección 3 mostramos que estos valores no producen el resultado ideal para la imagen del satélite Sentinel-1, correspondiente a la costa de Sudáfrica (ver Figura 2), con tiempo de adquisición en 2016-10-23 03:29:30 (UTC time). En la sección 4 se presentan los resultados de una serie de ejecuciones del algoritmo realizadas variando el valor de estos parámetros a fin de encontrar la mejor configuración de los mismos para esta imagen.

3 Resultados

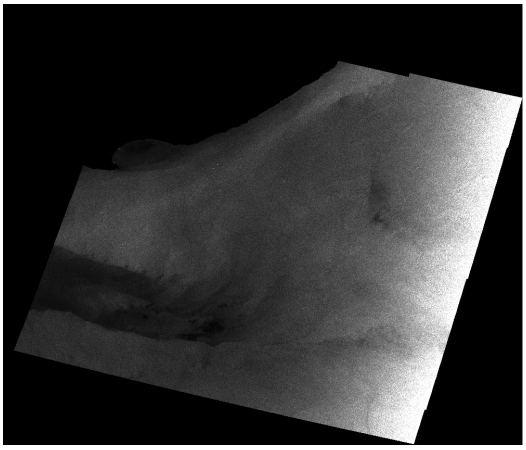


Fig. 2. Imagen SAR de la costa de Sudáfrica usada en los experimentos.

La Figura 2 muestra la imagen SAR utilizada durante los ensayos del sistema, ya con el enmascaramiento de tierra aplicado. Como se indicó previamente, la imagen corresponde a una porción de la costa de Sudáfrica y fue obtenida del catálogo abierto de la misión Sentinel-1 de la Agencia Espacial Europea (ESA).

Durante la operación regular, esta imagen es ingresada a nuestro sistema, el cual procede a reproyectarla, aplicarle la máscara de tierra y correr el algoritmo de detección CA-CFAR, obteniendo los polígonos de detección que indican la presencia de buques en la zona. En la Figura 3 se muestra un segmento de la imagen procesada, con dos buques detectados, marcados los píxeles que permitieron la detección.

Utilizando como entrada los objetivos detectados en el paso anterior y los puntos del sistema AIS para el tiempo de adquisición de la imagen, se realiza la correlación de estos datos. Se retorna información tanto sobre la correlación de puntos AIS con objetivos detectados en la imagen SAR como aquellos objetivos que no tienen correlación con datos AIS. La Figura 4 muestra el resultado

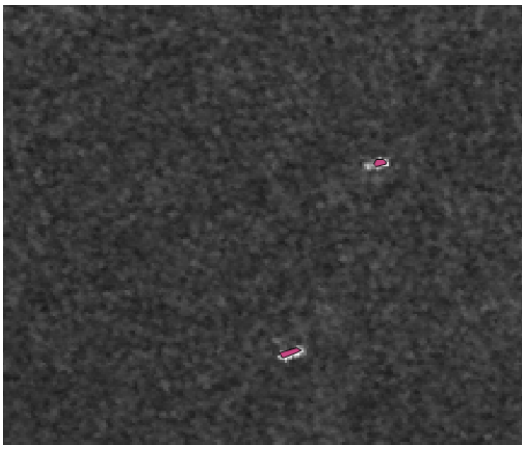


Fig. 3. Buques detectados en imagen SAR.

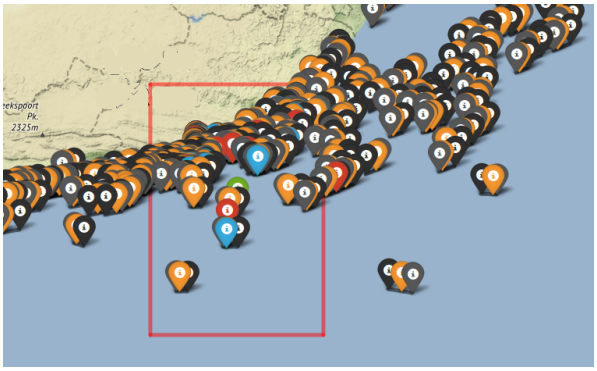


Fig. 4. Correlación entre puntos AIS y buques detectados.

visualmente a gran escala, tal como lo recibe un operador, mostrando el área procesada y los puntos detectados, correlacionados y no correlacionados.

La Figura 5 muestra un acercamiento sobre la Figura 4. Los íconos más oscuros, que indican el inicio y fin de cada segmento, identifican los puntos AIS tomados antes y después del momento de adquisición de la imagen SAR. Los íconos más claros, ubicados sobre el segmento y superponiéndose en diferente grado, representan la estimación de la posición del buque al tiempo exacto de la adquisición de la imagen SAR (calculada en base a los puntos AIS) y los objetivos detectados en la imagen SAR. Que ambos íconos se superpongan significa que el objetivo SAR y los puntos AIS cercanos han sido correlacionados. En el centro de la Figura puede verse un ícono oscuro (originalmente rojo en la imagen) no asociado a ningún segmento, lo cual indica que existe un objetivo detectado

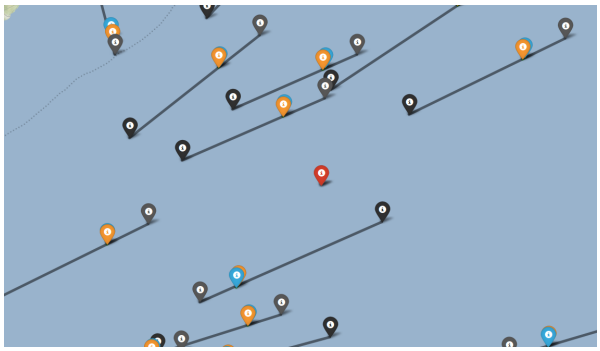


Fig. 5. Detalle de correlación entre puntos AIS y buques detectados.

sobre la imagen SAR que no tienen correlación con los datos reportados por el sistema AIS. Es decir, se ha detectado una anomalía.

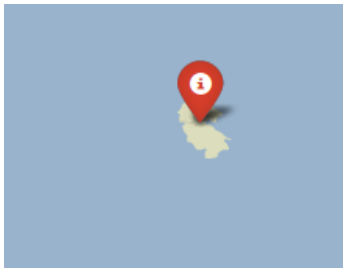


Fig. 6. Falso positivo.

La Figura 6 muestra, a través de un acercamiento mayor, que la anomalía detectada resulta, en este caso, ser una isla. Este falso positivo en particular no es falla del ajuste del algoritmo, sino que es una sección de tierra no enmascarada en el paso correspondiente del análisis de la imagen SAR.

Inicialmente ejecutamos el sistema utilizando los valores por defecto para los parámetros del algoritmo de CFAR, tal como se los define en el trabajo de Ranjan y Moorth [11]. Con dichos valores nuestro sistema detectó apenas 39 buques en la imagen SAR del área de interés, mientras que el sistema AIS registra las posiciones de 116 buques en la misma área en ese momento. A partir de la observación de esta discrepancia se plantearon los experimentos presentados en la próxima sección, con la finalidad de obtener una cantidad de detecciones sobre la imagen SAR que se aproxime a la cantidad de buques reportados por AIS y establecer las bases para el desarrollo de un algoritmo de ajuste dinámico.

Table 2. Resultados obtenidos

Parámetros CFAR					Resultados	
Tño Objetivo	Tño Guarda	Tño Fondo	Tño Nave	Pfa	Naves Detectadas	Posibles Objetivos
3	15	31	15	6	39	1
3	15	31	7	6	50	3
3	15	31	4	6	52	3
1	7	31	4	6	67	9
1	5	31	4	6	69	9
1	3	15	4	6	49	4
1	3	19	4	6	64	8
1	3	21	4	6	67	9
1	3	23	4	6	69	10
1	3	27	4	6	70	11
1	3	31	4	4	299	230
1	3	31	4	6	74	13
1	3	41	4	6	74	12

4 Discusión

Se ejecutaron 13 corridas del algoritmo modificando cada vez sus parámetros de entrada. En el Cuadro 2 se muestran los resultados de detecciones y de asociaciones para las distintas configuraciones de parámetros del algoritmo CFAR. En la sección "Resultados" de este cuadro, el término "Naves Detectadas" se refiere a la cantidad de buques detectados en la imagen SAR, mientras que "Posible Objetivo" son los puntos SAR que el algoritmo de correlación no pudo asociar con ningún punto AIS estimado en el tiempo de adquisición de la imagen SAR (116 para la imagen del ejemplo). La primera línea del cuadro se corresponde con los parámetros por defecto presentados en [11]. Nótese que en esa primera ejecución hay pocos buques detectados en relación a los puntos AIS disponibles, la correlación arroja sólo 1 posible candidato a barco que no registra su posición AIS (lo cual puede deberse a varios motivos: apagó su dispositivo AIS, no tiene AIS, etc.). Verificamos que este blanco detectado a simple vista parecería ser correcto. Para tener la certeza absoluta, habría que tomar una imagen óptica al momento de adquisición SAR en la supuesta posición del barco, y validar que efectivamente se trata de un barco y no ruido (podría ser basura flotando, una isla pequeña no marcada en el mapa, etc.).

En el mismo Cuadro observamos que conforme aumenta la diferencia entre el tamaño de pixel de guarda y el tamaño de pixel del fondo, aumenta considerablemente la cantidad de detecciones. El aumento en las detecciones trae aparejado también un aumento en la cantidad de candidatos a blancos. Para la imagen analizada, al variar el parámetro "Tño Nave" (tamaño de nave), el cual sirve como filtro de la salida del algoritmo de detección, se verifica una mayor cantidad de detecciones. Este fenómeno puede deberse a que con la resolución de esta imagen SAR, un tamaño de 15 pixeles para "Tño Nave" ocasiona un filtrado excesivo de buques considerados *pequeños*. La diferencia entre el tamaño de guarda y tamaño de fondo genera un mayor contraste y, dado que el algoritmo de CFAR funciona básicamente detectando la variación en la respuesta de los píxeles de guarda con el fondo, produce una mayor cantidad de detecciones.

En las imágenes de salida se pudo verificar que muchos de estos "buques" nuevos detectados corresponden a falsas alarmas. Ya vimos, por ejemplo, en la Figura 6 que un punto detectado como "buque" es una pequeña isla que la resolución de nuestro archivo DEM (del inglés *Digital Elevation Model* o Modelo Digital de Elevación) utilizado para máscara de tierra no permitió identificar como tal, por lo que no fue enmascarada y se utilizó en el procesamiento como si fuera parte del mar.

Es dable notar que la mayoría de los blancos nuevos detectados se identifican en zonas costeras. Sin embargo, estos blancos no resultan de interés en la aplicación de nuestro sistema, ya que la costa es una zona fácilmente monitoreable usando otros mecanismos, tales como buques de prefectura o pequeñas naves autónomas provistas de cámaras.

Un resultado interesante que arroja el ajuste de parámetros realizado es que al modificar el valor de la probabilidad de falsa alarma (Pfa) y por ende alejarlo del valor teórico obtenido por Ranjan y Moorthy [11], las detecciones obtenidas crecen a valores no razonables. Por ejemplo, en la fila 11 se reporta que al establecer este parámetro en un valor de 4 se realizan 299 detecciones de buques, mientras que se han registrado sólo 116 buques con posición AIS. Entonces, es claro que el algoritmo detecta "buques" en zonas donde la variación en la intesidad de los píxeles no hace suponer que los haya. La correlación con estos datos espúreos arroja 230 candidatos a blancos o anomalías.

Los valores de parámetros que permiten obtener mejores resultados corresponden a las dos últimas filas de la tabla. Ambas configuraciones detectan 74 candidatos. Comparando las detecciones de una configuración y otra con un tamaño de fondo de 41 píxeles, se pueden detectar buques que se corresponden con puntos AIS aún en las zonas luminosas de la imagen SAR que son escasamente perceptibles al ojo humano. Con el aumento del área a considerar en el contraste también aparecen nuevos "buques" que son falsas alarmas, mientras que desaperecen otras falsas alarmas que aparecian con tamaño de fondo 31.

Nuestra implementación del algoritmo CFAR está paralelizada, con 3 hilos de ejecución. Aunque no reportamos los tiempos de ejecución en este trabajo, cabe mencionar que cuando el tamaño del fondo crece, el tiempo de ejecución del algoritmo también crece, pero en el peor de los casos el tiempo de ejecución no supera los 10 minutos para esta imagen.

5 Conclusiones y Trabajo Futuro

El trabajo aquí presentado demuestra el potencial que presenta la fusión de imágenes satelitales con datos del sistema AIS para el monitoreo de grandes áreas costeras o mar abierto. En particular, los resultados enunciados demuestran que las imágenes SAR son de especial utilidad, ya que con una única adquisición pueden ser cubiertas grandes extensiones con resoluciones lo suficientemente precisas como para detectar blancos de interés para el proyecto, siendo estas detecciones independientes de la hora del día y la cobertura nubosa del área.

Mediante los experimentos realizados con distintos valores para los parámetros del algoritmo, concluimos que el valor de la probabilidad de falsa alarma no debería ser modificado del valor teórico ($Pfa = 6$) y que los valores óptimos para la imagen con la que se trabajó son: *Target Size* $= 1$, *Guard Size* $= 3$, *Backgorund Size* $= 41$ y *Vessel Size* $= 4$. Es de destacar que el procedimiento realizado para ajustar los parámetros para esta imagen puede automatizarse para mejorar el procesamiento de otras imágenes. Lo cual nos acerca a nuestro objetivo de desarrollar un sistema que ajuste los valores de los parámetros en forma dinámica.

Las mejoras futuras a ser implementadas incluyen la capacidad de procesamiento de imágenes de múltiples proveedores (sean SAR o de otro tipo), mejorar los algoritmos de estimación de datos AIS a partir de datos históricos, e implementar algoritmos de optimización de parámetros de detección y correlación en función de la imagen a procesar. El siguiente paso inmediato es realizar experimentos similares de ajuste de parámetros aplicados a otras imágenes, de manera tal de poder compararlos con los obtenidos en este trabajo.

References

1. G. Pallotta, M. Vespe, K. Bryan, *Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction*, Entropy, 15, 2013.
2. IMO, *Guidelines for the Onboard Operational Use of Shipborne Automatic Identification Systems (AIS)*, Resolution A.917(22), International Maritime Organization, November 2001.
3. ITU, *Technical Characteristics for an Automatic Identification System Using Time-Division Multiple Access in the VHF Maritime Mobile Band (ITU)*, Recommendation ITU-R M.1371-5, International Telecommunication Union, February 2014.
4. IMO, *Safety of Life at Sea (SOLAS) Convention Chapter V*, International Maritime Organization, July 2002.
5. Jackson, C. R. and Apel, *Synthetic Aperture Radar Marine User's Manual, Chapter 1. Principles of Synthetic Aperture Radar*, National Oceanic and Atmospheric Adminstration, U.S. Department of Commerce, Washington DC, September 2004.
6. I. Cumming, F Wong, *Digital Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*, Artech House, Boston, MA, USA, 2005.
7. D.J Crisp, *The State-of-the-Art in Ship Detection in Synthetic Aperture Radar Imagery*, DSTO-RR-0272, 2004.
8. D. Pastina, F. Pico, P. Lombardo, *Detection of Ship Targets in COSMO-SkyMed SAR Images*, University of Rome, 2011.
9. F. Mazzarella, M. Vespe, C. Santamaria, *SAR Ship Detection and Self-Reporting Data Fusion Based on Traffic Knowledge*, IEEE Geoscience and Remote Sensing Letters, Vol. 12, August 2015.
10. A. Busccella, A. Cechich, P.R. Fillottrani, *Integración de sistemas de información geográfica*, IX Workshop de Investigadores en Ciencias de la Computación, pp. 397-401, Trelew, Chubut, Mayo de 2007.
11. D. Ranjan, H. Moorthy, *Development of Adaptive Algorithm for CFAR in non-homogenous environment*, ISO 9001:2008, July 2013.
12. Behar.V, *CA CFAR radar signal detection in pulse Jamming*, Compt. Rend. Acad. Bulg. Sci., vol. 49 (12), pp. 57-60. 1996.

Recognition of Surface Irregularities on Roads: a machine learning approach on 3D models

Rodrigo Huincalef[1], Guillermo Urrutia[1],
Gabriel Ingravallo[1], and Diego C. Martínez[2]

[1] Departamento de Informática Trelew, Facultad de Ingeniería,
Universidad Nacional de la Patagonia San Juan Bosco.
[2] Instituto de Ciencias e Ingeniería de la Computación, CONICET-UNS

Abstract. Roads are composed of various sorts of materials and with the constant use they expose different kinds of cracks or potholes. The aim of the current research is to present a novel automated classification method to be applied on these faults, which can be located on rigid pavement type. In order to collect proper representation of faults, a Kinect device was used, leading to three-dimensional point cloud structures. Images descriptors were used in order to establish the type of pothole and to get information regarding fault dimensions.

1 Introduction

The pavement of a street or highway is a structure composed of a set of layers of materials processed on the ground, whose function is to distribute the load of vehicles to the subsoil and allow the constant transit. The structure of the pavement should provide a surface of acceptable quality for the circulation of vehicles, adequate slip resistance, a lower level of noise, a waterproof, structural strength and a long life cycle with low maintenance cost. However, cracks are usual due to several factors. A *fissure* is a long, narrow opening in a slab of material that can be a corner fissure, longitudinal (if it extends along a slab) or transversal (if it extends perpendicular to the overturning of the slab material). The repair method for this type of failure consists of sealing joints and fissures, and repairing the entire thickness. On the other hand, a *pothole* is defined as a cavity, generally rounded due to the loss of the pavement on a part of the surface. The repair method for this type of failure depends on its deterioration, and is special for each case. In Figure 1 both faults are shown.



Fig. 1. A pothole and a longitudinal fissure, common in roads.

Due to the constant degradation of the different types of faults, these must be detected and repaired as soon as possible. The task of registering defects and depressions in the pavement has always been crucial to adopt a precise strategy for the maintenance and repair of roadways. However, manual measurement is a costly task both in time and resources. In this work we present an automated, AI-based classification of pavement faults, mainly potholes and fissures, through machine learning techniques using proper, accessible 3D models. There are previous works that apply different technologies for this purpose [2–12] but they require special devices and a good degree of important human intervention. Our interest is to formalize a process that is both accessible and cost-efficient. As a consequence, we implement an assistive application for the detection and classification of faults in the pavement, which will be useful for cost estimates. Our proposal make use of accessible domestic technologies, such as the 3D sensor of videogame consoles, to obtain a model of pavement fault that us appropriate for machine learning algorithms. A real implementation that offers an appropriate visualization and automatic geo-location of faults is also presented.

The work is organized as follows. Section 2 introduces the problem of road faults modelization. Section 2 explains the 3D sensors, the libraries to process the data obtained by these sensors, the management of point clouds, the descriptors that can be obtained from the analysis of these data and the machine learning technique used to generate the classification model. In section 3 we describe a prototype application for a vehicle that is used to collect the data through the Kinect sensor. Finally, conclusions and future work are discussed.

2 Surface Recognition

As stated before, we are interested in the automated classification of faults. This requires the use of adequate data structures representing real potholes and fissures. Hence, surface recognition is the first part of the process towards intelligent evaluation of faults. Ideally, the task of collecting data from real pavement should be also be automated, by using an autonomous vehicle. However, the construction of such a kind of vehicle is beyond the goals of our project. We are interested in the use of accessible hardware towards a practical solution for city governments and contractors.

In order to classify depressions then it is mandatory to get a proper mathematical model using some surface measure device. Here 3D sensors are used to analyse real world objects or environments and obtain their relevant physical properties, such as colors or shape, which later can be used to produce three-dimensional digital models. There are diverse, expensive sensors that may be used to attain this purpose, but we are interested in the use of domestic, accessible resources. Hence, we use a Kinect Microsoft sensor that has the capabability of generate Range Images, that keeps information regarding distance of each image pixel to the device capture point. We have used this device to record faults in a moving vehicle. The device should be mounted as shown in Figure 2, although that structure was not built during this research proyect.

The scanning process must produce 3D model representations. We use *point cloud structures*, which are point sets on a coordinate systems (optionally with RGB point information), being the most frecuently used the cartesian three-dimensional system (X,Y,Z). These representations can be renderized, inspectioned and converted to polygonal or triangular meshes.

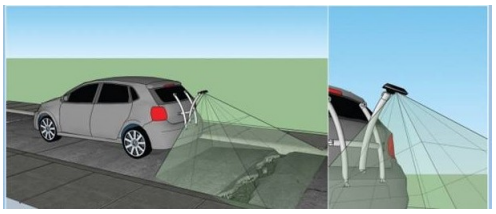


Fig. 2. The Kinect device can be easily installed on a vehicle.

There are several drivers and libraries (OpenNI, Freenect, OpenKinect,Point Cloud Library) that allow users to interact with the Microsoft Kinect sensor. Point Cloud Library (PCL) is an independent, open-source, multiplatform (also available in Windows, Linux, MacOS and AndroidOS), C++-written solution for sensing, geometrical point cloud processing and storing in 2D or 3D dimensions. The library PCL offers different standalone modules with algorithms that may be combined into a pipeline to spot several types of objects. These algorithms are meant to be applied on a wide range of tasks which is important for a correct object detection, for instance outliers filtering (filter point with values out of a certainrange), point cloud reading, storing, format conversion, decomposition (in order to perform searches) and concatenation, perform segmentation on specific parts of a complete point cloud capture, keypoint extraction and geometric descriptors computation with the aim of identifying distinct sorts of objects. In Figure 3 a point cloud of a real pavement fault is shown.

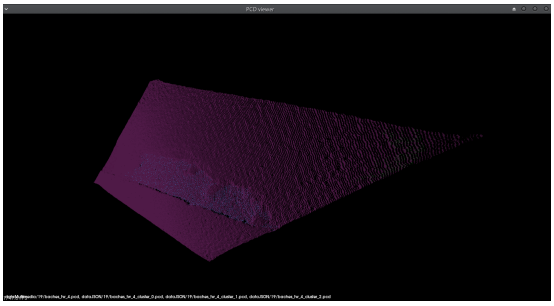


Fig. 3. Point cloud visualized with pcl_viewer tool from PCL. Full point cloud points are depicted with violet whereas points from a pavement depression, which have been isolated with classifier application, are with ligthblue and blue.

The process of obtaining a proper 3D model for automated classification is important and somehow difficult. Regarding the PCL pipeline for object recognition, it consists of a point cloud pre-processing phase with the goal of getting rid of noise. Then an object segmentation step is performed which allow us to obtain clusters which can be associated with potholes. Lasty, a descriptor generation phase is executed with the goal of getting information about the object geometry. PCL descriptors are divided in two classes: local and global. On the one hand, local descriptors describe surface geometry around a point without taking into account the complete geometry of the object the point belongs to. Hence when the local descriptors computation is performed, it is necessary to previously filter keypoints which belongs the object under study. Local descriptors applications include objects recognition and registration, which is a technique to detect whether common areas in several point clouds exists. On the other hand, global descriptors describe geometry of a whole point cluster which represents an object. Therefore if there is a need to generate this type of descriptor it would be compulsory to perform previously a pre-processing step in order to isolate the cluster from the point cloud. Applications of this type of descriptors are the object recognition and classification, position estimation and geometry surface analysis (type of object, shape, etc.).

The process of collecting samples took several days. Also, during this phase it was observed that some cracks does not possess significant depth to be measured by the Kinect device. Althoug this seems to be a drawback, those cracks are not of real interest for evaluation costs of repair (at least not for the moment, because faults tend to increase in size as time goes by). Due to this fact we decided to classify only those types of cracks which have enough depth to be isolated and described by descriptors which make use the of geometrical information regarding angles between normals of a surface. Because of this, descriptors were selected taking into account processing hardware capabilities and normal associated descriptor properties. These are the following:

- Global Radious-based Surface Descriptor (GRSD): This descriptor does not use any sort of pre-processing, for instance normals, yet it consists in traversing each of the point cloud points and for each select three random points and compute three ratio functions: D2, D2 ratio, D3 and A3. Every function produces histograms which describe the geometric relation among points that are part of the main figure.
- Ensamble Shape of Functions (ESF): Global descriptor that is based on Radio-based Surface Descriptor (RSD), which consists in geometrical surface shape by calculating radial information through neighbour points. Therefore this algorithm consist in setting up a formula using normals angle relation λ, the distance between them d and the surface radious r defined in $d = r * \alpha$. The latter formula is applied to every point p, that belongs to the point cloud, what gives as a result a set of ratios which describes each of the spheres that contains p with every one of the neighbours. Finally, only the minimal and maximal ratios are selected to be included in the final descriptor for that point.

The outcome of these complex processes is a data structure properly representing relevant pavement faults. There is, however, the need of determining whether these faults are either potholes or fissures, as they require different treatments. In the following section we discuss the use of machine learning techniques to help this task.

3 SVM

Support Vector Machine is a technique used for supervised learning to solve classification and regression problems based on hyperplanes. In this method the p features associated with a certain amount of training samples are arranged in a *p-dimensional* (being each sample a p dimensional vector) where each feature represents the coordinates with regard to one specific sample. In this way the SVM model generation is based on computing, testing and selecting from several hyperplanes which separate input data in order to get the best. Thus, SVMs allow researchers to divide data in sets with linear hyperplanes. Nonetheless, there are situations in which input data cannot be linearly divided, causing an unsatisfactory performance. The solution offered by SVMs to this problem is the use of kernel decision functions which convert the feature input data space to a higher dimension space being these quadratic, cubic, polynomial or higher dimensions space with the goal of achieving the finding of a brand-new hyperplane that separate samples with a better precision. Therefore there are a diverse range of kernels employed to attain this objective, mainly Linear, RBF and Polynomial.

As a result of SVM being a classification binary algorithm, various techniques have been developed to attain multiclass classification which allow researchers to assign samples among two or more classes. Two of these methods are:

- One versus One (One-vs-One, OvO): Given N classes this algorithm perform the $N(N-1)/2$ binary classifiers training getting each for every class combination using the training dataset. Then during the prediction phase a voting scheme where every classifier gives a result for the same input data sample. Hence the class which has a greater amount of positive votes it is the result of the prediction process.
- One versus Rest (One-vs-Rest, OvR, OvA): Given N classes N classifiers are trained using the full training dataset, choosing as positive those classes that belong to the classificator whilst the others are chosen as negatives. Accordingly, by receiving an input sample every classifier generate a decimal value which is a confidence score that indicates the probability of that sample of belonging to that type of element. Consequently the class with the highest score is considered the class in which the element is assigned.

In our current research project, an SVM implementation using C++ programming language was implemented for constructing the classification model (converting the data point cloud samples into SVM compatible format). Throughout this process, it was necessary to assign a class to every type of sample (crack

or pothole) using OvO technique and testing the same dataset with Linear and RBF kernel.

3.1 Building a classification model

The process of construction of the classification model consisted in the following steps:

1. Pavement surface samples collection: During this step an estimated amount of 1000 samples were captured using the Kinect sensor, which sensed different pavement streets in the city of Trelew.
2. Samples pre-processing phase: The main target of this phase is the preparation of the raw samples in order to improve quality. Several PCL algorithms for discarding noise were studied. Consequently they were tested against several samples. Next, the amount of point cloud points was reduced because of the hardware limitations and high samples quantity. Thus, *downsampling* techniques were implemented. Finally, it was necessary to obtain geometric properties in order to describe with higher precision the sensed pavement surface. Here, PCL normal computation algorithms were employed.
3. Sample segmentation: Algorithms for segmentation PCL were analyzed employing the samples previously captured with the aim of isolating point cloud clusters which belong to the depressions on the pavement and the rest of the street. As a consequence, Planar and Euclidean Segmentation algorithms were selected and concatenated in the instructions pipeline.
4. Curvatures computing: A deeper comprehension of fault dimension is obtained. The average curvature for clusters isolated earlier, which are thought to be part of a pavement depression, were computed so that allowed us to get curvatures minimum and maximum ranges, which let us get rid of those segments whose average curvature level was out of these ranges. Therefore this helped to improve the classifier application ability to filter or keep clusters based on alike geometry surface depressions.
5. Feature building: Lastly, diverse PCL surface description methods were studied and compared with regard to the size in bytes for each sample and amount of samples to be processed, histogram shape associated with different types of samples and sort of descriptor (local or global). Later it was investigated whether PCL offered machine learning models support combined with descriptors. Due to the fact the library had machine learning SVM support (making use of libsvm tool) this method was chosen for the classification application and each one of the point cloud samples were converted to libsvm compatible format.

After faults were detected by the sensor and raw data was obtained, a complex process of data preparation is triggered. Although it was an integral part of the project, it is not, however, the intention of this paper to describe the process in detail, other than the stages mentioned above. In the following section we address the automatic classification of faults.

4 Automatic classification

The first step toward the construction of the model was to assign 76 % of samples to training partition leaving the remaining for testing partition. Previously mentioned descriptors were tested in several experiments with various types of samples and as its precision was not enough it was decided to add depressions properties (for example length, width, depth and volume) along with the descriptors. Potholes properties were computed by PCL algorithms which are based on the minimum and maximum Cartesian axis values (AABB, OBB y "boundary cloud"). Thus statistical size depression values were compared as well as arithmetical relationships among them (averages, arithmetical difference between size and average sizes) with the goal of finding a way of make a difference among different types of depressions.

By observing these values it was seen that the width-height difference in the training dataset was similar for the same type of sample. Then a re-classification sample step was performed. This consisted in establishing a limit value to distinguish cracks from potholes being the samples that have a $|length - width| > 0.49$ difference value considered as cracks whereas others were considered as potholes, since cracks have considerably higher length than width. Thus the previous formula utilized these values for computing the difference using PCL Oriented Bounding Box (OBB) algorithm on X and Y axis.

Having observed the classification precision increased by reclassifying sample training dataset the same technique was applied to the testing dataset which allowed us to filter 806 out of 1000 samples (753 samples for training and 53 for testing). Afterwards new tests were performed using the (now) reclassified samples, obtaining 89 % accuracy with Linear kernel and 71 % with RBF (gamma 0.0000002 and cost C 1500) employing a five iteration cross validation approximation with GRSD descriptor. Next the length-width difference was re-tested, but with ESF descriptor. Using the same parameters, the process obtained a 98 % precision score for Linear kernel and 54 % for RBF kernel.

A comparison of classification metrics concerning the selected descriptors (using the original dataset partition with 50 samples) was made, with the goal of verifying the classification performance of ESF against others. These values can be found in Figure 4 and 5.

In the following section we describe the use of a vehicle in the collecting samples stage, as shown in Figure 2, and the embedded system for registering faults and their position by geo-localization.

	Kernel Linear			Kernel RBF		
Tipo de muestra	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Baches	1.0	1.0	1.0	0.0	0.0	0.0
Grietas	1.0	1.0	1.0	0.17	1.0	0.29
avg/total	1.0	1.0	1.0	0.03	0.17	0.05

Fig. 4. ESF descriptor metrics with Linear and RBF kernel.

	Kernel Linear			Kernel RBF		
Tipo de muestra	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Baches	0.83	1	0.91	0.00	0.00	0.00
Grietas	0.23	0.78	0.36	0.17	1.00	0.29
avg/total	0.80	0.53	0.58	0.03	0.17	0.05

Fig. 5. GRSD descriptor metrics with Linear and RBF kernel.

5 Application for a vehicle prototype

Ideally the sensing prototype device consists of a vehicle with a shell fixed in its back part, as it is seen in Figure 2, which allows the operators to hold the device from an appropriate high and take samples directly from above the pavement surface. The sensor is connected to an electrical supply device as well as a notebook in which the pavement depression measuring application is running. This application is set up to work with a GPS device that provides the fault location (latitude and longitude). Thus the pothole measuring process workflow is to stop the vehicle when a pavement depression is spotted, then the application user performs the measuring of it and the car moves to the following one and so forth. This process allows the sensing and registration of big areas of a city in acceptable time. The application main architecture is composed of the following software modules:

- Kinect device: It is the main sensing device which let the application get video and depth frames, which are compulsory to the point cloud file building. The frames are continuously read, rendered and visualized in real time from the depression capture application screen.
- Geofencing module: Its purpose it is to compute and return GPS coordinates in which a pavement depression is.
- APIClient: This component contains the main data exchange class between capture and web application.
- Client capture application: It offers the functionality with respect to capturing depression process and sending to the web application.

5.1 Pothole classification application

The pothole classification application is developed in C++ and its operation consists in reading configuration parameters from a .json file which holds information regarding the built model and also a database file that keeps a previous processed samples log.

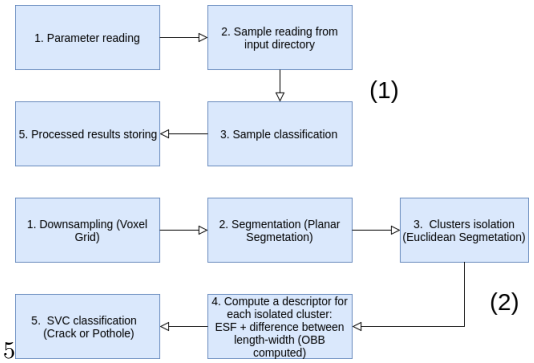


Fig. 6. (1) Classification application (2) Cropping pipeline.

Thus the general workflow starts reading input point cloud files (from a earlier setted up directory) and applying the sample cropping pipeline in order to get possible pavement depressions clusters. Then for each one build the custom descriptor and use the previously trained machine learning model, which is read from hard disk drive, to get the pavement depression type. As a result, every isolated cluster are stored in point cloud format in an output directory together with their computed width, length and depth. This process is pictured in Figure 6. In Figure 5.1 the general architecture is described, and in Figure 5.1 a screenshot of the application is shown.

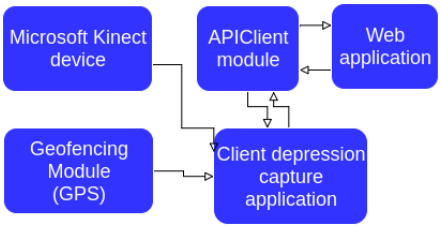


Fig. 7. General software modules from fault-capturing application.

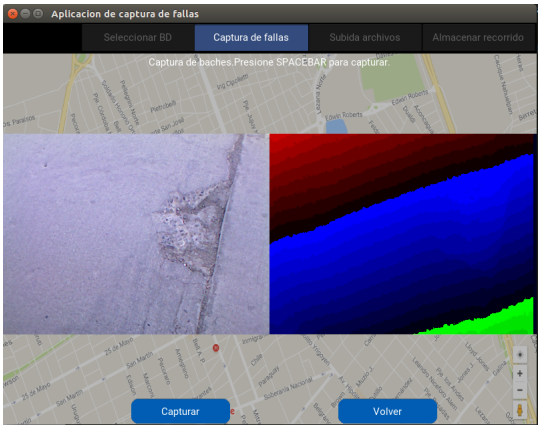


Fig. 8. Pavement depression measuring application interface (Left: RGB video view. Left: Nigth vision view.).

6 Conclusions and future work

In this work we have described a system for the collection and intelligent classification of potholes and cracks in pavement surfaces, using a low-cost commercial device along with the open-source PCL algorithms connected in an easily configurable pipeline. Moreover the estimated properties related to the pavement depression (width, length, depth), which may be useful for performing a material budget estimation in the repairing process, were successfully computed and stored for each isolated cluster. The system successfully classified collected data as potholes or cracks with a proper accuracy.

There exists other proposals on pavement depressions detection and analysis, which encompass researching and image processing with 3D videos and images, the following investigations have been made [2–6]. Alternatively, studies have been performed utilizing vibration devices such as accelerometers in [7–10]. Otherwise, studies in which the main focus is on specific road potholes features with aim of getting to process three-dimensional points coordinates as it is seen on [11, 12]. However none of them uses domestic devices nor classification algorithms based on machine learning. The use of automated learning is interesting since pavement faults tend to observe similar qualities within a city, as a consequence of climatic variations, used materials and quality of human work.

Future work has several directions. A continuous capture is desirable since it will make possible to avoid stopping the car. This requires a proper adaptation of dynamic gathering of data, since it demands real-time algorithms for point-cloud noise reduction. We are also interested in the estimation of materials and costs of repairing pavement faults. Since there is a 3D model of every fault a close,

reliable estimation can be produced. This can be done after processing a whole street or neighbourhood, hence helping to anticipate financial requirements and proper planning for e-governance. Even more, after long periods of time historic data can be mined to eventually detect areas that require special solutions other than normal patches to pavement. Finally, the addition of more than one type of sensor can be helpful to allow classification of faults which does not have enough depth to be fully captured by one sensing device.

References

1. Point Cloud Library (PCL), `https://www.pointclouds.org`
2. Korch y Brikalis: Pothole detection in asphalt pavement images. Advanced Engineering in Informatics (2011)
3. Buza, Omanovic, Huseinovic: Pothole detection with image and spectral clutering. Recent Adances in Computer Science and Networking (2013)
4. Seing-Ku Ryu , Taehyong Kim, Young-Ro Kim: Image-Based Pothole Detection System for ITS Service and Road Management System. Mathematical problems in Engineering (2015)
5. S Nieber , M Brooysen, R Kroon: Detection Potholes using Simple Processing Techniques and Real-World Footage. Proceedings of the 34th Southern African Transport Conference (2015)
6. Tomoyuki Yamaguchi , Shingo Nakamura, Ryo Saegusa, Shuji Hashimoto: Image-based crack detection for real concrete surfaces. Wiley InterScience (2008)
7. Eriksson, Girod, Hull, Newton, Madden, Balakrishnan: The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring. MobiSys (2008)
8. K. Chen, M. Lu, X. Fan: Road Condition monitoring using on-board three-axis acelerometer. 6th International ICST Conference on Communications and Networking in China (2011)
9. A. Tecimer, Z. Taysi, A. Yavuz, M. Karsligil Yavuz: Assessment of vehicular transportation quality via smartphones. Turkish Journal of Electrical Engineering and Computer Science (2013)
10. D. Casas-Avellaneda, J. López-Parra: Detection and localization of potholes in roadways using smartphones. DYNA Magazine (2016)
11. Sikai Xie: 3D pavement surface reconstruction and cracking recognition using Kinect-Based solution. The University of New Mexico (2015)
12. S Mathavan, M. Rahman, K. Kamal, S. Usman, I. Moazzam: Metrology and Visualization of Potholes using the Microsoft Kinect Sensor. 16th International IEEE Annual Conference on Intelligent Transportation Systems (2013)
13. H. Brink, J. W. Richards, M. Fetherolf: Real-World Machine Learning. Mannnig Publications (2017)
14. Scikit Learn (Support Vector Machines), `http://scikit-learn.org/stable/modules/svm.html`
15. Tom M. Mitchell: Machine Learning. McGraw-Hill (1997)

Video Surveillance for Road Traffic Monitoring

Guillermo Torres[1] ⋆, Ivan Caminal[2], and Cristina Maldonado[2] Marc Górriz[2]

[1] Universidad Nacional del Comahue, 8300 Neuquén, Argentina
[2] Universidad Politècnica de Catalunya, 08034 Barcelona, España
guille.torres,ivancaminal72,cmmsmcgma,algayon2}@gmail.com

Abstract. This work proposes a framework for road traffic surveillance using computer vision techniques. After a foreground estimation, post processing techniques are applied to the detected vehicles in motion to generate blobs. Then, a tracking approach based on Kalman filters is used to extract instantaneous information throughout a video sequence, including speed and trajectory estimation and imprudent driving detection. The system has been developed in Python and can be launched in real-time using a standard CPU. The code is available at github: https://github.com/mcv-m6-video/mcv-m6-2018-team3.

Keywords: traffic control · road monitoring · foreground segmentation · vehicle tracking · kalman filter · speed estimator.

1 Motivation

Full roads and congested traffic relates a common problem nowadays. Consequently, there is a need to develop intelligent traffic surveillance systems that play an important role in road monitoring. Its purpose is to provide numerical data in real time about traffic activity and to point out potentially anomaly situations, such as accident detection or dangerous driving. Video analytics is defined as surveillance based computer vision algorithms and systems to extract appropriate information from video. This can assist the detection of events of interest, which is useful for traffic management due to additional data available. Video cameras have been used for a long time for traffic monitoring purposes, because they provide a rich information for human understanding. Automatic preprocessing allows efficient control for the operators to process and collect statistics from live videos, with the aim to improve traffic flow.

The article starts describing the related work in section 2, followed by a detailed explanation of the methodology in section 3 and finally, the overall framework evaluation at section 4.

2 Related work

Detecting vehicles in images acquired from roads is a challenging problem and has become increasingly popular in the last years, replacing current approaches

⋆ This research is part of the project Altas Prestaciones Parte III (04/F012) of the Universidad Nacional del Comahue, Argentina.

such as RADAR or mobile surveillance configurations. The solutions have to deal several adverse situations and uncontrolled environment with a high variability (background, illumination), and in a reduced processing time.

Regarding the background subtraction, several approaches provides very good results, including frame differencing [5], gaussian mixtures [3], median filter [7] or non-parametric kernel-density estimation [16]. For the case of vehicle tracking, deep learning has overcome the state of the art of the methodology, both for prediction quality and computational efficiency. Some powerful approaches for detection include Yolo [10] and Faster R-CNN [11], while specifically for tracking, Zhang [14] proposed multi-layer occlusion detection and processing framework for the mutual occlusion between two vehicles and Faro [13] proposes a further improvement by introducing curvature scale space to segment occluded region accurately.

3 Methodology

In this section we will present the different datasets, methods, configurations and applications that enabled us to obtain the final pipeline, see figure 1.



Fig. 1: Pipeline followed to make our application.

3.1 Datasets

During this project we worked with one downloaded video [3], named "Custom" in this paper, and with three other datasets obtained from *ChangeDetection* [15] corresponding to the year 2014: Traffic, Fall and Highway. The Custom has no annotations while the ones from *ChangeDetection* have. Hence, we used the annotated ones to asses quantitatively the quality of the obtained masks in subsections 3.2 and 3.3. Regarding the results obtained in subsections 3.4 and 3.5 all datasets could be used because we only asses them qualitatively, however, only Traffic, Highway and Custom were finally used.

A split was done to the four sequences to obtain train and test sets, see Table 1.

[3] `https://www.youtube.com/watch?v=nt3D26lrkho`

3.2 Background subtraction

Background subtraction is the first step in order to detect moving objects in a sequence. For this task was tried three approaches: *Non adaptive*, *Adaptive* and *MOG*. After considering the results, that are presented in the section 4, the *Adaptive* one was selected for the final pipeline.

Non adaptive This approach, models the background of the sequence using a single Gaussian for each pixel in the train set. When the model is estimated, we threshold every pixel in the test set by (1) to obtain binary masks.

$$|I_i - \mu_i| \geq \alpha \cdot (\sigma_i + 2) \tag{1}$$

Where i is the pixel position, μ is the mean, σ is the standard deviation and α is a constant. The result is a binary mask where high pixels belong to foreground and low ones to background. The optimum value of α was found by maximizing f1-score metric.
Illumination changes and errors in the scene are the main limitations of this approach. Therefore, is only useful in high-supervised applications.

Adaptive This other approach, is equal to the previous but the parameters of the model evolve over time following (2) and (3) after each prediction.

$$\mu_i = \rho \cdot I_i + (1 - \rho) \cdot \mu_i \tag{2}$$

$${\sigma_i}^2 = \rho \cdot (I_i - \mu_i)^2 + (1 - \rho) \cdot {\sigma_i}^2 \tag{3}$$

Where ρ is a constant within [0,1] that works as a trade-off between the current and accumulated value of mean and variance. Grid search was used maximizing again f1-score metric to find the optimal combination of α and ρ parameters. Adaptive methods, are chosen for more applications due to the limitations of non-adaptive ones.

	Train	**Test**
Highway	1050-1200	1201-1350
Fall	1460-1510	1511-1560
Traffic	950-1000	1001-1050
Custom	1-359	1-359

Table 1: Split (in frames) of the used sequences to obtain test and train sets.

MOG MOG [4] is an improved adaptive Mixture Model (implemented in *OpenCV*). It uses a method to model each background pixel by a mixture of K Gaussian distributions (K = 3 to 5). The weights of the mixture represent the time proportions that those colors stay in the scene. The probable background colors are the ones which stay longer and more static.

3.3 Foreground improvement

In this part, we will try to improve the previous masks with some video preprocessing and mask post-processing.

As pre-processing, we used two stabilization methods only in the Traffic set (which had camera jitter). The first was a own implementation using block matching algorithm and the second was a web application [12].

For mask post-processing we tried shadow removal using an available *OpenCV* function called MOG2 [4] that detects shadows. Furthermore, some morphological operators were used in different combinations and structuring elements.

After applying different combinations of the mentioned processing methods, we discarded the ones that did not improve the results. The final improvement pipeline can be seen in Table 2.

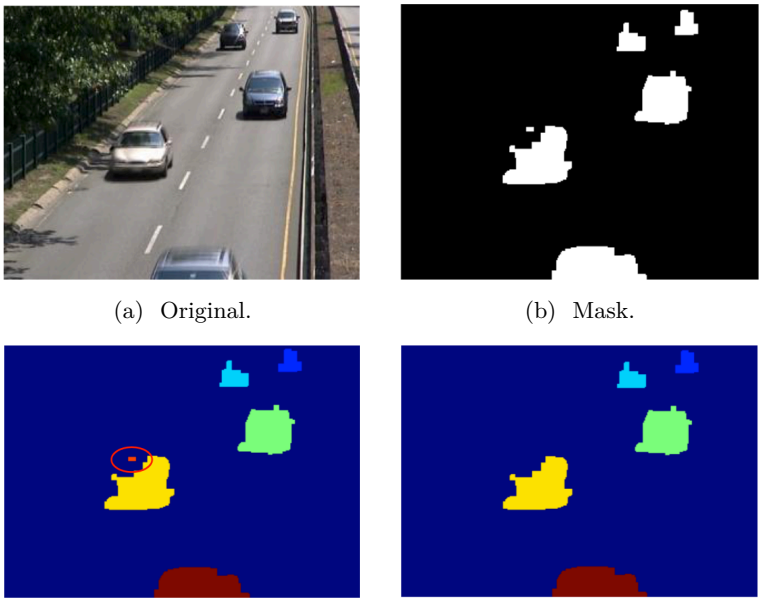
	Highway	Traffic	Custom
1-Opening	s.e. 4x4	s.e. 11x11	s.e. 3x3
2-Dilation	x4 iter.	x8 iter.	x3 iter.
3-Hole Filling	4 connect.	4 connect.	4 connect.
4-Erosion	x4 iter.	x8 iter.	x3 iter.

Table 2: Best improvements applied to the masks obtained in section 3.2 ordered up to down. The opening was done with a square structuring element (s.e.). The iterations (iter.) in dilation and erosion were done with a square 3x3 structuring element.

3.4 Object tracking

In order to identify a vehicle along the sequence and extract real time features we used different tracking algorithms: Kalman filter [1], Kernelized Correlation filter (KCF) [2], Median Flow filter [8] and Boosting filter [9]. We implemented the first approach while we took the *OpenCV*[4] implementations [cite] for the others. Once the foreground masks are generated, tracking is performed following the pipeline described in figure 2. First, connected components are extracted from

[4] `http://opencv.org`



(a) Original. (b) Mask.

(c) Connected comp. (d) Area filtering.

Fig. 2: Tracking preprocessing pipeline.

the masks. Then, area filtering is applied in order to remove the noisy objects. Finally, each foreground object is tracked throughout the sequence.

The filters are able to do tracking taking blobs as input and determining the next position of the objects using spatial measurements. After some test results (see section 4), Kalman filter seems to be the most accurate approach. Is robust against illumination changes in shadows and in difficult situations such as horizontal displacements when the vehicle changes the land.

3.5 Speed computation

For the speed estimation, in the first step we use a projective transformation to obtain a bird's-eye perspective and compute the new location of the car centroids. In this way, the 8-point algorithm [6] was implemented using only 4 points to obtain a fundamental matrix, which is computed only once. It is important to point out that the fundamental matrix is only applied to the position of the centroids, in contrast to be applied to the whole image, to save computation resources [5].

[5] Team 1 credits from last year.

In the second step, the speed computation takes advantage of the birds'-eye perspective and assume that the movement of the vehicles is only in y-direction, which is shown in the figure 3.

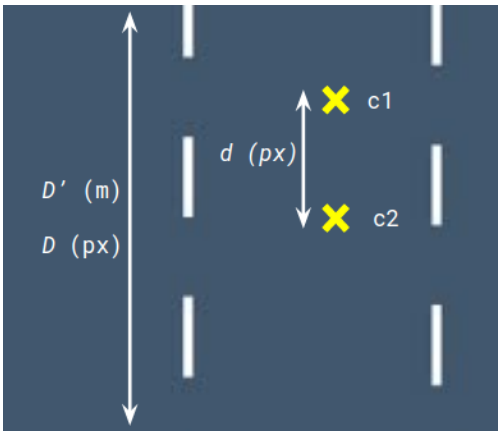


Fig. 3: Measures used to compute the speed estimation: D and D' represent the relationship between pixels and meters respectively, and d represents the movement of a centroid from the position $c1$ to $c2$

Thus, the speed is computed in (4). Where *fps* stands for frames per seconds and 8 denotes the frequency in which speed is updated. This frequency was arbitrary selected to decrease the speed variation of the vehicles, specially, when a new vehicle begins to be tracked.

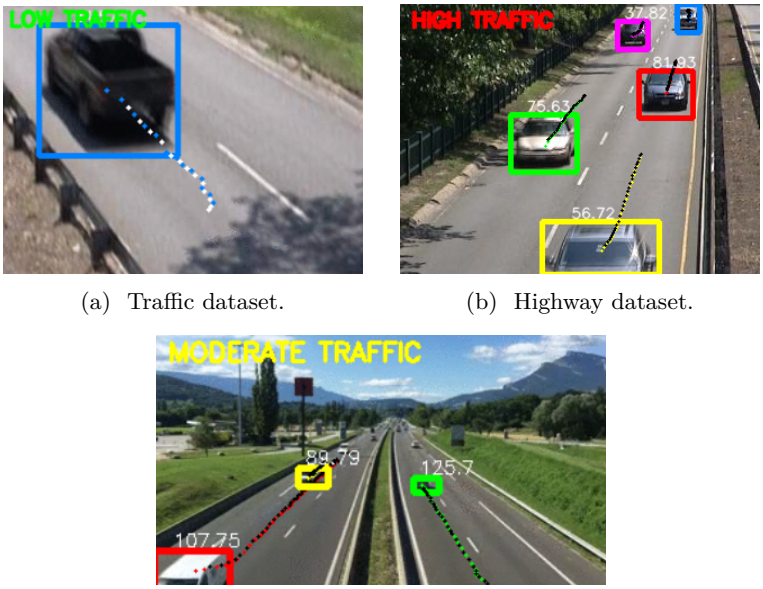
$$Speed = \frac{fps}{8} * \frac{D'}{D} * d \tag{4}$$

3.6 Application

With all the techniques described before, we build our Road Traffic Monitoring system. This application is the end part of the pipeline showed in the figure 1 and was built with the best methods and configurations evaluated in the previous sections.
For this reason, the application was configured to use the Kalman filter. The sets employed were Traffic, Highway and Custom, which are described in section 3.1.

This application does the following:

- density of the road: considering 1 or 2 vehicles as low traffic, 3 or 4 vehicles as moderate traffic, and 5 o more as high traffic;



(a) Traffic dataset. (b) Highway dataset.

(c) Custom dataset.

Fig. 4: The Road Monitoring Traffic in action using the Kalman filter in the (a) Traffic, (b) Highway, and (c) Custom datasets. The tracked vehicles are closed with a color bounding box and its speed is shown over it. Also the vehicles's trajectory are drawn with color for the centroids and with black/white for the filter predictions. And a label with low/moderate/high traffic is shown for the density of the road.

– detects when vehicles exceed speed limit; and
– obtain a trajectory estimation of the vehicles.

A picture of the application in action is shown in the figure 4.

4 Results

In this section, we present the quantitative results for the two first parts of the pipeline. The table 3 shows the best results obtained for each background subtraction. The adaptive approach get highest maximum F1-Score for the three datasets evaluated, and in second place it is followed by non adaptive. The particular parameters values used for the non adaptive and adaptive are listed in table 4.

Figure 5 shows the foreground improvement (green) reached by the adaptive approach in background subtraction (black) for the traffic and highway test sets. In the figure 5a along the recall axis, always it is obtained a precision

improvement. And the figure 5b shows that the precision get an improvement from 0.5 onwards in the recall axis. Below that value it works worst (red), which means that the approach has a negative impact in the foreground and in the forward pipeline steps it will worsen the vehicles detection.

Also, a qualitative results for the other parts can be seen in the figure 4.

	Non Adaptive	*Adaptive*	*MOG*
Highway	0.45	**0.73**	0.35
Fall	0.67	**0.71**	0.39
Traffic	0.48	**0.68**	0.57

Table 3: Maximum F1-Score background subtraction approaches from section 3.2

	Non Adaptive	*Adaptive*	
	α	α	ρ
Highway	1.9	2.89	0.21
Fall	4.5	3.2	0.05
Traffic	1.8	3.55	0.16

Table 4: Optimal parameters, sections 3.2 and 3.2

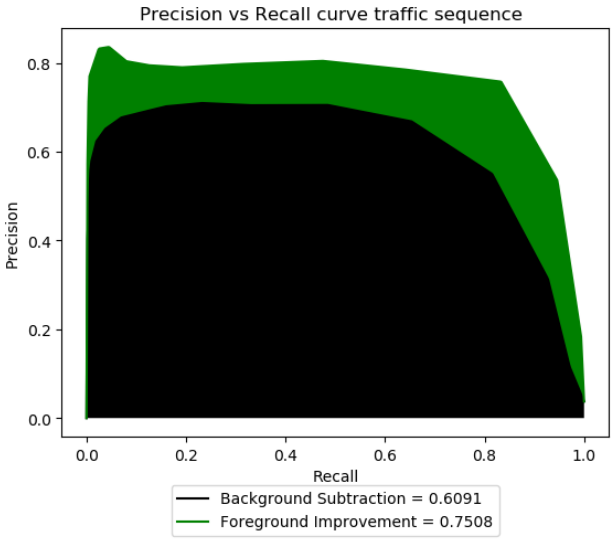
.

5 Conclusions

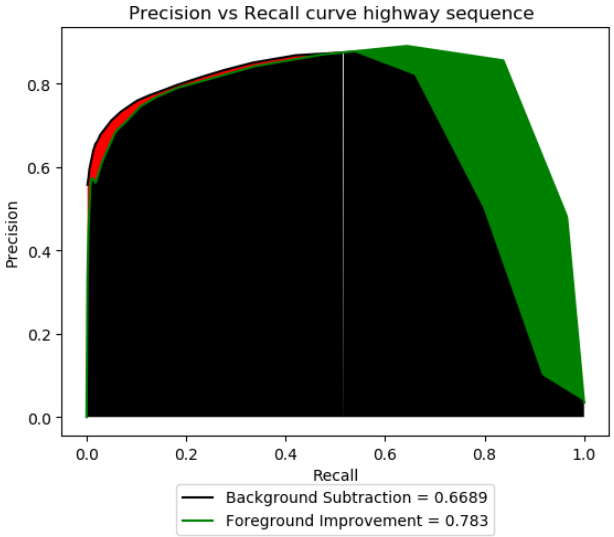
We developed a traffic monitoring system modularized by a pipeline to track cars, count them, estimate their trajectory and speed. We learned that errors in the first stages of the pipeline can produce bad results in subsequent stages. It is very important to have a robust foreground detection system in order to avoid errors in object tracking.

For object tracking we used different algorithms and we found that Kalman filter is the best method to track the cars in the video and gives the best qualitative results. With this technique there are less erroneous detections and the trajectories are estimated with more precision than other filters.

As future work, the application could incorporate a metric of imprudent driving taking into account the distance between vehicles, its trajectory and



(a) Traffic test set



(b) Highway test set

Fig. 5: Visual improvement of the precision and recall curves and area under the curve of two of the analyzed test sets in sections 3.2 and 3.3.

speed. Also is interesting to advance in direction of deep learning to track the vehicles using a neural network like Yolo [10] or other. A measurement of the

required time for each step and in the over all process is needed to know the viability of the techniques, because it is a real time application.

References

1. Bishop, G., Welch, G., et al.: An introduction to the kalman filter. Proc of SIGGRAPH, Course **8**(27599-23175), 41 (2001)
2. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence **37**(3), 583–596 (2015)
3. Jin, C., Zhang, Y., Balakrishnan, S., J. Wainwright, M., Jordan, M.: Local maxima in the likelihood of gaussian mixture models: Structural results and algorithmic consequences (09 2016)
4. Kaewtrakulpong, P., Bowden, R.: An improved adaptive background mixture model for realtime tracking with shadow detection (2001)
5. Kartika, I., Mohamed, S.S.: Frame differencing with post-processing techniques for moving object detection in outdoor environment. Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on (March 2011)
6. Longuet-Higgins, H.C.: A computer algorithm for reconstructing a scene from two projections. Nature **293**(5828), 133 (1981)
7. Mallikarjuna Rao, G., , C.: Object tracking system using approximate median filter, kalman filter and dynamic template matching **6** (04 2014)
8. Oflazer, K.: Design and implementation of a single-chip 1-d median filter. IEEE transactions on acoustics, speech, and signal processing **31**(5), 1164–1168 (1983)
9. Okuma, K., Taleghani, A., De Freitas, N., Little, J.J., Lowe, D.G.: A boosted particle filter: Multitarget detection and tracking. In: European conference on computer vision. pp. 28–39. Springer (2004)
10. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection. ArXiv e-prints (Jun 2015)
11. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp. 91–99 (2015)
12. Unknown: Video Stabilize, `https://video-stabilize.com/`
13. W. Zhang, Q.W.e.a.: Adaptive background modeling integrated with luminosity sensors and occlusion processing for reliable vehicle detection. IEEE Transactions on Intelligent Transportation Systems **12**(4), 1398–141 (Dec 2011)
14. W. Zhang, Q.W.e.a.: Multilevel framework to detect and handle vehicle occlusion. IEEE Transactions on Intelligent Transportation Systems **9**(1), 161–174 (Feb 2008)
15. Wang, Y., Jodoin, P.M., Porikli, F., Konrad, J., Benezeth, Y., Ishwar, P.: CDnet 2014: An Expanded Change Detection Benchmark Dataset `https://www.cv-foundation.org/openaccess/content{_}cvpr{_}workshops{_}2014/W12/papers/Wang{_}CDnet{_}2014{_}An{_}2014{_}CVPR{_}paper.pdf`
16. Zanin Zambom, A., Dias, R.: A Review of Kernel Density Estimation with Applications to Econometrics (Dec 2012)

Workshop Bases de Datos y Minería de Datos (WBDMD)

Coordinadores

Laura Lanzarini (UNLP) - Hugo Alfonso (UNLPam) - Nora Reyes (UNSL)

Temas

Modelo y Diseño de Datos
Verificación y Validación de Software
Bases de Datos Objeto
Bases de Datos Difusas
Bases de Datos Temporales, Espaciales y Espacio-Temporales
Bases de Datos Geográficas
Bases de Datos de Conocimiento
Bases de Datos Textuales y Multimedia
Bases de Datos en Web
Base de Datos y XML
Otro tipo de Bases de Datos
Data Warehouse
Data Mining
Datos: Procesamiento y Exploración
Clasificación y Análisis de Asociación
Clustering: Agrupamiento de datos
Detección de Anomalías
Aprendizaje Automático

Data Sources Integrator for Management by Indicators

Gustavo Illescas, Iris Figini, Mariano Martinez.

Research Institute in Advanced Information Technology, Computer Management- Centre Associate CIC Group. Faculty of Exact Sciences. National University in the centre of Buenos Aires Province. Paraje Arroyo Seco, Tandil (7000), Argentina.
illescas@exa.unicen.edu.ar, imfigini@slab.exa.unicen.edu.ar, mmartinez@slab.exa.unicen.edu.ar.

Abstract. This current work exposes the research and development of a tool for the integration of data sources in the context of management by indicators. The main benefits of this research are the administrators of databases that will work with a powerful tool that will allow them to effectively manage the updating of indicators previously defined by information managers. These information managers are the most interested in that the indicators are updated in the databases for the taking of decisions which will directly benefit them.
This work gives rise to a thesis in Systems Engineering and the result could be applied in the TIC direction of the Faculty of Exact Sciences as it is expressed in the case study, taking part of the Strategic Plan of Engineers Formation promoted by the Secretary of University Policies.

Keywords: Database, ETL process, Indicators, Management Control.

1 Introduction

Frequently the managerial sectors of organizations diagnose the need to improve some aspect of their management, which establishes a possible hypothesis that should be able to validate and execute the strategies and policies defined. To perform this task formally, indicators and targets will be settled which will allow a measure of expected improvement.

In this sense, we can define an indicator as a compact description of a particular observation. These may relate to a specific theme or can be summarized as a certain number of linked issues [1]. More specifically, an indicator is a variable or factor that provides a simple and reliable means to measure achievement, reflect changes associated with an intervention or help you evaluate the results in an organization. It is a sign, signal, or value that allows establishing differences, behaviors, and trends, its measurement can be quantitative or qualitative, and in a certain period of time.

The results obtained by the measurement allow improved planning, since it is possible to observe events in real time, making decisions with greater certainty and reliability.

To collaborate in this regard, a prototype was developed within the framework of a thesis in systems engineering (thesis 1)[1] called "Tool to Support the Management by Indicators" [2], which manages the creation and modification of indicators (Figure 1a) and present the results through control panels (Figure 1b) by registering them at an own database, the database of indicators (DBI).



Figure 1 a - Display of an indicator. **Figure 1 b** - Control panel.
Figure 1 – Display of an indicator and control panel. The data shown are simulated [2].

This DBI does not have any sense on its own, but it is "fed" by the systems of organization that are in their transactional databases (TDB) as it is from the primary process as well as another sources, for example, from external organisms, credit bank entities, organisms of state control, and others.

As an example, we can mention some management systems (accountants, ERP-Enterprise Resource Planning, CRM-Customer Relation Management, and so on) and as a particular case in our home study the Guarani, Kolla, Araucano, Kune, Majen, among other systems.

Due to the complexity of the integration processes from different data sources, the necessity of counting with a tool arouses which will allow implementing an orderly sequence of activities over these processes.

In [3] it is defined the lifecycle of indicators by business intelligence tools implementation:

a) definition and computerization of indicators
b) conformation of specific DBI from the available databases
c) informatization of indicators in intelligent tools of businesses: Balanced SocoreCard- BSC [4], Control Tables [1] On-line Analytical Processing (OLAP)
d) implementation of tools of control and follow-up, construction of usage profiles and responsibility
e) maintenance of indicators, add/remove/modify to selected tools, monitor the use and
f) the decisions of the organization related to the function provided by the tools.

[1] Thesis in preparation: Federico Etchepare and Agustín Servat under direction of Dr. Gustavo Illescas

In Figure 2, we can appreciate a simplified scheme of the mentioned cycle that allows a better visualization of the prototype to develop in which specifications are included [5].

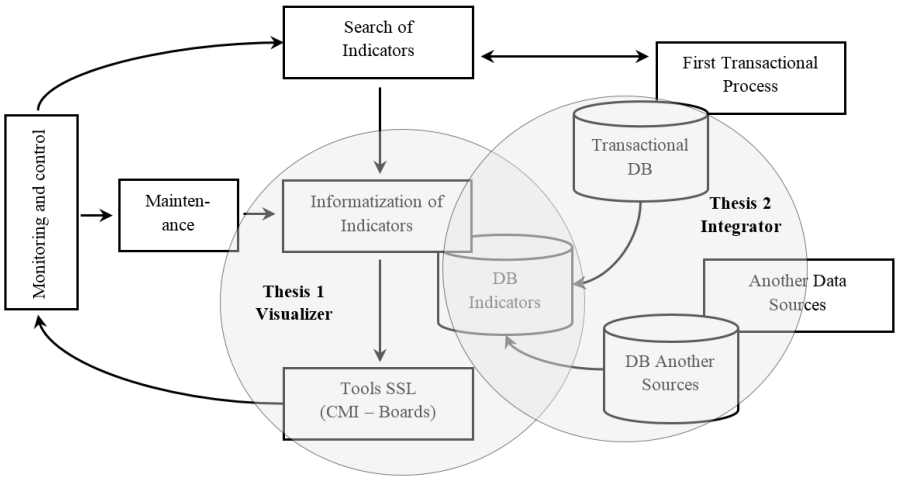


Figure 2 - Simplified scheme with the lifecycle indicators in business intelligent tools [3].

In order to arrive at an integral solution, it is proposed to add an integrative dispositive (thesis 2) to allow the actualization of values of the indicators defined in the DBI. The dispositive will be in charge of providing standard formats to the previous data validation so that the values can be added for update or insertion. For data actualization of an indicator, the information to incorporate is based on the date and the value of the indicator. From the TDB or the DB of other sources can come data from the indicator of different periods. On the other hand, we proceed to classify the states of registered values according to the already defined thresholds in DBI for a certain period.

Once validated the data and incorporated to the DBI, it proceeds to register the value of the last date of actualization to be used when the decision-maker visualizes the indicators.

To summarize, the objective of this work can be observed in the scheme presented in Figure 3.

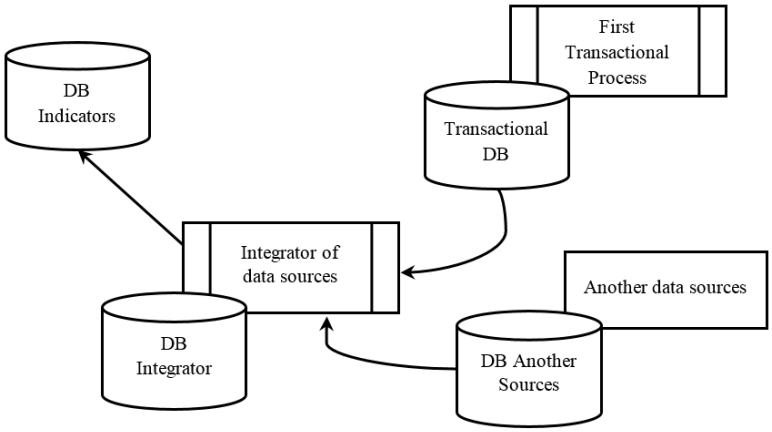


Figure 3 - Scheme of the design of the data sources integrator.

2 Development

The motivation of the current work is a proposal for future works of [3], together with the necessity and institutional interest of the Faculty of Exact Sciences that analyses great volumes of data coming from different sources.

Furthermore, the development of this tool is viewed in the Strategic Plan of Engineers Formation (PEFI – Plan Estratégico de Formación de Ingenieros), which deals with a plan of improving academic indicators for the formation of engineers. The PEFI is driven by the Secretariat of University Policies with the purpose of incrementing the number of engineer graduates to guarantee the quantity and quality of necessary human resources and to make Argentina a developed country [6].

The prototype presented in this work has the characteristics of what we know as the process of Extraction, Transformation, and Load (ETL) [7]. We find different tools as IBM Websphere DataStage [8], Pentaho Data Integration [9], Oracle warehouse Builder [10], Microsoft SQL Server Integration Services [11], among others but that analyzed in detail does not satisfy the specific needs that this work tries to accomplish.

Although there exists at the market many tools that visualize indicators, in our composition of the state-of-art we have not found a tool that allows the extraction of indicators from different and varied sources of data and the integration of them with a visualized solution.

In general, all of them require a manual process to extract the transactional registered information and the supply of such systems, besides, with the possibility that the visualized information will not be find synchronized in the management area, and with the inconvenient of manual tasks and direct manipulation over the data. Thus, it was raised the necessity to build a prototype that automates those tasks.

Then, the current work is based in developing a prototype that acts as an interface between the transactional systems in the organization and the tool of visualization of indicators (mentioned in 1), and through different functions, implement the programming of automated tasks of actualization of indicators (Figure 4).

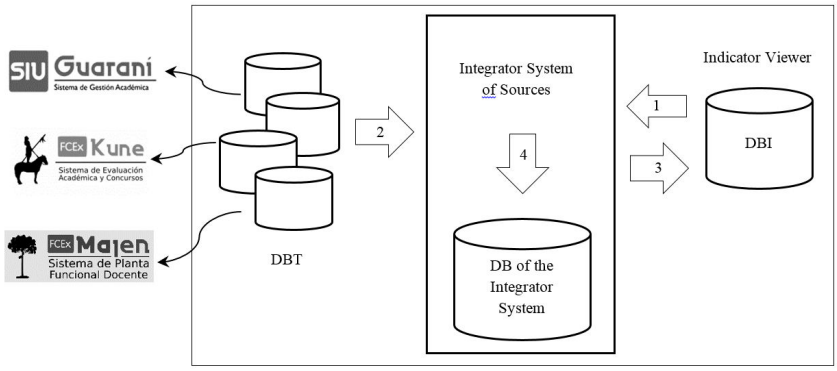


Figure 4 - Reference flow and actualizations to DBI

This prototype may be used in every organization that has multiple data sources and the necessity of evaluating the information from indicators. As a principal case of study, of application and demonstration, the current work is devoted to extract and integrate values for indicators of the different systems that are used in the Faculty of Exact Sciences of the UNCPBA as being the system of SIU-Guarani, of Teaching Plant (Majen) or systems where the teaching contests are managed (Kune).

That is to say, the case study will be first applied in an educative entity but it can be applied without any problems in every field that has recollected data through the database and has the interest to make an analysis to improve some kind of aspect in their organization or business.

The sequence or basic procedure for the use of the tool is shown in Figure 4 and it includes: 1) the lecture of the indicator to actualize from the DBI, 2) the configuration of the connection to the TDB where the values of the indicator are extracted (Figure 5) and the generation to a query to the TDB in an editor of source code (Figure 7) together with the programming of the calendar in the execution of the query, 3) the actualization of the DBI with the obtained values and 4) the register for the audit of the result of the execution in the DB of the integrator.

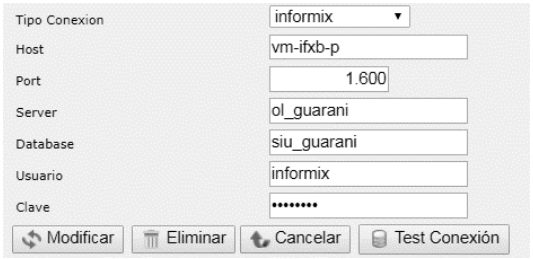
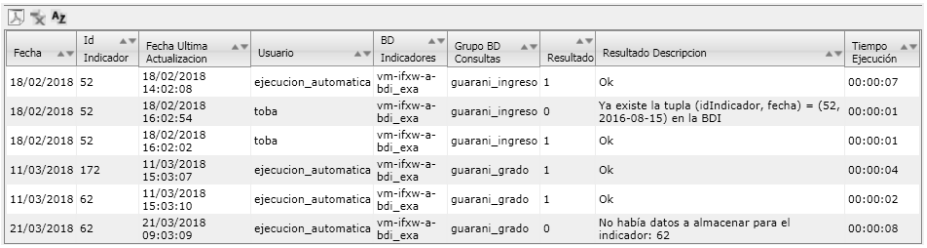


Figure 5 - Configuration of a connection to a TDB



Fecha	Id Indicador	Fecha Ultima Actualizacion	Usuario	BD Indicadores	Grupo BD Consultas	Resultado	Resultado Descripcion	Tiempo Ejecución
18/02/2018	52	18/02/2018 14:02:08	ejecucion_automatica	vm-ifxw-a-bdi_exa	guarani_ingreso	1	Ok	00:00:07
18/02/2018	52	18/02/2018 16:02:54	toba	vm-ifxw-a-bdi_exa	guarani_ingreso	0	Ya existe la tupla (idIndicador, fecha) = (52, 2016-08-15) en la BDI	00:00:01
18/02/2018	52	18/02/2018 16:02:02	toba	vm-ifxw-a-bdi_exa	guarani_ingreso	1	Ok	00:00:01
11/03/2018	172	11/03/2018 15:03:07	ejecucion_automatica	vm-ifxw-a-bdi_exa	guarani_grado	1	Ok	00:00:04
11/03/2018	62	11/03/2018 15:03:10	ejecucion_automatica	vm-ifxw-a-bdi_exa	guarani_grado	1	Ok	00:00:02
21/03/2018	62	21/03/2018 09:03:09	ejecucion_automatica	vm-ifxw-a-bdi_exa	guarani_grado	0	No había datos a almacenar para el indicador: 62	00:00:08

Figure 6 - Audit of the result of the execution in the DB of the integrator

2.1 Code Editor

As a basic requirement of the tool, it was proposed a code editor that will allow writing the necessary queries in SQL enriched with vocabulary programming as PHP. Those queries can be stored in flat files in the server to be later processed in a manual or automatic way according to the calendar, with the frequency of the defined actualization in each indicator.

When a query is defined for the calculation of an indicator for the first time, it is charged to a predefined code to a template, which includes comments according to the DBI guide fields: Name, Methodological Tab, Formula, Frequency of Actualization and Unit of Measure. This is done to orientate the user when writing the necessary code, as it is shown in figure 7.

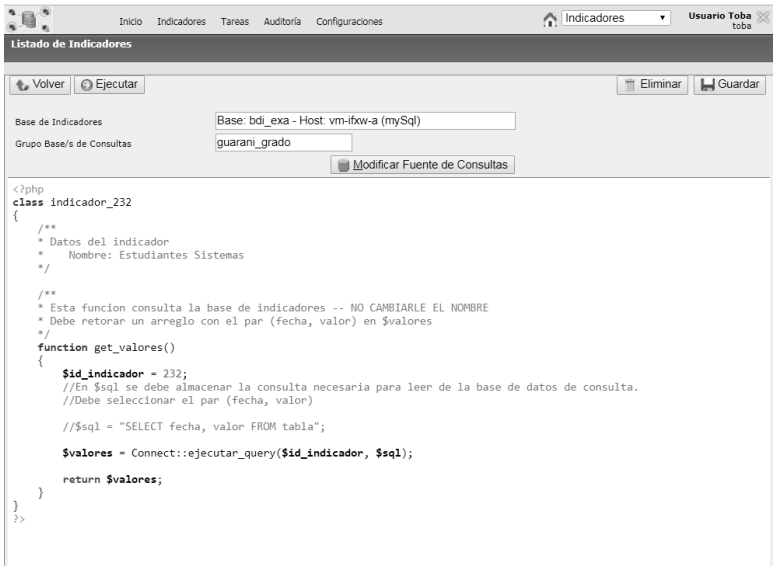


Figure 7 - Basic template for the code editor

Furthermore, it is now necessary to define the query source; that is to say what type of database will be connected to effectively make the query to obtain the values of the indicator we are talking about.

The template includes an obligatory function called “get values” which must be defined and should return in every case with the field date and value of the indicator. If this is not the case, the prototype includes the emission of exceptions to register the failure and notify it to the user.

2.2 Task modeling

To administrate and execute these queries automatically, the concept of task is applied so that the user can manage them in a simple way, allowing the creation of new ones, modifying the previous ones, consulting on the execution of some of them.

2.3 Organizer

A task organizer (scheduler) brings the possibility to coordinate the tasks, to revise and select them. The tasks are shown on a priority list. Once the organizer selects the tasks, it takes them from the priority list.

Besides, every task has a possible state of execution:

- Pending
- Active
- Successfully executed
- Successfully executed without storage of new data
- Executed with a failure

Each of these states has a different colour that represents them intuitively and visually in the calendar.

2.4 Tasks calendar

Another relevant aspect when dealing with the actualization of tasks is the possibility of using a calendar with visual and detailed information about the state of the execution of done queries as programming tasks (Figure 8).

If the task is not successfully executed, it will be red. If the task is pending or should be done in the near future, it will be yellow. If the task is successfully executed, it will be green. If the task is an actualization of data, it will be dark green. If the actualization is successful without an actualization of values, the task will be light green. Last but not least, if the task is being executed at the moment, it will be blue. So, with a simple visualization at the calendar, the user will have a clear panorama of the state of the executions.

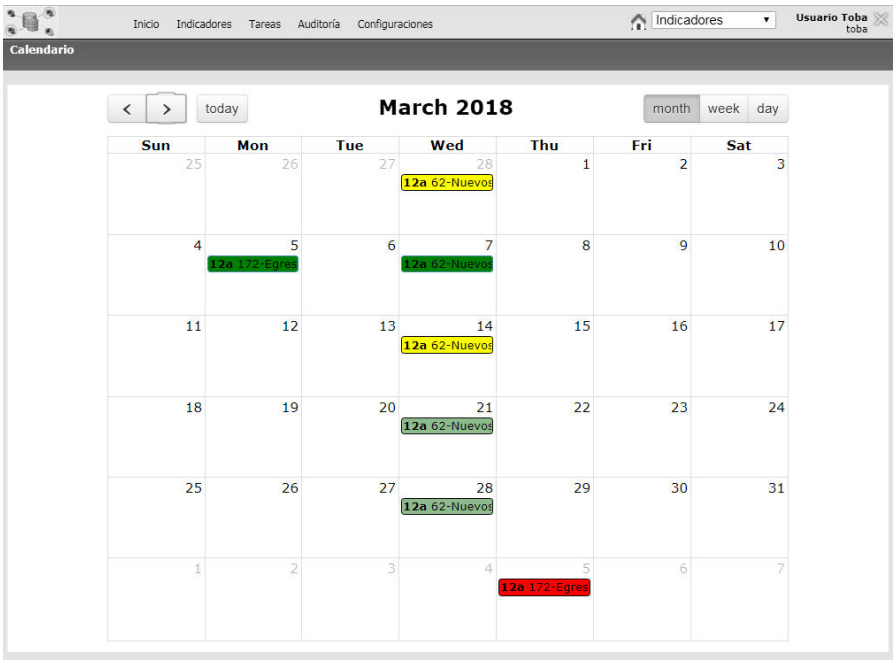


Figure 8 - Tasks calendar

3 3. Conclusions and Future Work

The main idea of this project has been the development of an intuitive prototype simple to use that helps to integrate the different data sources for the management of indicators.

It is achieved the building of a tool which allows the extraction of information and that acts as an interface between the transactional systems of the institution and the prototype that supports the management of indicators (DBI) together with functions that allow the implementation and follow up of automated tasks and register of done operations.

It was possible to install and test the prototype in the Faculty of Exact Sciences of UNCPBA achieving the values of the different systems used in the academic unit to integrate and actualize them in the DBI.

As a test drive, it will be selected (applicants, enrolled pupils, re-enrolled pupils, graduates) for the basic indicators and graduation of each career. On the other hand, the efficiency of indicators (achievement level), efficiency (effort level to get the desired result), efficiency and effectiveness by cohort, breakdown, among others have been defined [2].

Then, it is presented some future work which could be incorporated into the tool after knowing the result of this investigation:

- It allows creating, modifying, moving and deleting tasks from the calendar in an intuitive way.

- It extends the function of connecting with a new database apart from Informix, Postgres, and MySQL.
- It notifies by mail in case of a failure in the execution phase of any automatic task.
- It reports the state of the tasks.
- It continues with the development to improve the use and performance of the prototype which will be the principal matter for a future work.
- It will have a coded access to a REST type web-service for the actualizations in the DBI.

References

1. Ballvé A.: Tablero de control. Ed. Macchi. Buenos Aires. (2000)
2. Illescas G, Santiago M, Martínez M.: Procedimientos y herramientas para la mejora de indicadores académicos en facultades de ingeniería. 20vo Workshop de Investigadores en Ciencias de la Computación (WICC 2018). 1130--1134 (2018)
3. Illescas G. (2014), Tesis Doctoral: Aplicación de métodos matemáticos en el control de gestión por indicadores". En repositorio digital de la biblioteca central UNICEN. (2014)
4. Kaplan R., Norton D.: Cuadro de mando integral. 2da edición. Ediciones GESTION 2000. (2000)
5. AENOR. Guía para la implantación de indicadores. Norma Española UNE66175. Asociación Española de Normalización y Certificación. Madrid, España. (2003)
6. Plan Estratégico de Formación de Ingenieros (PEFI) http://pefi.siu.edu.ar, downloaded 2018-02-19
7. Procesos ETL, https://www.datasciencecentral.com/profiles/blogs/10-open-source-etl-tools, downloaded 2017-07-06
8. IBM Websphere DataStage, https://www.ibm.com/support/knowledgecenter/es/SSZJPZ_8.1.0/com.ibm.swg.im.iis.productization.iisinfsv.overview.doc/topics/cisodsoverview.html, downloaded 2017-07-21
9. Pentaho Data Integration (Kettle ETL), http://www.pentaho.com/product/data-integration#data-integration-, downloaded 2017-05-21
10. Oracle Warehouse Builder, https://docs.oracle.com/cd/B28359_01/owb.111/b31278/concept_overview.htm#BABJGDAF, downloaded 2017-05-11
11. Microsoft SQL Server Integration Services (SSIS), https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services, downloaded 2017-06-09

Descubriendo reglas de asociación en bases de datos del sector retail usando R

Juan Manuel Báez Acuña, Clara Anuncia Paredes Cabañas,
Gustavo Sosa-Cabrera, María Elena García

Facultad Politécnica
Universidad Nacional de Asunción
San Lorenzo, Paraguay
{juanmanuelbaez,cparedescabanas,gdsosa,mgarcia}@pol.una.py
http://www.pol.una.py

Resumen A día de hoy, la alta competitividad en los negocios del tipo retail les exige buscar nuevas estrategias para garantizar su supervivencia. A estos efectos, las organizaciones han comprendido que los datos residentes en sus bases de datos transaccionales pueden ser utilizados como materia prima para impulsar el crecimiento del negocio, si es que los mismos pueden explotarse con propiedad. Este trabajo de investigación tiene como objetivo principal aplicar técnicas de Minería de Datos para el descubrimiento de reglas de asociación, tomando como período de estudio datos transaccionales netamente comerciales, en un intervalo de 10 años de una entidad retail de electrodomésticos y muebles. Se describe la fase de selección y preparación de los datos así como también su costo en horas/hombre. En la fase de modelado se ha corrido los algoritmos *Apriori* y *Eclat* implementados en el paquete *arules* de la herramienta *R* donde tanto las asociaciones resultantes como el tiempo de ejecución fueron comparados entre ambos algoritmos. Los resultados demuestran patrones relevantes en el comportamiento de compra de los clientes como ser aquellos que relacionan artículos y precio de accesorios.

Palabras Claves: Minería de Datos; reglas de asociación; retail; bases de datos transaccionales; herramienta R, Apriori, Eclat

1. Introducción

El crecimiento desenfrenado de las bases de datos en los últimos años, especialmente aquellas del tipo de datos sobre actividades cotidianas como las elecciones de los clientes, lleva a la minería de datos a la vanguardia de las nuevas tecnologías comerciales. Los datos analizados de manera inteligente son un recurso muy valioso y puede conducir a nuevos conocimientos y, en entornos comerciales, a ventajas competitivas para las empresas. En la economía de hoy, altamente competitiva, centrada en el cliente y orientada a los servicios, los datos son la materia prima que impulsa el crecimiento del negocio, si es que puede explotarse con propiedad, creando un importante valor agregado en el negocio.

Asimismo, cada día cobra más relevancia el estudio de obtención de conocimiento útil desde datos almacenados en grandes repositorios, ya que la misma es reconocida como una necesidad básica en muchas áreas, especialmente aquellas relacionadas con los negocios del sector conocido como *retail* [1].

El presente estudio de caso involucra a una empresa del sector retail pionera en la venta de electrodomésticos a nivel país y que ha logrado posicionarse como una de las compañías más reconocidas a lo largo de sus más de 6 décadas de existencia en el mercado [2]. Dicha empresa inició la incursión de medios informáticos para el desarrollo de sus actividades en la década de 1990 y desde entonces ya ha pasado por 3 versiones de software de gestión en retail, cada una con su base de datos relacional con estructura propia, registrando sus movimientos por medio de transacciones. En la actualidad, la empresa cuenta con más de 16 sucursales por lo que ha experimentado un crecimiento vertiginoso tanto de ventas como de cantidad de información. Por otra parte, también ha crecido la necesidad de buscar nuevas estrategias para garantizar la supervivencia de los negocios, esto es, como una consecuencia de la globalización y la alta competitividad del sector.

Surge, por tanto, la necesidad de la inteligencia de negocios que pueda transformar los datos en conocimiento, para que este pueda ser usado oportunamente en la toma de decisiones, propiciando acciones que resulten en una ventaja competitiva para la empresa [8,17].

Este trabajo tiene como objetivo exponer los resultados obtenidos de la aplicación de técnicas de descubrimiento de asociaciones donde se toma como período de estudio los registros que pertenecen a las transacciones netamente comerciales de la empresa, realizadas en un intervalo de tiempo de 10 años.

2. Aspectos teóricos

La minería de datos es el proceso de extraer conocimiento útil y comprensible desde grandes cantidades de datos almacenados en distintos formatos [5]. Este proceso forma parte de una secuencia iterativa de etapas para el descubrimiento de conocimiento en bases de datos [4]. Se distinguen 2 tipos de tareas, las predictivas (clasificación y regresión) y las descriptivas (agrupamiento y las reglas de asociación) [3].

En el descubrimiento de reglas de asociación, se pretende obtener conocimiento interesante como ser los hábitos de compra de los clientes mediante por ejemplo, la relación de los diferentes artículos en sus "cestas de compras" [10].

Es conocido que los algoritmos esenciales en la búsqueda de reglas de asociación en bases de datos son el Apriori, el Eclat y el FP Growth [11]. Sin embargo, en [12] se considera que únicamente el Apriori y el Eclat son las dos grandes familias ya que se puede incluir al FP Growth como miembro del Eclat.

El algoritmo Apriori [13] busca primero todos los conjuntos frecuentes unitarios (contando sus ocurrencias directamente en la base de datos), se mezclan estos para formar los conjuntos de ítems candidatos de dos elementos y seleccionan entre ellos los frecuentes. Considerando la propiedad de los conjuntos de ítems frecuentes, se vuelve a mezclar estos últimos y se seleccionan los frecuentes

(hasta el momento ya han sido generados todos los conjuntos de ítems frecuentes de tres o menos elementos). Así sucesivamente se repite el proceso hasta que en una iteración no se obtengan conjuntos frecuentes.

En contraste, el algoritmo Eclat [14] se basa en realizar un agrupamiento (clustering) entre los ítems para aproximarse al conjunto de ítems frecuentes maximales y luego emplean algoritmos eficientes para generar los ítems frecuentes contenidos en cada grupo. Para el agrupamiento proponen dos métodos que son empleados después de descubrir los conjuntos frecuentes de dos elementos: el primero, por clases de equivalencia: esta técnica agrupa los itemsets que tienen el primer ítem igual. El segundo, por la búsqueda de cliques maximales: se genera un grafo de equivalencia cuyos nodos son los ítems, y los arcos conectan los ítems de los 2-itemsets frecuentes, se agrupan los ítems por aquellos que forman cliques maximales.

Finalmente, *R* es un lenguaje de programación y un entorno que proporciona una amplia variedad de técnicas estadísticas y gráficas. Además, es altamente extensible a través de paquetes que se encuentran disponibles través de sitios de Internet de CRAN que cubren una amplia gama de estadísticas modernas.

3. Materiales y Métodos

3.1. Datos

La empresa ha sido informatizada en la década de 1990, generando desde entonces grandes cantidades de información que fueron almacenados en bases de datos, esto es, de acuerdo a las distintas versiones de software implementadas y con el proceso de migración de registros consecuente. Además, ante la ausencia de una base de datos dedicada a la inteligencia de negocios, para este estudio, se montó una base de datos "espejo" a la transaccional donde se tomaron en cuenta los movimientos correspondientes a las ventas de electrodomésticos y muebles de los periodos cerrados contablemente 2008 al 2017, debido a que a partir del año 2008 la empresa retail amplió su catálogo de productos, incorporando la venta de muebles [6].

Fase de selección. Las entidades registradas en las tablas de historiales y que fueron identificadas como factores que intervienen en la venta se muestran junto con sus atributos en la Figura 1.

Fase de preprocesamiento. Se describe a continuación por entidad:

- *Vendedores*: tenía como atributos nombre, apellido, tipo de vendedor (vendedor, cobrador o ambos), para el estudio se requería incluir variables como sexo, edad, y tipos más específicos a los existentes. De un total de 1292 registros de vendedores, quedaron 780 con movimientos en el período de estudio. Para asignar el sexo, se tuvo en cuenta el género del nombre, con una efectividad del 90 %, luego se aplicó una verificación individual hasta lograr que el

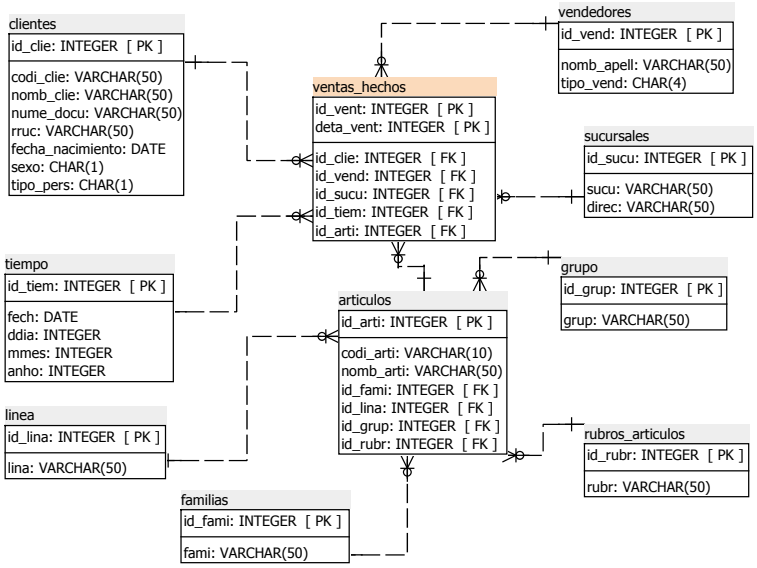


Figura 1. Esquema estrella de la base de datos de la entidad retail.

100 % de los registros cuente con sexo asignado. Para la variable edad se ha cruzado los datos de los vendedores con los registros de recursos humanos, identificando 132 vendedores, luego se realizó comparaciones con planillas de pago del seguro social, tarea que no se pudo culminar por falta de información completa, por tanto ésta variable fue descartada para el estudio, por último teniendo en cuenta los códigos presentes en el nombre del vendedor, en forma manual se asignaron 11 nuevas categorías.

- *Artículos*: contiene los datos de los productos comercializados, clasificados en "familia", "línea" y "grupo" y distribuidos en 13 rubros de artículos. De un total de 37413 registros, 26189 se encontraban en estado "Discontinuado", es decir que no contaban con una clasificación válida para el estudio, razón por la cual debían ser recategorizados. Se ha excluido 17037 registros, por no poseer movimientos en el período de estudio. Se ha excluido 106 registros por no pertenecer al rubro de venta de electrodomésticos o muebles, quedando 20270 registros para la muestra. Luego de haber excluido los registros mencionados, la cantidad de artículos en estado "Discontinuado", se redujo a 14335. Teniendo en cuenta la descripción del artículo, se logró asignar una nueva clasificación a 14074 registros, debido a la falta de estándares, en la carga de los artículos en la base de datos, se ha realizado la verificación indi-

vidual, corrigiendo manualmente 6971 registros y finalmente se ha excluido 261 por no pertenecer a los rubros de electrodomésticos y muebles quedando 20009 registros válidos para el estudio. Debido a que los artículos contaban con categorías muy generales, se realizó una comparación con los artículos expuestos en el portal web de la empresa retail, en la cual se encontró que la clasificación de "Familia", "Línea" y "Grupo" no era la misma que se disponía en la base de datos, se pudo observar que eran más específicas y teniendo en cuenta la misma se logró ampliar tal clasificación, seguidamente utilizando el atributo descripción, se asignó una nueva clasificación a todos los artículos preprocesados.

- *Clientes*: con 454737 registros, se ha excluido 182449 registros, por no poseer movimientos en el período de estudio, quedando 272288. Seguidamente se ha comparado el atributo "Número de Documento" con los registros de la base de datos proveída por el Departamento de Identificaciones de la Policía Nacional de Paraguay, encontrando 247766 coincidencias, a las cuales se ha actualizado los atributos "fecha de nacimiento", "tipo de persona" y "sexo". Luego, teniendo en cuenta el atributo "RUC" [7], omitiendo el dígito verificador [7], se encontró 2247 registros correspondientes a personas físicas, se ha identificado además otros 46 registros con caracteres anómalos, a los cuales se actualizaron los atributos "fecha de nacimiento", "tipo de persona" y "sexo", 14072 registros correspondientes a empresas, diplomáticos y personas con documento extranjero, teniendo en cuenta el formato de RUC, se actualizó estos últimos, "sexo" y "tipo de persona", descartando el atributo "fecha de nacimiento" para estos casos. Teniendo en cuenta el atributo "nombre del cliente", se ha identificado 490 registros correspondientes a empresas [9], 126 clientes registrados como "Lista de Bodas", aplicando a los registros la misma actualización anterior, se ha aplicado el mismo procedimiento utilizado para la entidad "Vendedores" para la identificación de sexo por género del nombre a los 7541 registros restantes, obteniendo los siguientes resultados: 3263 eran de sexo femenino, 3758 eran de sexo masculino, se corrigieron manualmente 25 registros que correspondían a empresas, obteniendo un acierto del $93,1\,\%$ con el procedimiento. Por último quedaron 495 registros sin actualizar, para los cuales se realizó un trabajo manual e individual utilizando los nombres, comparando los mismos con búsquedas realizadas en el buscador de *Google*, identificando 162 registros con sexo femenino, 290 registros con sexo masculino y 20 registros de empresas. Finalmente, únicamente se descartaron 23 registros por falta de información adecuada para realizar la actualización. Totalizando 272265 registros óptimos para las pruebas.
- *Ventas*: para esta entidad existe un ínfimo grupo de transacciones no señalizadas en la base de datos, que pertenecerían a otorgamientos en concepto de regalo de artículos a clientes por alguna promoción y no así una venta como tal, en este sentido, se ha entrevistado a los departamentos de Marketing, Ventas, Contabilidad y Auditoría de la empresa quienes han avalado que dichas ocurrencias era despreciable para los fines del estudio. Además, la relación existente entre bajo número de variables y el tamaño muestral hace que el estimador estadístico sea consistente [19].

Importante. Si bien el modelo tradicional de descubrimiento de conocimiento en bases de datos no contempla actividades para la gestión del proyecto [18], en este trabajo se ha llevado a cabo una planificación y medición del tiempo donde las fases de selección y preprocesamiento de los datos, tuvieron en conjunto un costo aproximado de 880 horas/hombre.

Vistas minables. Para finalizar la preparación de los datos, se realizaron diferentes combinaciones para obtener las vistas minables (Cuadro 1) donde cada una contiene los registros de ventas con un detalle por cada artículo vendido.

Entidades	**Vistas Minables**	
Artículos	1. Familia 2. Linea	3. Grupo
Artículos/Vendedores	1. Sexo/Grupo 2. TipoVend/Grupo 3. Sexo/Familia 4. TipoVend/Familia 5. Sexo/Linea 6. TipoVend/Linea	7. Sexo/TipoVend/Grupo 8. Sexo/TipoVend/Familia 9. Sexo/TipoVend/Linea 10. Sexo/Articulos 11. TipoVend/Articulos 12. Sexo/TipoVend/Articulos
Artículos/Clientes	1. Edad/Grupo 2. Edad/Familia 3. Edad/Linea 4. Sexo/Grupo 5. Sexo/Familia 6. Sexo/Linea	7. TipoPers/Grupo 8. TipoPers/Familia 9. TipoPers/Linea 10. Edad/Sexo/Grupo 11. Edad/Sexo/Familia 12. Edad/Sexo/Linea
Artículos/Tiempo	1. Dia de la Semana/Grupo 2. Mes / Grupo 3. Picos de Venta / Grupo	4. Trimestre/Grupo 5. Semestre/Grupo 6. Estación/Grupo
Artículos/Tiempo/Clientes	1. Edad/SemanaDia/Grupo 2. Edad/Estacion/Grupo 3. Sexo/SemanaDia/Grupo	4. Sexo/Estacion/Grupo 5. Sexo/Edad/SemanaDia/Grupo 6. Sexo/Edad/Estacion/Grupo
Artículos/Clientes/Vendedores	1. Edad/SexoVend/Grupo 2. Edad/TipoVend/Grupo 3. TipoPers/TipoVend/Grupo	4.TipoPers/SexoVend/Grupo 5. SexoCliente/SexoVend/Grupo 6. SexoCliente/TipoVend/Grupo

Cuadro 1. Lista de vistas minables.

3.2. Inducción de Reglas de Asociación

Primeramente se aplicó la técnica *Market Basket* (cesta de compra) [16] utilizada para descubrir asociaciones entre artículos. Se describe a continuación los pasos del tratamiento aplicado a las vistas minables, para lograr el formato *basket* (Cuadro 2), necesario para la aplicación de los algoritmos Apriori y Eclat.

```
> df_vista<-read.csv("arti_clie_gral.csv", header=TRUE, sep=";")
> df_itemList <- ddply(df_vista, c("id_venta","rango_edad_clie"),
    function(df1)paste(df1\$grupo, collapse=";"))
> df_itemList <- unite(df_itemList, items,
    c(V1, rango_edad_clie), sep=";",  remove=TRUE)
> df_itemList\$id_venta <- NULL
> write.csv(df_itemList, "ItemList.csv", row.names=TRUE)
```

id_venta	rango_edad	grupo
100101	Adulto	Cocina
100101	Adulto	Cafetera
100101	Adulto	Horno Eléct.
100102	Joven	Celular
100103	Joven	Cafetera
100103	Joven	Televisor

$\Longrightarrow$

id_venta	rango_edad	V1
100101	Adulto	Cocina; Cafetera; Horno Eléct.
100102	Joven	Celular
100103	Joven	Cafetera; Televisor

Cuadro 2. Transformación de registros al formato basket.

Posteriormente, para la inducción de reglas de asociación, se ha utilizado el paquete arules [15] que contiene la implementación de los algoritmos "A PRIORI" y "ECLAT". Para determinar el grado de "significancia" e "interés" de las reglas, se ha utilizado los conocidos umbrales mínimos de "soporte" y "confianza" respectivamente, a saber:

$$Sop(X) = \frac{|X|}{|D|} \qquad \text{y} \qquad Conf(X \implies Y) = \frac{Sop(X \cap Y)}{Sop(X)} = \frac{|X \cap Y|}{|X|}.$$

Importante. Para este estudio se utilizaron 5 medidas de soporte: 20 %, 10 %, 5 %, 1 % y 0,1 %, debido a la cantidad de registros de transacciones que se dispone la empresa retail, se tuvieron en cuenta soportes muy bajos para no descartar ítems frecuentes que se pueden perder por la cantidad de registros existentes. Asimismo, el porcentaje de la confianza utilizado fue de 1 %, para no excluir ninguna regla, debido a que para este estudio todas las reglas generadas dignas de ser analizadas.

Corrida de los algoritmos. Para la generación de las reglas de asociación, las corridas de los algoritmos se ejecutaron de la siguiente forma:

```
> datos = read.transactions(file="ItemList.csv",
    rm.duplicates=TRUE, format="basket",
    sep=";",cols=1)
> reglas <- apriori(datos, parameter= list(supp=0.1,
    conf=0.01, target="rules"))
> reglas <- ruleInduction(eclat(datos, parameter=list(supp= 0.1)),
    datos, confidence=.01)
> inspect(reglas)
```

4. Resultados

4.1. Reglas Encontradas

Dada la variable principal "Grupo" del artículo, en el cuadro 3 se observan las reglas encontradas más relevantes, separados en grupos de 10 años por un lado y del último año por el otro.

2008-2017			2017		
Antecedente	**Consecuente**	**Confianza**	**Antecedente**	**Consecuente**	**Confianza**
Base, Colchones, Mesas de Noche	Cabeceras	0,97	Colchones	Base	0,96
Colchones, Mesas de Noche	Cabeceras	0,95	Base, Colchones, Mesas de Noche	Cabeceras	0,96
Colchones	Base	0,93	Colchones, Mesas de Noche	Cabeceras	0,95
Mueble p/PC	Webcam	0,85	Licuadoras	Acc. p/Licuad.	0,78
Monitores, CPU	Muebles p/PC	0,73	Base	Colchones	0,74
Monitores, CPU	Webcam	0,72	Base, Colchones	Mesas de Noche	0,68
Base, Colchones	Mesas de Noche	0,62	Caminadoras	Banco p/Abds	0,64
Base	Colchones	0,62	SMART TV	Soporte p/TV	0,59
Cunas	Colchones	0,55	Hornos de Mesa, Planchas	Licuadoras	0,57
Caminadoras	Banco p/Abds	0,53	Hornos de Mesa	Balanza de Cocina	0,55

Cuadro 3. Las 10 reglas con mayor porcentaje de confianza por período.

Los resultados muestran que en ambos periodos existe una fuerte relación *[Colchón — Base de Colchón]* y sus diferentes accesorios, obteniendo un porcentaje de confianza mayor al 95 %, otra regla que aparece en ambos periodos es la de *[Banco para Abdominales — Caminadoras]*, con porcentajes de confianza mayor al 50 % en ambos casos, lo que indica que existe alta probabilidad de venta de dichos artículos en conjunto. Además de mencionar que estas combinaciones de artículos se mantuvieron con el paso del tiempo, lo interesante de estos datos es que, teniendo en cuenta el factor "precio de venta del artículo" y tomando todos los resultados mostrados en el Cuadro 3, se refleja que la compra de los artículos de mayor costo, inducen a la compra de sus respectivos accesorios o complementos, siempre y cuando su precio es menor o igual al 25 % del artículo principal, información muy valiosa que puede ser utilizada por el área de marketing para obtener beneficios e incrementar las ventas de artículos de menor costo, o la venta de artículos que tienen mayor cantidad de complementos, realizando campañas de promociones armando combos de dichos artículos. Su utilidad es extensible al área comercial, como herramienta de conocimiento para vendedores, al momento de ofrecer artículos al cliente.

Importante. Estos resultados han sido interpretados y evaluados por personal especializado del área de ventas de la empresa retail.

Comparación de algoritmos En el cuadro 4, se refleja la comparación de los algoritmos utilizados para el experimento, teniendo en cuenta parámetros como "Cantidad de Observaciones" e "Ítems", midiendo el tiempo de ejecución y cantidad de reglas arrojadas.

Tiempo de ejecución Todas las pruebas fueron realizadas sobre una hardware con procesador Intel® Core(TM) I3-4005U CPU @ 1,70GHz y memoria RAM de 4Gb. Para el 95 % de las pruebas realizadas, el algoritmo Apriori posee un menor tiempo de ejecución en comparación al Eclat en la generación de reglas. Ademas, en el cuadro 4, se puede apreciar que mientras mayor sea la cantidad

de ítems a ser combinados por los algoritmos, es mayor la brecha en cuanto al tiempo de ejecución.

Cantidad de reglas encontradas Las diferencias encontradas, se deben a que el algoritmo Apriori genera reglas con 1 ítem frecuente que cumple con el soporte especificado como parámetro, estas reglas no poseen antecedentes, las mismas indican que existe la probabilidad de venta de ese artículo sin tener en cuenta otro ítem involucrado.

Descripción	Cant.de Obs.	Cant.de ítems	Variables	Sop.	Conf.	APRIORI		ECLAT	
						Tiempo Ejecución (seg)	Cant Reglas	Tiempo Ejecución (seg)	Cant. Reglas
Registros de venta del período 2017	124.386	330	Grupo del Artículo	0,001	0,01	2,97	218	3,36	185
				0,01	0,01	0,95	35	1,93	2
				0,05	0,01	0,69	3	0,71	0
				0,1	0,01	0,67	0	*	*
				0,2	0,01	0,68	0	*	*
Registros de venta del período 2008 al 2017	1.007.064	543	Grupo del Artículo	0,001	0,01	2,46	267	2,67	234
				0,01	0,01	2,25	33	2,14	0
				0,05	0,01	1,76	5	1,52	0
				0,1	0,01	1,41	0	*	*
				0,2	0,01	1,41	0	*	*
Registros de venta del período 2017	1.125.448	20.265	Articulo	0,0001	0,01	4,78	478	384,6	476
				0,001	0,01	2,87	4	11,7	2
				0,01	0,01	2,29	2	10,2	0
				0,05	0,01	2,55	0	*	*
				0,1	0,01	2,4	0	*	*
				0,2	0,01	2,39	0	*	*

*Sin generación de reglas debido a que el *ECLAT*, no encontró ítemsets frecuentes.

Cuadro 4. Rendimiento en cuanto a tiempo de ejecución y reglas encontradas.

5. Conclusión

Este estudio ha contribuido a evidenciar los resultados de la aplicación de técnicas de descubrimiento de asociaciones en bases de datos transaccionales del sector retail mediante el uso de herramientas de software libre.

Las reglas de asociación encontradas para la venta de artículos en conjunto (con medidas de confianza superiores al 50 %), indica la existencia de asociaciones relevantes y que se pueden considerar como un patrón de comportamiento válido de los clientes.

La segunda contribución más importante ha sido la medición del gran tiempo invertido para la preparación de los datos. Esto es, debido a las numerosas tareas de preprocesamiento que ha requerido la base de datos transaccional caracterizada por varias actualizaciones de versiones y migraciones desde la década de 1990.

Esta contribución es muy valiosa y debe ser considerada para investigaciones similares futuras, como ser, la aplicación de nuevas técnicas para la reducción del tiempo en la fase de preprocesamiento de los datos.

Referencias

1. Douglas, H.: retail — Origin and meaning of retail by Online Etymology Dictionary, https://goo.gl/zzwvu2 Accedido 25 May 2018.
2. Grupo González Giménez Desde 1952, https://goo.gl/MY3oVv Accedido 05 Jul 2018.
3. Witten, I., Frank, E.: Data mining. Morgan Kaufmann, San Francisco, Calif. (2005).
4. Fayyad, U., Irani, K.: Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning. 13th International Joint Conference on Artificial Intelligence. pp. 1022-1027 (1993).
5. Witten, I., Frank, E.: Data Mining: Practical machine learning tools with Java implementations. M. Kaufmann. (2000).
6. Ultima Hora: González Giménez expande sus productos y servicios, https://goo.gl/LcmVvn, (2008) Accedido 17 Jun 2018.
7. LEY N° 1352/88, http://www.impuestospy.com/Leyes/Ley%201352_88.php.
8. JlaweJ, H., Kamber, M.: Data Mining: Concepts and Techniques. Simón Fraser University. Morgan Kaufmann Publishers. (2002).
9. Subsecretaría de Estado de Tributación - Lista de pequeños contribuyentes, https://goo.gl/Fpqny5.
10. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. ACM SIGMOD Record. 22, 207-216 (1993).
11. Heaton, J.: Comparing dataset characteristics that favor the Apriori, Eclat or FP-Growth frequent itemset mining algorithms. SoutheastCon. pp. 1-7. IEEE (2016).
12. Schmidt-Thieme, L.: Algorithmic Features of Eclat. FIMI (2004).
13. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. Proc. 20th int. conf. very large data bases, VLDB. pp. 487-499 (1994).
14. Zaki, M., Parthasarathy, S., Ogihara, M., LI, W.: New algorithms for fast discovery of association rules. KDD. pp. 283-286 (1997).
15. Mining Association Rules and Frequent Itemsets [R package arules version 1.6-1], https://cran.r-project.org/package=arules Accedido 03 May 2018.
16. Han, J., Pei, J., Kamber, M.: Data mining: concepts and techniques. Elsevier (2012).
17. Kim, J., Ale, J.: Descubrimiento incremental de las reglas de asociación temporales. X Congreso Argentino de Ciencias de la Computación (2004).
18. Moine, J., Gordillo, S., Haedo, A.: Análisis comparativo de metodologías para la gestión de proyectos de Minería de Datos. Congreso Argentino de Ciencias de la Computación (2011).
19. Sosa-Cabrera, G., García-Torres, M., Gómez, S., Schaerer, C., Divina, F.: Understanding a Version of Multivariate Symmetric Uncertainty to assist in Feature Selection. Conference of Computational Interdisciplinary Science (2016).

Proceso para el Descubrimiento de Asociaciones Espaciales Mediante Subgrafos Frecuentes

Giovanni Daián Rottoli[1,2,3][0000−0002−7623−2591], Hernán Merlino[3], and Ramón García-Martínez[3†]

[1] Programa de Doctorado en Ciencias Informáticas, Facultad de Informática, Universidad Nacional de La Plata (UNLP). La Plata, Buenos Aires, Argentina.
[2] Programa de Beca Doctoral para el Fortalecimiento de Áreas I+D+i. Universidad Tecnológica Nacional (UTN). Argentina
rottolig@frcu.utn.edu.ar
[3] Grupo de Investigación en Sistemas de Información (GISI), Universidad Nacional de Lanús (UNLa). Lanús, Buenos Aires, Argentina.
hmerlino@unla.edu.ar

Resumen Las asociaciones espaciales constituyen uno de los tipos de regularidades más utilizadas para la inteligencia de negocio en relación a los datos espacialmente referenciados. Debido a las particularidades de este tipo de información, distintos acercamientos han sido propuestos a lo largo del tiempo para el minado de las mismas, sin embargo, no existe una forma definida de realizar esta actividad que sea adaptable a las situaciones particulares de los distintos problemas de negocio que se presentan. Por este motivo se propone en el presente trabajo un proceso de explotación de información que permita modelar distintos escenarios y relaciones espaciales haciendo uso de la teoría de grafos, y la búsqueda de los patrones de asociación mediante el uso de algoritmos de búsqueda de subgrafos frecuentes. Una prueba de concepto de la propuesta utilizando datos reales es llevada a cabo para ilustrar la implementación del mismo.

Keywords: Asociación Espacial, Minería de Datos Espaciales, Explotación de Información Espacial, Subgrafos Frequentes

1. Introducción

La explotación de información espacial busca encontrar patrones interesantes y novedosos en un conjunto de datos espacialmente referenciados para así apoyar la toma de decisiones estratégicas de una organización [1]. Entre estos patrones interesantes, las asociaciones espaciales constituyen uno de los más utilizados [2], sin embargo, debido a las características de este tipo particular de información, diferentes desafíos se presentan [3].

En primer lugar, $D = \{\{s_j\}, \{v_k\}\}$ es un dato espacial si está compuesto por dos clases de atributos, un conjunto de n atributos espaciales $\{s_j\}, j = 1..n$, y un conjunto de m atributos no espaciales $\{v_j\}, k = 1..m$. Los últimos mencionados almacenan mediciones sobre el dominio del problema, mientras que los primeros vinculan estas mediciones con un lugar específico del espacio [4].

A partir de esta dependencia espacial de los atributos se pueden extraer relaciones entre las distintas instancias de los datos tales como relaciones geométricas basadas en la distancia euclideana, como la vecindad o la lejanía, relaciones topológicas, como las intersecciones e inclusiones, relaciones de ubicación, como que un objeto se encuentre al norte o debajo de otro, y relaciones mixtas como la perpendicularidad o el paralelismo.

Estas relaciones se calculan de diversas maneras en función del dominio del problema y del tipo de dato espacial que estemos utilizando: puntos, lineas o regiones. En consecuencia, las asociaciones en el espacio no solo contemplarán los valores de los atributos no espaciales, sino que además deberán considerar las relaciones entre los objetos espaciales, a diferencia de los patrones de asociación transaccionales tradicionales [5, 6].

Por otro lado, es necesario considerar además el fenómeno de la heterogeneidad espacial: un patrón espacial puede tener distinta probabilidad de ocurrencia en distintas ubicaciones del espacio, por lo que tiene sentido la búsqueda de patrones locales [7]. De esta forma, las asociaciones deberán ser analizadas particularmente en secciones del espacio estudiado.

Todas estas características deben ser tenidas en cuenta para encontrar patrones de asociación en datos espacialmente referenciados, sin embargo, los distintos trabajos realizados hasta la fecha solucionan los distintos problemas que se presentan de manera aislada: Saleb y Christel (2000) [8] proponen la consideración de múltiples jerarquías de datos espaciales; Sutjipto y otros [9] utilizan acercamientos basados en el algoritmo A-Priori, Malerba y Lisi (2001) [10], Malerba y Esposito (2002) [11] y Appice y otros (2003) [6] proponen la utilización de algoritmos basados en Programación Lógica Inductiva, mientras que Ladner y otros (2003) [12], y Bai y otros (2014) [13] incorporan el uso de lógica difusa (*fuzzy logic*) para el modelado de las relaciones entre los objetos espaciales, y Zhang (2017) [14] y Goudarzi y Ghadiri (2017) [15] la utilización de metaheurísticas como algoritmos genéticos y programación evolutiva. Por otro lado, Lee (2004) [16] propone un acercamiento basados en transformación de datos y Yang y otros (2005) [17] incorpora el uso de estructuras como grafos. Por último, distintos fenómenos son considerados: la localidad de los patrones por Ding y otros (2006, 2011) [18, 19] y Sha y otros (2010, 2015) [20, 21], y la autocorrelación espacial, el efecto vecindario y el efecto derrame por Dao y Thill [22].

Esta variedad de soluciones hace que aumente la complejidad de modelar el problema correctamente e implementar de forma adecuada una solución para el mismo. Por este motivo, es necesario un proceso para la búsqueda de patrones de asociación que permita el modelado de distintas relaciones espaciales simultáneamente entre distintas clases de objetos espaciales, la posibilidad de buscar los patrones localmente, y la búsqueda de patrones alrededor de objetos espaciales de referencia, para enfocar la búsqueda según el problema considerado.

Para esto, se presenta el proceso de explotación de información propuesto analizando cada una de sus etapas (sección 2) y se realiza una prueba de concepto del mismo sobre datos reales para ilustrar la implementación del mismo (sección 3). Finalmente se presentan conclusiones (sección 4).

2. Proceso para el Descubrimiento de Asociaciones Espaciales

Un patrón de asociación consiste en un predicado compuesto que posee una alta probabilidad de ocurrencia. Una asociación es llamada espacial si por lo menos uno de sus predicados simples componentes involucra una relación espacial entre objetos espaciales. Por ejemplo, sean A una ciudad perteneciente al conjunto de datos, X una iglesia ubicada en A, Y una plaza ubicada en A, entonces se verifica que en A, las iglesias se encuentran cerca de las plazas:

$$EsCiudad(A) \wedge EsIglesia(X) \wedge EsPlaza(Y) \wedge Incluido(X, A) \wedge Incluido(Y, A) \wedge Vecinos(X, Y)$$

Como se puede observar en este predicado, $Incluido(X, C)$, $Incluido(Y, C)$ y $Vecinos(X, Y)$ son predicados espaciales relacionados con relaciones topológicas y geométricas. Múltiples relaciones distintas deben ser tenidas en cuenta para encontrar asociaciones complejas que resulten interesantes para la inteligencia de negocio. Además, estas relaciones deben ser calculadas localmente debido al fenómeno ya mencionado de la heterogenidad espacial [7].

En el presente trabajo, en consecuencia, se introduce un proceso de explotación de información que permite extraer asociaciones espaciales contemplando la posibilidad de incorporar múltiples relaciones espaciales, la heterogeneidad espacial, y el poder limitar la búsqueda alrededor de tipos de datos de interés para el dominio del problema. Este proceso (Figura 1) consta de 5 fases: preparación de datos (Sección 2.1), definición de vecindarios (Sección 2.2), modelado gráfico de relaciones espaciales (Sección 2.3), minado de subgrafos frecuentes (Sección 2.4), y evaluación de resultados (Sección 2.5).

2.1. Preparación de los datos

La primera etapa del proceso de explotación de información consiste en la preparación de los datos espacialmente referenciados. Estos datos, disponibles en distintas fuentes, deben ser codificados de tal forma que las relaciones entre los mismos puedan ser extraídas en etapas posteriores.

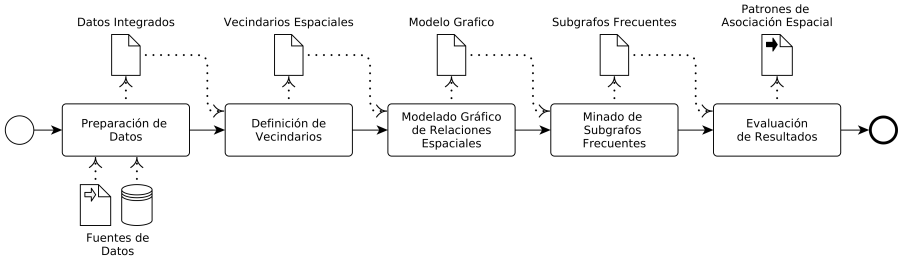


Figura 1. Proceso de Explotación de Información para el descubrimiento de Asociaciones Espaciales.

Dos conjuntos de datos datos deben ser determinados: los datos objetivo, alrededor de los cuales se desea encontrar patrones de conocimiento, y los datos relevantes, que se relacionarán con los datos objetivos y proveerán información útil para el dominio del problema [23]. Estos conjuntos deberán ser limpiados y preparados, resolviendo errores, inconsistencias, datos nulos y anomalías. Además, nuevos atributos derivados podrán ser generados en caso de que se lo requiera, o incluso nuevos datos espaciales a partir de los disponibles en las distintas fuentes de datos de entrada.

Esta etapa requiere un esfuerzo considerable y posiblemente varias iteraciones, por lo que es recomendable implementar todo el proceso mediante una metodología apropiada, tal como CRISP-DM [24] o MoProPEI [25].

2.2. Definición de Vecindarios

Como se ha mencionado anteriormente, los patrones encontrados en el conjunto de datos pueden poseer una alta probabilidad de ocurrencia en solo algunos lugares del espacio contemplado, por lo que una búsqueda local o regional debe ser realizada [7].

Para esto se propone la definición de particiones del conjunto de datos, a las que denominaremos *vecindarios*, y la ejecución de la búsqueda de asociaciones dentro de cada una de ellas. Estos vecindarios pueden ser definidos previamente, correspondiéndose a características del dominio del problema, o bien pueden ser definidos mediante técnicas de segmentación o *clustering* espacial.

En el presente trabajo se recomienda la utilización de algoritmos de agrupamiento o bien basados en densidad o bien basados en distancia sobre los atributos espaciales. Esta preferencia está fundamentada en la Primera Ley de la Geografía, la cual establece que existe una mayor relación entre los objetos espaciales que se encuentran cercanos entre si [26–29]. Opciones a ser utilizadas en la implementación del proceso son los algoritmos basados en densidad de la familia DBSCAN [30–35].

2.3. Modelado Gráfico de Relaciones Espaciales

Una vez generados los vecindarios espaciales es necesario explicitar por cada uno de ellos las relaciones espaciales entre los objetos espaciales objetivo y los objetos de referencia. En función del dominio del problema distintos tipos de relaciones pueden ser consideradas [6]: (i) Relaciones Geométricas, si están basadas en la distancia euclidiana; (ii) Direccionales, basadas usualmente en el concepto de vecindario en grillas, utilizándose el modelos de 8 direcciones (Norte, Noreste, Este, Sudeste, Sur, Suroeste, Oeste, Noroeste); (iii) Topológicas, independintes del concepto de distancia y aplicables a cualquier tipo de objeto espacial, se utiliza normalmente el modelo de 9-intersecciones [36–38]; (iv) Híbridas, si responden a múltiples categorías anteriores, como el paralelismo y la perpendicularidad.

Para este modelado se hace uso de la teoría de grafos por su estrecha relación con los predicados y el tipo de patrón que se busca [11] (Figura 2). Sea entonces

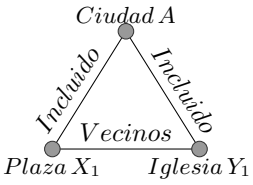


Figura 2. Ejemplo simplificado del modelado de relaciones espaciales mediante multigrafos. Se puede apreciar una instancia del patrón de asociación mencionado en la sección 2.

un multigrafo sin bucles $G = (V, E, L, K, \phi, \mathrm{P}_l, \mathrm{P}_k)$ en donde V es un conjunto de vértices que se corresponden con las instancias de los objetos espaciales, E es un conjunto de aristas que se corresponden con las relaciones entre los objetos espaciales, L es un conjunto de características o valores de los atributos de los objetos espaciales, K es un conjunto de características o atributos de las relaciones espaciales, $\phi : E \rightarrow \{\{v_1, v_2\} | v_1 \in V \ \wedge v_2 \in V \ \wedge v_1 \neq v_2\}$ es la función de incidencia de las relaciones espaciales sobre los objetos espaciales, $\mathrm{P}_l \subseteq V \times L$ es la relación de etiquetado de características de los objetos espaciales, y $\mathrm{P}_k \subseteq E \times K$ es la relación de etiquetado de características de las relaciones espaciales.

De esta manera, al poseer múltiples aristas por cada par de vértices, se pueden modelar distintos tipos de relaciones, que a su vez pueden ser etiquetadas con diversos atributos y, de la misma manera, varios atributos de los objetos espaciales pueden ser tenidos en cuenta. Por otro lado, si la situación lo amerita, estas relaciones pueden ser explicitadas mediante el uso de lógica difusa [12].

2.4. Minado de Subgrafos Frecuentes

Con el propósito de extraer las asociaciones en el espacio con alta tasa de ocurrencia se propone la búsqueda de subgrafos frecuentes dentro de los grafos asociados a cada vecindario espacial. Dado un multigrafo $G = (V, E, L, K, \phi, \mathrm{P}_l, \mathrm{P}_k)$ como el descrito en la sección anterior, el problema de la búsqueda de subgrafos frecuentes en un único gran grafo consiste en encontrar subgrafos $G_i \subset G$ que sean recurrentes en G o, dicho de otra manera, que se repitan múltiples veces dentro del grafo original (Figura 3). Es necesario remarcar que para que un grafo sea isomorfo con otro, las etiquetas de los nodos y las aristas también deben coincidir.

Estos subgrafos frecuentes representan aquellas relaciones entre tipos de objetos espaciales que poseen una alta tasa de ocurrencia dentro del modelo gráfico de los objetos espaciales, correspondiendo así a asociaciones espaciales. Al poseer múltiples relaciones y etiquetas, diversas situaciones pueden ser exploradas simultáneamente.

Múltiples algoritmos han sido diseñados para la búsqueda de subgrafos frecuentes en grandes grafos, teniendo además diversas formas de medir la relevan-

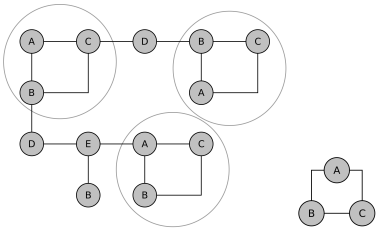


Figura 3. Ejemplo de subgrafo frecuente en un grafo simple sin etiquetas en las aristas.

cia de cada subgrafo frecuente en función de parámetros como la recurrencia o el índice de compresión de información. Algunos ejemplos de algoritmos bien conocidos son IncGM+, FSSG, SUBDUE, HSIGRAM, GREW, entre otros [39–41]

Como resultado de esta fase se obtiene un conjunto de subgrafos frecuentes por cada vecindario, que deberán ser analizados con el fin de obtener conocimiento útil para la toma de decisiones.

2.5. Análisis de Resultados

Como paso final resta el filtrar aquellos subgrafos frecuentes triviales (aquellos no novedosos en el dominio del problema) y traducir los subgrafos frecuentes restantes a predicados compuestos a fin de poder ser utilizados por la inteligencia de negocios.

Este paso puede realizarse de manera automatizada o bien ser realizado por el analista con conocimiento en el dominio del problema o con la ayuda de un experto.

3. Prueba de Concepto

La prueba de concepto efectuada en esta sección pretende ilustrar el funcionamiento del proceso de explotación de información propuesto mediante el uso de distintas herramientas de programación y minería de datos para su implementación.

Como entrada del proceso se dispone de 10 archivos con la ubicación de distintos tipos de edificios de la Ciudad Autónoma de Buenos Aires y alrededores (Bibliotecas, Clínicas, Oficinas de Correo, Instalación Deportiva, Discotecas, Escuelas, Estaciones de Servicio, Iglesias, Museos y Estaciones de Policía). Por cada uno de ellos, en la etapa de preparación de los datos, se generaron mediante la herramienta QGis[1] objetos espaciales puntuales que consisten en dos atributos espaciales con información sobre la Latitud y la Longitud, y un atributo no espacial con el tipo de edificio. Posteriormente se seleccionaron solo aquellos que se encuentran dentro de la Ciudad Autónoma de Buenos Aires para acotar el dominio de búsqueda, resultando 742 puntos (Figura 4, Izquierda).

[1] http://qgis.org/ [10 de julio de 2018]

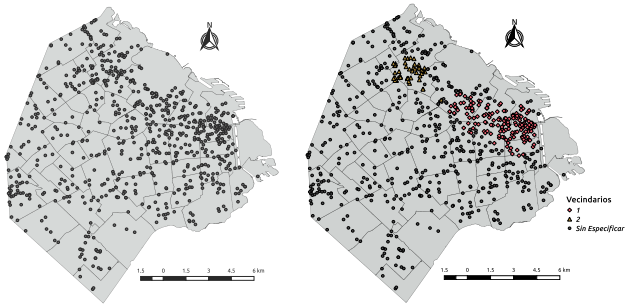


Figura 4. Objetos espaciales generados para la prueba de concepto (Izquierda). Vecindarios generados para la prueba de concepto utilizando HDBSCAN (Derecha).

Posteriormente,en la fase de definición de vecindarios, se utilizó el algoritmo HDDBSCAN [35] de la libreria 'dbscan'[2] de R sobre los atributos espaciales de los datos, genérandose dos vecindarios con un mínimo de 50 puntos en cada uno (Figura 4, Derecha), para reducir el número de vecindarios generados por fines ilustrativos.

Luego, por cada uno de los vecindarios se explicitaron las relaciones geométricas de vecindad entre los objetos espaciales mediante un grafo, generándose un vértice por cada uno de ellos etiquetado con el tipo de edificio al que corresponde, y un arista etiquetado con la etiqueta "Cerca" si la distancia entre los vértices adyacentes resulta menor a 150 metros. Como resultado se obtuvo un grafo con 71 vértices y 45 aristas para el vecindario 1, y un grafo con 15 vértices y 11 aristas para el vecindario 2.

Para la obtención de los subgrafos frecuentes en cada uno de los vecindarios se utilizó el algoritmo SUBDUE implementado en la aplicación Subdue Graph Miner [3], obteniéndose para el vecindario 1 un subgrafo con una tasa de compresión del 15,5 % del que se extrajo el predicado $Correo(X) \wedge Discoteca(Y) \wedge Cerca(X,Y)$, y dos subgrafos del vecindario 2, con una tasa de compresión del 27,2 %, de los que se extrajeron los predicados $Clinica(X) \wedge Correo(Y) \wedge Cerca(X,Y)$ y $Correo(X) \wedge Deporte(Y) \wedge Cerca(X,Y)$.

Como se puede observar, el patrón observado en el primer vecindario no se evidencia en el vecindario número 2, y viceversa, por lo que son patrones que solo poseen una alta probabilidad de ocurrencia local.

4. Conclusión

En el presente trabajo se presenta el diseño de un proceso de explotación de información para el descubrimiento de patrones de asociación en datos espacialmente referenciados. Este proceso resulta suficientemente flexible para poder

[2] https://cran.r-project.org/web/packages/dbscan/dbscan.pdf [10 de julio de 2018]
[3] http://ailab.wsu.edu/subdue/ [10 de julio de 2018]

contemplar multiplicidad y variedad de relaciones espaciales extraidas de diferentes tipos de datos espaciales, haciendo uso de estructuras de grafo para el modelado de las mismas. Se tiene en cuenta además los fenómenos de heterogeneidad y autocorrelación espacial definiendo vecindarios en los cuales se realizará la búsqueda de este tipo de regularidad.

Por otro lado, se presenta una prueba de concepto utilizando datos reales en la cual se pone en evidencia como se realiza la implementación del proceso utilizando distintas herramientas y técnicas para cada una de las etapas propuestas.

En futuros trabajos se hará énfasis en estrategias de implementación de cada una de las fases según el dominio del problema a fin de reducir los tiempos computacionales de ejecución de las actividades para grandes volúmenes de datos con gran cantidad de relaciones entre los mismos. Además se evaluarán distintas estrategias de implementación de extracción de relaciones difusas.

Agradecimientos

La investigación presentada en este artículo fué parcialmente financiada por el Programa de Doctorado para el Fortalecimiento de Áreas I+D+i (2016-2020) de la Universidad Tecnológica Nacional y por el Proyecto de Investigación 80020160400001LA de la Universidad Nacional de Lanús.

Referencias

1. Mennis, J., & Guo, D. “Spatial data mining and geographic knowledge discovery—An introduction.” *Computers, Environment and Urban Systems* 33.6 (2009): 403-408.
2. Koperski, Krzysztof, and Jiawei Han. “Discovery of spatial association rules in geographic information databases.” *International Symposium on Spatial Databases*. Springer, Berlin, Heidelberg, (1995).
3. Rottoli, Giovanni Daián. “Formulación de Procesos para una Ingenieria de Explotación de Información Espacial.” *XXI Ibero-American Conference on Software Engineering*. Doctoral Symposium. Bogotá, Colombia, (2018).
4. Leung, Yee. “Knowledge Discovery in Spatial Data.” Ed. Manfred M. Fischer et al. 1st ed. Vol. 45. Hong Kong: Springer, (2011).
5. Agrawal, Rakesh, and Ramakrishnan Srikant. “Fast algorithms for mining association rules.” *Proc. 20th int. conf. very large data bases*(VLDB). Vol. 1215. (1994).
6. Appice, Annalisa et al. “Discovery of Spatial Association Rules in Geo-Referenced Census Data : A Relational Mining Approach.” *Intelligent Data Analysis* 7 (2003): 541–566.
7. Simpson, Edward H. ”The interpretation of interaction in contingency tables.”*Journal of the Royal Statistical Society*. Series B (Methodological) (1951): 238-241.
8. Saleb, Ansaf, and Christel Vrain. “An Application of Association Rules Discovery to Geographic Information Systems.” *European Conference on Principles of Data Mining and Knowledge Discovery*. (2000). 613–618.

9. Sutjipto, Suci Sri Utami, Imas Sukaesih Sitanggang, and Baba Barus. "Potential Usage Estimation of Ground Water using Spatial Association Rule Mining." *Telecommunication Computing Electronics and Control* (TELKOMNIKA) 15.1 (2017): 504-511.
10. Malerba, Donato, and Francesca A Lisi. "Discovering Associations between Spatial Objects : An ILP Application." *International Conference on Inductive Logic Programming* (2001): 156–163.
11. Malerba, Donato, and Floriana Esposito. "Mining Spatial Association Rules in Census Data." *An International journal for research in official statistics* (ROS) 5.1 (2002): 19.
12. Ladner, Roy, Frederick E Petry, and Maria A Cobb. "Fuzzy Set Approaches to Spatial Data Mining of Association Rules." *Transactions in GIS* 7.1 (2003): 123–138.
13. Bai, Hexiang et al. "A Method for Extracting Rules from Spatial Data Based on Rough Fuzzy Sets." *Knowledge-Based Systems* 57 (2014): 28–40.
14. Zhang, Anshu. "Uncertainty-Based Spatial Association Rule Mining." The Hong Kong Polytechnic University, 2017.
15. Goudarzi, Amir Hossein, and Nasser Ghadiri. "A Hybrid Spatial Data Mining Approach Based on Fuzzy Topological Relations and MOSES Evolutionary Algorithm." *arXiv preprint arXiv:1704.06621* 2017: 1–25.
16. Lee, Ickjai. "Mining Multivariate Associations within GIS Environments." *Innovations in Applied Artificial Intelligence* 3029 (2004).
17. Yang, Hui, Srinivasan Parthasarathy, and Sameep Mehta. "Mining Spatial Object Associations for Scientific Data." *IJCAI International Joint Conference on Artificial Intelligence II* (2005): 902–907.
18. Ding, Wei et al. "A Framework for Regional Association Rule Mining in Spatial Datasets." *Data Mining, 2006. ICDM'06. Sixth International Conference on.* (2006). 851–856.
19. Ding, Wei et al. "A Framework for Regional Association Rule Mining and Scoping in Spatial Datasets." *GeoInformatica* 15.1 (2011): 1–28.
20. Sha, Zongyao, and Xiaolei Li. "Mining Local Association Patterns from Spatial Dataset." *Proceedings - 2010 7th International Conference on Fuzzy Systems and Knowledge Discovery* (FSKD) (2010): 1455–1459.
21. Sha, Z., X. Tan, and Y. Bai. "Localized Spatial Association: A Case Study for Understanding Vegetation Successions in a Typical Grassland Ecosystem." *Communications in Computer and Information Science* 482 (2015): 33–45.
22. Dao, Thi Hong Diep, and Jean Claude Thill. "The SpatialARMED Framework: Handling Complex Spatial Components in Spatial Association Rule Mining." *Geographical Analysis* 48.3 (2016): 248–274. DOI: 10.1111/gean.12094
23. Bogorny, Vania, Paulo Martins Engel, and Luis Otavio Alvares. "GeoARM: An Interoperable Framework to Improve Geographic Data Preprocessing and Spatial Association Rule Mining." *18th International Conference on Software Engineering and Knowledge Engineering* (SEKE) (2006): 79–84.
24. Wirth, Rüdiger, and Jochen Hipp. "CRISP-DM: Towards a standard process model for data mining." *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining.* (2000).
25. Martins, Sebastian, Patricia Pesado, and Ramón García-Martínez. "Information Mining Projects Management Process." *28th International Conference on Software Engineering & Knowledge Engineering* (SEKE). (2016): 504-509. ISBN 1-891706-39-X.
26. Tobler, Waldo R. "Cellular geography." *Philosophy in geography. Springer, Dordrecht.* (1979): 379-386.

27. Miller, Harvey J. "Tobler's first law and spatial analysis." *Annals of the Association of American Geographers* 94.2 (2004): 284-289.
28. Tobler, Waldo. "On the first law of geography: A reply." *Annals of the Association of American Geographers* 94.2 (2004): 304-310.
29. Duan, Jiangli, Lizhen Wang, and Xin Hu. "The effect of spatial autocorrelation on spatial co-location pattern mining." *Computer, Information and Telecommunication Systems (CITS), 2017 International Conference on*. IEEE. (2017).
30. Sander, Jörg, et al. "Density-based clustering in spatial databases: The algorithm gdbscan and its applications." *Data mining and knowledge discovery* 2.2 (1998): 169-194.
31. Smiti, Abir, and Zied Elouedi. "Fuzzy density based clustering method: Soft DBSCAN-GM." *Intelligent Systems (IS), 2016 IEEE 8th International Conference on*. IEEE. (2016).
32. Birant, Derya, and Alp Kut. "ST-DBSCAN: An algorithm for clustering spatial–temporal data." *Data & Knowledge Engineering* 60.1 (2007): 208-221.
33. Zhu, Ye, Kai Ming Ting, and Mark J. Carman. "Density-ratio based clustering for discovering clusters with varying densities." *Pattern Recognition* 60 (2016): 983-997.
34. Sharma, Arvind, R. K. Gupta, and Akhilesh Tiwari. "Improved Density Based Spatial Clustering of Applications of Noise Clustering Algorithm for Knowledge Discovery in Spatial Data." *Mathematical Problems in Engineering 2016* (2016).
35. Campello, Ricardo JGB, et al. "Hierarchical density estimates for data clustering, visualization, and outlier detection." *ACM Transactions on Knowledge Discovery from Data* (TKDD) 10.1 (2015): 5.
36. Egenhofer, Max J., and Robert D. Franzosa. "Point-set topological spatial relations." *International Journal of Geographical Information System* 5.2 (1991): 161-174.
37. Egenhofer, Max J., and John Herring. "Categorizing binary topological relations between regions, lines, and points in geographic databases." The 9.94-1 (1990): 76.
38. Egenhofer, Max J., Eliseo Clementini, and Paolino Di Felice. "Topological relations between regions with holes." *International Journal of Geographical Information Science* 8.2 (1994): 129-142.
39. Abdelhamid, Ehab, et al. "Incremental Frequent Subgraph Mining on Large Evolving Graphs." *IEEE Transactions on Knowledge and Data Engineering* 29.12 (2017): 2710-2723.
40. Kavitha, D., V. Kamakshi Prasad, and J. V. R. Murthy. "Finding Frequent Subgraphs in a Single Graph based on Symmetry." *International Journal of Computer Applications* 146.11 (2016).
41. Bianco, Santiago. "Análisis Comparativo de Algoritmos de Minería de Subgrafos Frecuentes." *Revista Latinoamericana de Ingenieria de Software* 4.2 (2016): 111-142.

D3CAS: un Algoritmo de Clustering para el Procesamiento de Flujos de Datos en Spark

Roberto Molina [1,2] y Waldo Hasperué [1,3]

[1] Instituto de Investigación en Informática (III-LIDI), Facultad de Informática – Universidad Nacional de La Plata
[2] Becario CIN-EVC
[3] Investigador Asociado - Comisión de Investigaciones Científicas (CIC)
rpmolina94@gmail.com, whasperue@lidi.info.unlp.edu.ar

Abstract. En este trabajo se presenta una prueba de concepto de un algoritmo de clustering basado en densidad, denominado D3CAS, el cual fue implementado para ser ejecutado bajo el framework Spark Streaming y que permite el procesamiento de flujos de datos. La principal característica del algoritmo presentado es que es dinámico, es decir selecciona automáticamente el número de clusters del flujo de datos. El algoritmo fue probado datasets de CLUTO, midiendo la calidad de los clusters obtenidos. Los resultados, obtenidos en un ambiente virtualizado, fueron comparados con otro algoritmo de clustering (CluStream), demostrando que D3CAS arroja mejores resultados.

Keywords: Clustering, Spark, Streaming processing.

1 Introducción

Últimamente el avance tecnológico ha conseguido que las organizaciones y empresas puedan generar y almacenar grandes cantidades de datos. Cuando estos datos se generan de manera continua y rápida en el tiempo, siendo imposible almacenarlos en su totalidad, se los denominan 'Flujos de datos' (Data streams).

El procesamiento y análisis de flujos de datos es una característica que se encuentra presente en muchas aplicaciones y sistemas de software de la actualidad. Debido al crecimiento de internet, la automatización de sistemas, el aumento de la conectividad social y el avance en la tecnología, las aplicaciones generan flujos de datos potencialmente infinitos, volátiles y continuos, lo que requiere un procesamiento en tiempo real, simple y rápido.

En muchos casos, estos grandes volúmenes de datos pueden ser minados para obtener información interesante y relevante en una amplia variedad de aplicaciones, como lo son la de detección de intrusiones en la red, flujos de transacciones, registros telefónicos, redes sociales, monitoreo de sensores, internet de las cosas (IoT), en aplicaciones médicas, monitoreo del clima, entre otras.

Una tarea comúnmente realizada sobre flujos de datos es el clustering, la cual consiste en dividir o agrupar la información de tal manera que los datos en cada grupo sean similares y los datos entre grupos sean disímiles. Con flujos de datos potencialmente infinitos, la tarea de clustering es un desafío atractivo ya que los

algoritmos de clustering tradicionales suelen iterar sobre el conjunto de datos más de una vez, es por ello que las restricciones de tiempo de ejecución, uso de la memoria y no poder contar con el conjunto de datos completo deben considerarse cuidadosamente en el contexto del análisis de los flujos de datos. Además, por lo general, la distribución de los datos del flujo cambian continuamente, fenómeno conocido como Concept-drift [1]. Por ello resulta de interés contar con algoritmos de clustering dinámicos, donde la cantidad de clusters en un momento dado dependerá de la distribución de los datos del flujo [2] [3].

En este trabajo proponemos un algoritmo de clustering dinámico para el procesamiento de flujos de datos con funcionamiento distribuido denominado D3CAS (Distributed Dynamic Density based Clustering Algorithm for dataStreams).

El resto del paper se organiza de la siguiente manera. En la sección 2 se lleva a cabo la presentación del modelo de procesamiento de flujos de datos, el estado del arte de las técnicas de clustering sobre flujos de datos y un resumen de las características principales del framework Spark Streaming. En la sección 3, se presenta el algoritmo propuesto. En la sección 4 se presentan los experimentos realizados junto a la comparación con una de las técnicas más importantes dentro de esta área. Finalmente en la sección 5 se presentan las conclusiones y trabajos a futuro.

2 Clustering sobre Flujos de Datos en el Entorno Spark

Spark [4] [5] es un framework distribuido cuya finalidad es ejecutar de manera eficiente operaciones sobre grandes conjuntos de datos, como así también el procesamiento de flujos de datos. Una de las características principales de Spark es la velocidad de ejecución, debido a que tiene la capacidad de ejecutar operaciones sobre los datos en la memoria principal. Spark Streaming es un componente de Spark que permite el procesamiento de flujos de datos, como lo pueden ser datos generados por servidores web o colas de mensajes publicadas por usuarios de un servicio web.

Spark Streaming utiliza una arquitectura denominada "micro-batch", donde el procesamiento del flujo de datos se trata como una serie continua de computaciones sobre pequeños batch de datos, los cuales se crean a intervalos de tiempo regulares. Cada batch de entrada forma un RDD y se procesa mediante operaciones específicas de la API de Spark para generar otros RDD transformados o resultados finales que luego pueden enviarse a sistemas externos.

2.1 El Modelo de Flujos de Datos

En el modelo de flujo de datos, los datos de entrada con los que se van a operar no están permanentemente disponibles para el acceso aleatorio desde la memoria o para recuperar desde un disco, sino que llegan como una o más secuencias continuas de datos temporales [3]. Los flujos de datos son potencialmente ilimitados, lo que lleva a que los datos sean procesado de manera 'on-the-fly', es decir, una vez recibido cierto elemento, este es procesado es ese mismo momento [2] [6].

Existen dos enfoques al momento de procesar un flujo de datos. El aprendizaje incremental donde el modelo evoluciona incrementalmente para adaptarse a los

cambios que presentan los datos entrantes en el flujo [7] y el aprendizaje en dos fases, también conocido como *online-offline learning*, donde la idea básica es tener una primera fase (*online*) la cual va generando un resumen de los datos que llegan en tiempo real. En la segunda fase (*offline*), el proceso de minería se realiza utilizando las síntesis generadas en la etapa online [8].

Como los flujos son potencialmente infinitos, sólo es posible procesar una parte de su totalidad. A la porción de datos sobre la cual se va a trabajar se la define como ventana temporal: $W[i,j] = (x_i, x_{i+1}, x_{i+2}, \ldots, x_j)$, donde i y j son puntos en el tiempo con la propiedad de que $i < j$. En la literatura es posible encontrar diferentes tipos de ventanas de tiempo: *landmark window* [9], *sliding window* [9], *fading window* [10], y *tilted time window* [8], cada una con sus ventajas y desventajas.

El problema de agrupamiento sobre flujos de datos, o *data stream clustering*, es un proceso que divide los datos de forma continua teniendo en cuenta las restricciones de memoria y tiempo. En la literatura, la mayoría de los algoritmos de agrupamiento de flujo de datos usan un esquema de dos fases. Muchos de los algoritmos de agrupamientos de flujo de datos existentes modifican los métodos tradicionales de clustering para usar el enfoque de dos fases propuesto en [8]. Por ejemplo, DenStream [10] es una extensión del algoritmo DBSCAN [11], StreamKM++ [12] es una extensión de k-means++ [13] y StrAP [14] es una extensión de AP.

3 D3CAS

El algoritmo propuesto está basado en las estructuras de datos (micro-clusters) y la metodología de procesamiento online-offline usadas por CluStream [8] y por DenStream [10]. Un micro-cluster, es la estructura de datos que representa y resume a un conjunto de datos proveniente del flujo. Un micro-cluster es una tupla de elementos donde principalmente se almacena: el número total de datos del micro-cluster, la suma lineal y cuadrática de los elementos del conjunto de datos. La metodología de procesamiento online-offline propone separar el proceso de clustering en dos etapas, la etapa online se encarga de generar los micro-clusters y la etapa offline de utilizar los micro-clusters para realizar la tarea de clustering.

El objetivo es implementar un algoritmo que funcione en una arquitectura distribuida, específicamente, bajo el modelo Spark Streaming, lo cual representa la característica diferencial más importante con respecto a las técnicas mencionadas, ya que ambas están diseñadas para trabajar en un ambiente no distribuido.

Para realizar el tratamiento del flujo de datos, se utiliza el modelo de ventana *fadding window* [10], el cual posee la ventaja de aumentar la importancia a los datos más recientes, lo cual facilita la detección temprana de cambios en la distribución de los datos. Para ello, cada dato debe almacenar un atributo que determine su importancia en el tiempo. Este valor es calculado por medio de una función de envejecimiento, en la que el valor de cada dato disminuye exponencialmente a medida que pasa el tiempo (Ec. 1).

$$f(dt) = 2^{-\lambda.dt} \quad \textbf{(1)}$$

donde $\lambda > 0$ y *dt* es la diferencia entre el timestamp de llegada del dato y un instante de tiempo mayor al de llegada, generalmente representado por el instante de tiempo actual. Como puede observarse en la ecuación (1) cuanto mayor sea el valor de λ, menor será la importancia de un dato en comparación con los datos más recientes.

En D3CAS se utiliza el enfoque *online-offline* para el tratamiento del flujo. La etapa online se ejecuta paralelamente en los *workers* de Spark debido a que la tarea de generar los micro-clusters se ajusta al tipo de procesamiento que realizan las operaciones *map* de Spark, mientras que la etapa offline se ejecuta en el nodo *master*, ya que para llevar a cabo la tarea de detección de agrupaciones necesita conocer todos los micro-clusters generados.

3.1 Etapa Online

En esta etapa se utiliza la técnica de micro-clusters, la cual consiste en reducir el volumen de datos a un modelo más pequeño y representativo. Un micro-cluster en un instante de tiempo *t* para un conjunto *P* de *n* puntos de *d* dimensiones $p_1, ..., p_n$ y sus respectivos timestamps $T_1, ..., T_n$, se define como una 7-tupla $MC = \{ LS, SS, w, n, d, t_c, t_m \}$ donde *LS* representa la suma lineal ponderada del grupo de datos, *SS* representa la suma cuadrática ponderada del grupo de datos, *w* representa el peso o importancia del micro-cluster en el tiempo, *n* representa la cantidad de puntos en el conjunto *P*, *d* representa la dimensionalidad de los puntos en el conjunto *P*, t_c representa el timestamp del instante de tiempo cuando fue creado el micro-cluster y t_m representa el timestamp de la última actualización del micro-cluster.

Ante la llegada de un nuevo dato p_i de la partición de datos P_k, se buscan todos los puntos q_j cuya distancia a p_i sea menor a *e* (parámetro del algoritmo que representa el radio del micro-cluster). Si no se encontró ningún punto q_i, entonces se crea un micro-cluster para representar el punto p_i. Si se encontró un conjunto de puntos cercanos a p_i entonces, sea el conjunto *Q* que se encuentra formado por p_i y por el conjunto de puntos cercanos a p_i, se crea un micro-cluster de la siguiente forma: $MC(LS(Q), SS(Q), t_q, cant(Q), cant(Q), timestamp(Q), timestamp(Q))$.

Luego de este proceso, los nodos *workers*, se encargan de enviar el conjunto de micro-clusters generados al nodo *master*, quien será el encargado de realizar el proceso de actualización y llevar a cabo el proceso de detección de clusters.

3.2 Etapa Offline

En esta etapa el proceso de clustering se lleva a cabo mediante la utilización de una técnica basada en densidad, particularmente DBSCAN [11], la cual permite detectar agrupaciones sin especificar cuantas formar, proporcionando la posibilidad de detectar agrupaciones con formas arbitrarias y detectar ruido o outliers presentes en los datos.

Para este proceso se utilizan los centros de todos los micro-clusters generados en la etapa previa cuyo resultado final se ve afectado por dos parámetros pertenecientes a DBSCAN: *eps* y *minpoints*. *eps* representa el radio de un punto *i* en donde buscar sus

puntos vecinos y *minPoints* representa la cantidad mínima de puntos vecinos en el radio *eps* que tiene que tener el punto *i* para considerarse un cluster.

En esta etapa además se lleva a cabo el proceso de actualización del peso de los micro-clusters, ya que recién en esta etapa se tienen todos los micro-clusters del flujo, debido a que los *workers* solo conocen sus respectivos micro-clusters.

La actualización de los micro-clusters se realiza periódicamente y se actualizan los atributos de peso *w* y los atributos *LS* y *SS* que también están influenciados por la importancia del tiempo. Sea *t* el instante de tiempo actual, t_m el timestamp de última modificación del micro-cluster y *dt* la diferencia entre *t* y t_m, entonces cada micro-cluster actualiza sus atributos con las ecuaciones (2) (3) y (4).

$$w = 2^{-\lambda.dt}.w \quad (2)$$

$$LS = 2^{-\lambda.dt}.LS \quad (3)$$

$$SS = 2^{-\lambda.dt}.SS \quad (4)$$

Como puede observarse, dada la característica decreciente de la función de envejecimiento utilizada, los pesos de los micro-clusters van disminuyendo gradualmente. Luego del proceso de actualización del peso, se lleva a cabo el proceso de eliminación de los micro-clusters que se consideran viejos según su peso. Un micro-cluster se considera viejo y debe ser descartado si su peso *w* es menor a un umbral μ, siendo $0 < \mu < 1$.

4 Resultados

Se realizaron dos experimentos. En una primera instancia se probó con flujos de datos, los cuales forman clusters esféricos y bien separados entre sí, utilizando tres datasets generados por el script GENERATEDATA [15] [16], los cuales en el texto son nombrados como 100K10C, 100K40C y 100K14C. El segundo experimento consistió en la detección de clusters en datasets con agrupaciones irregulares.

Para todas las ejecuciones se utilizó el mismo juego de parámetros. En la componente online: $e = 2$ y en la componente offline: *eps*=3.0 y *minPoints* = 4.0, los cuales se corresponden con los parámetros originales de DBSCAN [11]. Para controlar la ventana deslizante utilizada se utilizaron los parámetros $\lambda = 0.25$ y $\mu = 10$ qué son los parámetros que se utilizan generalmente en este tipo de ventanas [10].

Las comparaciones se realizaron contra los resultados arrojados por CluStream [8] utilizando la implementación sobre Apache Spark creado por “HUAWEI Noah's Ark Lab” [17] que se encuentra dentro de su proyecto de minería de flujos de datos StreamDM [18]. Todos los experimentos se realizaron en un ambiente local, en una computadora con un procesador Intel i5 3210M y 8GB de RAM.

4.1 Datasets con clusters esféricos

Una de las características de D3CAS es la detección de clusters de forma dinámica. Para comprobar su correcto funcionamiento se realizó un primer experimento sobre un flujo de datos donde, en distintos instantes en el tiempo, el flujo posee diferentes cantidades de agrupaciones, es decir, con el correr del tiempo, los datos del flujo entrante conforman nuevos clusters.

Para esta prueba se utilizó el dataset denominado 100K10C, el cual está compuesto de 100.000 elementos que se encuentran agrupados en 10 clusters. Se evaluaron los resultados parciales al formarse 4 y 7 clusters y al finalizar la ejecución. En cada evaluación se midió el índice silhouette de cada cluster para evaluar la calidad de los grupos formados.

La tabla 1 muestra los valores de los índices silhouettes correspondientes luego de cada evaluación. Como era de esperarse, ambos algoritmos dan muy buenos resultados, lo cual indica que los clusters detectados por D3CAS son correctos.

Tabla 1. Valores del índice silhouette para las evaluaciones de 4, 7 y 10 clusters.

Número de cluster	Evaluación 1 (4 clusters)		Evaluación 2 (7 clusters)		Evaluación 3 (10 clusters)	
	D3CAS	CluStream	D3CAS	CluStream	D3CAS	CluStream
1	0.95	0.96	0.97	0.95	0.98	0.98
2	0.95	0.96	0.96	0.97	0.98	0.96
3	0.95	0.96	0.95	0.97	0.94	0.98
4	0.95	0.96	0.95	0.95	0.98	0.97
5			0.94	0.97	0.98	0.98
6			0.94	0.96	0.97	0.99
7			0.95	0.98	0.99	0.95
8					0.95	0.97
9					0.94	0.95
10					0.95	0.96

Al elevar el número de clusters por arriba de 10 (usando el dataset 100K40C) es posible observar que D3CAS ofrece mejores resultados con respecto a CluStream. En la figura 1 se puede observar que CluStream confunde clusters distintos agrupándolos en un único cluster, lo que queda evidenciado en la figura 2 donde se observa para CluStream un valor de silhouette promedio de 0.65 mientras que para D3CAS un valor promedio de 0.95. Para un número de clusters alto, D3CAS los detecta correctamente mientras que CluStream no.

4.2 Datasets con Formas Arbitrarias

En este experimento se realiza una comparación utilizando flujos de datos compuestos por datos que forman clusters con formas arbitrarias y aleatorias como por ejemplo se puede encontrar en flujos de datos de monitoreo de tráficos de redes [10] [11].

Para estas pruebas, se utilizan datasets conocidos para este tipo de análisis como los que ofrece el kit de herramientas “CLUTO” [19], desarrollado por el departamento

de Computer Science de la universidad de Minnesota. Los datasets de CLUTO que se van a utilizar son los siguientes:

- cluto-t4-8k: 8.000 datos, 6 clusters irregulares con ruido.
- cluto-t8-8k: 8.000 datos, 8 clusters irregulares con ruido.
- cluto-t7-10k: 10.000 datos, 9 clusters irregulares con ruido.

Para la validación de estas pruebas se utiliza el índice de pureza (Ec. 5).

$$pur = \frac{\sum_{i=1}^{K} \frac{|C_i^d|}{|C_i|}}{K} \tag{5}$$

donde K es el número total de clusters, $|C_i^d|$ representa la cantidad de elementos agrupados en el cluster i por el algoritmo de clustering y $|C_i|$ representa la cantidad total de elementos etiquetados para el cluster i. Por lo tanto, la pureza mide la cantidad de elementos que fueron correctamente agrupados según la etiqueta real del dataset, característica que poseen los datasets de CLUTO.

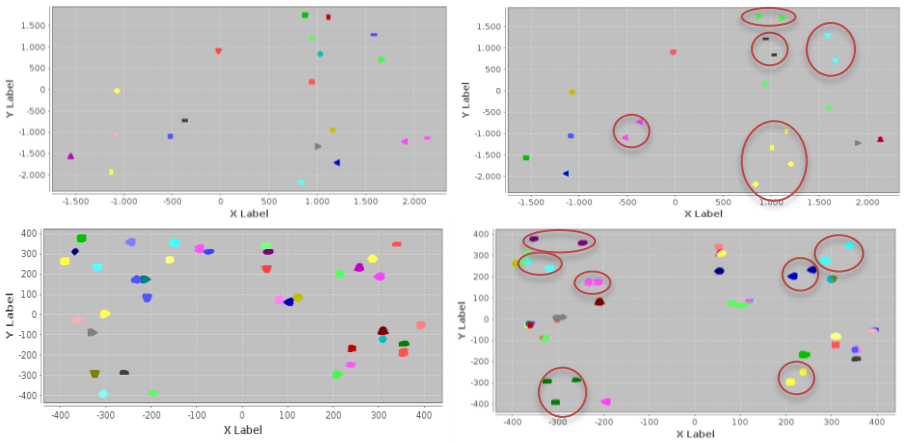


Fig. 1. Clusters hallados al finalizar la ejecución. En la columna de la izquierda los resultado de D3CAS y en la columna de la derecha los de CluStream. La fila de arriba son los resultados con 21 clusters y la fila de abajo corresponde a los resultados luego de presentar los 40 clusters. Los clusters mal agrupados están marcados con un óvalo.

Tabla 2. Valores de pureza para los datasets de CLUTO.

Dataset	D3CAS	CluStream
cluto-t4-8k	90 %	47 %
cluto-t8-8k	93 %	39 %
cluto-t7-10k	92 %	27 %

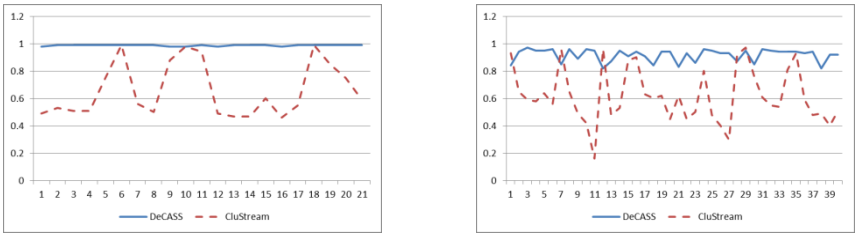


Fig. 2. Índices silouette conseguidos en los clusters hallados por D3CAS y CluStream. A la izquierda el resultado con de 21 clusters y a la derecha el de 40 clusters.

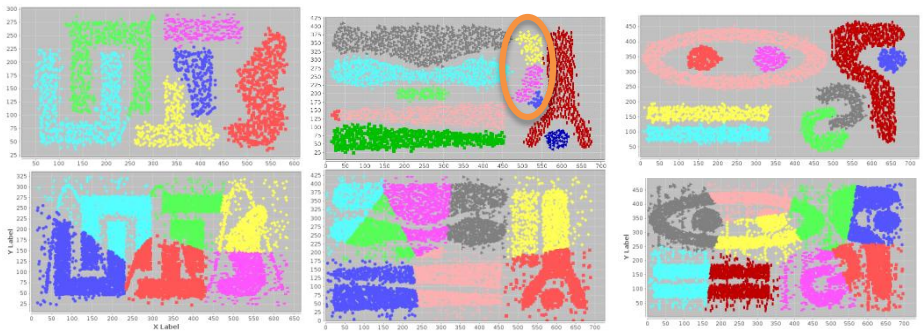


Fig. 3. Resultados hallados por D3CAS (gráficos de la fila superior) y CluStream (gráficos de la fila inferior) para distintos datasets (de izquierda a derecha: cluto-t4-8k, cluto-t8-8k y cluto-t7-10k). La mala agrupación de D3CAS aparece remarcada con un óvalo.

Como se puede ver en la figura 3, D3CAS detecta correctamente los clusters excepto en el caso de cluto-t8-8k, que uno de los grupos los separa en más de un cluster. Se puede observar también que CluStream no consigue detectar correctamente ninguno de los grupos. Todo esto queda reflejado en los resultados de pureza obtenidos para los dos algoritmos que se puede apreciar en la tabla 2. Por lo tanto, es posible afirmar que el D3CAS es capaz de detectar clusters con formas arbitrarias o irregulares y además obtiene mejores resultados que CluStream.

Otra característica de suma importancia para destacar es que D3CAS puede detectar y filtrar los datos que representan ruido, como se aprecia en la figura 3, en estos casos los puntos alejados de los clusters formados no forman parte de ellos. Esta característica no es satisfecha por CluStream, la cual, toma los datos que representa ruido como parte de los clusters.

4.3 Variando el parámetro *e*

En la implementación de D3CAS la etapa online es la encargada de procesar los datos que llegan al flujo y generar los micro-clusters que se utilizaran en la etapa offline. La cantidad de estos micro-clusters depende del parámetro *e*, el cual determina el radio del micro-cluster que representa a los datos originales que se encuentran dentro de su área, por lo que se desprende que a mayor radio menor cantidad de micro-clusters.

Como último experimento se propuso evaluar, para un mismo dataset, cual es la calidad lograda dependiendo del radio del micro-cluster. Para esta prueba se utiliza el dataset denominado 100K14C que posee 100.000 datos distribuidos entre 14 clusters con el objetivo de consumir el flujo con distintos valores de *e* (0.2, 0.25, 0.3, 0.35, 0.5, 1.0, 2.0, 4.0) a fin de generar distintos cantidades de micro-clusters y analizar la variación en los resultados.

Como se ve en la figura 4, utilizando 500 micro-clusters (*e* = 4.0) se obtiene un índice de silhouette de 0.81, lo cual prueba que se está detectando correctamente los distintos grupos, teniendo en cuenta además que 500 micro-clusters representa el 0,5% del dataset completo. Con 25.000 micro-clusters (*e* = 2.0), (25% del dataset), se observa un coeficiente de 0.86, marcando una diferencia de 0.05 con respecto a la prueba de 500 micro-clusters. Esto demuestra que una representación más específica con más micro-clusters no genera una diferencia significativa entre los resultados obtenidos.

Por último, es importante remarcar que es más conveniente tener la menor cantidad de micro-clusters posible para reducir el tiempo de ejecución de la etapa offline ya que, al ser la encargada de generar los modelos, tiene que iterar sobre los micro-clusters generados repetidas veces.

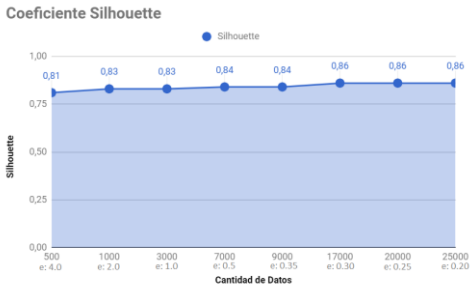


Fig. 4. Índice silhouette hallados por D3CAS para diferentes cantidades de micro-clusters.

5 Conclusión y Trabajo a Futuro

En este trabajo presentamos D3CAS, una prueba de concepto de un algoritmo de clustering dinámico basado en densidad para el procesamiento de flujos de datos que puede ser ejecutado en el framework Spark Streaming.

Los experimentos fueron llevados a cabo en un ambiente virtualizado no distribuido, midiendo únicamente la calidad de los resultados obtenidos sin medir eficiencia en los recursos necesarios durante la ejecución del algoritmo.

Los resultados obtenidos en los experimentos y comparaciones con otro algoritmo de clustering distribuido son alentadores, logrando una buena calidad tanto para los datasets de clusters esféricos como para datasets de formas arbitrarias.

Como trabajo futuro se propone la ejecución en un ambiente distribuido que permitan medir y mejorar el rendimiento del algoritmo presentado, con el fin de realizar comparaciones tanto a nivel de resultados como también a nivel de consumo de recursos como memoria, overhead de comunicación, consumo energético, etc.

Otro aspecto a estudiar es el de ejecutar pruebas en un entorno real, consumiendo un flujo de datos real, como por ejemplo, los flujos de datos que brindan los servicios de Twitter o de redes de sensores.

References

[1] Geoffrey I. Webb, Loong Kuan Lee, Bart Goethals, Franc¸ois Petitjean. *Understanding Concept Drift*. Computer Science. arXiv:1704.00362 (2017)

[2] Aggarwal C.C. *Data Streams: An Overview and Scientific Applications*. In: Gaber M. (eds) Scientific Data Mining and Knowledge Discovery. Springer, Berlin, Heidelberg (2009)

[3] Brian Babcock Shivnath Babu Mayur Datar Rajeev Motwani Jennifer Widom, *Models and Issues in Data Stream Systems*. Department of Computer Science, Stanford University Stanford, CA 94305 (2002)

[4]. Apache Spark. http://spark.apache.org/ . Accedido en 07/2018.

[5]. Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., McCauley, M., Franklin, M.J., Shenker, S., Stoica, I. *Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing*. In: Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation, pp. 1–14. USENIX Association (2012)

[6] Gama J, Rodrigues P. *An overview on mining data streams*. Studies in computational intelligence, Vol. 206 pp 29–45. Springer, Berlin. (2009)

[7] Gepperth, A., Hammer, B.. *Incremental learning algorithms and applications*. European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium (2016).

[8] Charu C Aggarwal, Jiawei Han, Jianyong Wang, and Philip S Yu. *A framework for clustering evolving data streams*. In Proceedings of the 29th international conference on Very large data bases-Volume 29, pages 81–92. VLDB Endowment, (2003)

[9] Zhang P, Zhu X, Shi Y, Wu X *An aggregate ensemble for mining concept drifting data streams with noise*, Lecture notes in computer science, 5476 pp 1021–1029. Springer, Berlin. (2009)

[10] Cao F, Ester M, Qian W, Zhou A. *Density-based clustering over an evolving data stream with noise*. In: Proceedings of the SIAM international conference on data mining, pp. 328–339. (2006)

[11] Ester, M., Kriegel, H.-P., Sander, J., Xu, X. *A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise.* Proceedings of the Second International Conference on Knowledge Discovery and Data Mining Pages 226-231 (1996)

[12] Ackermann, M. R., Märtens, M., Raupach, C., Swierkot, K., Lammersen, C., Sohler, C. *StreamKM++: A clustering algorithm for data streams*. ACM J Exp Algorithmics; 17(1):173–187. (2012)

[13] Arthur, D. and Vassilvitskii, S. *k-means++: The Advantages of Careful Seeding*. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms Pages 1027-1035 (2007)

[14] Zhang X, Furtlehner C, Sebag M. *Data streaming with affinity propagation*. In: ECML/PKDD (2). Berlin: Springer Berlin Heidelberg. p. 628–43. (2008)

[15] GENERATEDATA. https://github.com/FakenMC/generateData. Accedido en 07/2018.

[16] Fachada, N., Figueiredo, M.A.T., Lopes, V.V., Martins, R.C., Rosa, A.C., *Spectrometric differentiation of yeast strains using minimum volume increase and minimum direction change clustering criteria*, Pattern Recognition Letters, vol. 45, pp. 55-61 (2014)

[17] HUAWEI Noah's Ark Lab. http://www.noahlab.com.hk. Accedido en 07/2018.

[18] StreamDM. https://github.com/huawei-noah/streamDM. Accedido en 07/2018.

[19] CLUTO. http://www.cs.umn.edu/~cluto. Accedido en 07/2018.

Replicación de Bases de Datos en Modalidad "maestro-esclavo", caso de estudio: Firebird SQL Server

Cherencio Guillermo Rubén[1], Perello Mario Gerardo[1], Romero Juan Carlos[1]

[1] Universidad Nacional de Lujan, Departamento de Ciencias Básicas

Abstract. Los Sistemas de Gestión de Bases de Datos (SGBD) más evolucionados permiten la programación de desencadenadores o disparadores también llamados triggers, procedimientos y funciones en lenguaje PS/SQL o Transact-SQL. En nuestro caso práctico: Firebird/Interbase® (FB) incorpora el lenguaje de programación PL/SQL, el cual permite la programación de desencadenadores (triggers) y procedimientos almacenados (stored procedures). Como parte del código PL/SQL está disponible la instrucción EXECUTE STATEMENT, la cual permite la ejecución de comandos SQL tanto en forma local como remota. Además este motor de base de datos (FB) trabaja con un algoritmo de confirmación de dos pasadas (two-phase commit, 2PC) para este comando, lo cual es suma importancia porque permite el control de las transacciones distribuidas. De esta forma es posible determinar el éxito o fracaso de la ejecución distribuida de la transacción y de esta manera implementar una replicación de tipo homogénea maestro-esclavo basado en desencadenadores (triggers) con una granularidad de replicación a nivel de tablas. El presente trabajo pretende no solo hacer una replicación maestro-esclavo con Firebird, sino el desarrollo de una aplicación que permita a través de una interface automatizar la implementación de la replicación, generando de forma automática el código y las estructuras necesarias para la replicación y también permitir la posible recuperación en caso de fallas o anomalías. Se abren futuras líneas de investigación sobre nuevas funcionalidades a incorporar a la aplicación propuesta para facilitar la implementación de bases de datos distribuidas en Firebird, fortaleciendo esta funcionalidad no disponible en este motor de base de datos. Esta aplicación está desarrollada bajo licencia LGPL.

Palabras Clave: Java Firebird Interbase 2PC Replicación SQL PL/SQL

1 Introducción

En el marco del proyecto de investigación sobre la replicación de datos distribuidos denominado *"Replicación maestro-esclavo en bases de datos relaciones no distribuidas",* usando motores de bases de datos de libre distribución que no incorporan esta posibilidad seleccionamos el caso de Firebird/Interbase®. Este Sistema Gestor de Base de Datos[1] (SGBD) está basado en el modelo relacional [2], el cual posee una arquitectura de tipo cliente-servidor y está disponible para diferentes

plataformas. En nuestro caso haremos replicación tanto en plataforma Windows® como en Linux. Posee una serie de librerías (API) que ha permitido el desarrollo de *drivers (controladores de acceso)* para distintos entornos de desarrollo (lenguajes), dando la posibilidad de realizar aplicaciones *client-server* en entornos no distribuidos. A partir de la versión 2.5 incorpora el algoritmo 2PC[3] en la ejecución de comandos de tipo EXECUTE STATEMENT[4] haciendo posible la ejecución de comandos SQL en forma remota, indicando los datos para gestionar la conexión[1] es posible ejecutar el comando SQL en otro servidor Firebird/Interbase® y luego verificar el resultado de la transacción distribuida. De esta forma, es factible la implementación de un sistema de base de datos distribuido homogéneo[2] basado en desencadenadores (*triggers*) que permita la réplica de tablas en varios servidores Firebird/Interbase® bajo la modalidad del tipo maestro-esclavo. Es importante destacar que existen otras posibilidades de replicación, no obstante, el presente trabajo solo se limita a este enfoque de replicación.

Si bien existen múltiples soluciones de replicación para bases de datos de código abierto, podría citarse por ejemplo a Postgresql[5] ; no obstante, éste no es el caso de Firebird/Interbase®[3] y esto ha motivado el desafío del presente trabajo y otros desarrollos para el futuros en este lineamiento.

Replicar una tabla bajo este enfoque implica el desarrollo e implementación de un desencadenador o *trigger* que tome los eventos de actualización de la tabla como ser *insert, update, delete* y replique el cambio a todos los servidores esclavos. Por ej. sobre una tabla relativamente sencilla[4], se requirieron 75 líneas de código PL/SQL, de las cuales solo 5 líneas son de código específico para dicha tabla. Esto significa que en un escenario de implementación concreta y real, se requerirían miles de líneas de código no generalizables, dependientes de cada tabla, con alta tasa de probabilidad de error.

Se ha propuesto desarrollar una aplicación que permita generar todo el código PL/SQL necesario para implementar el esquema de replicación. La aplicación se ha denominado FBJReplicator[5] y está disponible en el portal SourceForge[6] bajo licencia LGPL.

1 Indicar: Servidor, usuario, contraseña y rol. No permite indicar puerto TCP/IP de conexión.

2 Los servidores deben ser Firebird/Interbase®, el comando no permite la ejecución remota en servidores de otros fabricantes.

3 Desde la página oficial Firebird/Interbase® http://www.firebirdsql.org/en/third-party-tools/ pueden vincularse sólo 2 alternativas, ambas para Windows®, una de ellas, posiblemente discontinuada y la otra es una aplicación propietaria. Otro caso es IBReplicator https://www.ibphoenix.com/products/software/ibreplicator, posiblemente la opción más evolucionada, vinculada al grupo de desarrollo de Firebird/Interbase® y con licencia comercial.

4 Una tabla con solo dos atributos, con una clave primaria simple (no compuesta).

5 Disponible en http://sourceforge.net/projects/fbjreplicator/

6 http://www.sourceforge.net

2 Recursos del trabajo y método utilizado

El presente trabajo se llevó adelante siguiendo una serie de etapas, tomando los lineamientos generales de Blanchette[6]:

1. Requisitos
2. Implementación de forma manual de la arquitectura propuesta.
3. Diseño y desarrollo de la aplicación.
4. Implementación de la aplicación y bases de datos de ejemplo. Testing.
5. Empaquetado, publicación y distribución.

2.1 Requisitos

La aplicación propuesta tiene los siguientes requisitos[7]:

2.1.1 Multiplataforma.
2.1.2 Libre distribución bajo licencia LGPL.
2.1.3 Interfaz gráfica de usuario con soporte multilingüe.
2.1.4 Estructuras de datos de replicación externas a base de datos maestra o esclavo.
2.1.5 Generación de *scripts* de replicación y/o de instalación/desinstalación de replicación en servidor maestro y esclavos.
2.1.6 Posibilidad de *activar y desactivar* la replicación (ya sea por tabla replicada o por servidor).
2.1.7 *Log de transacciones* para dar al sistema tolerancia a *fallos* y formas de recuperación en caso de caídas de servidores esclavos.
2.1.8 Posibilidad de chequear *en línea* la replicación.

2.2 Implementación de forma manual de la arquitectura propuesta.

Si bien Firebird/Interbase® brinda buena documentación, uno de los principales problemas fue contar con información técnica detallada del comando EXECUTE STATEMENT y la implementación del algoritmo 2PC en Firebird/Interbase® 2.5. Como forma de suplir este problema, se realizaron pruebas con 3 servidores Firebird/Interbase® 2.5, generando una base de datos maestra de testing y replicando una tabla de esa base de datos maestra en los restantes servidores que actuarían como esclavos. Para este caso el código PL/SQL de replicación fue generado manualmente, tratando de escribirlo lo más genérico y reutilizable posible, para tratar de ayudar a su automatización futura.

2.3 Implementación en la base de datos maestra

Se implementaron las estructuras básicas para la tolerancia a fallos y su forma de recuperación, los desencadenadores (*triggers*) implementados en la base de datos maestra se apoyan en las siguientes relaciones:

tbl_config_server (**id**, nserver, ndb, nuser, npwd, nrole, active)

tbl_config_table (**id, ntable**, ttable, ninsert_fmt, nupdate_fmt, ndelete_fmt, active)

7 Se plantean los requisitos generales, para ser considerados en el diseño y que los mismos puedan implementarse si el proyecto cuenta con los tiempos y recursos necesarios.

tbl_log (**nid**, nserver, ntable, nsql, ndb, nuser, npwd, nrole, ntr, napply, nsqlerror, ngdscode)

La relación *tbl_config_server* tendrá una tupla por cada servidor esclavo que replique una o más tablas. La relación *tbl_config_table* tendrá una tupla por cada tabla a replicar en un determinado servidor esclavo (vinculado con la clave extranjera *id* a *tbl_config_server*). Los atributos *nserver, ndb, nuser, npwd, nrole* se refieren a la identificación del servidor Firebird/Interbase® 2.5, la base de datos, el usuario, la contraseña y el rol del usuario respectivamente. El atributo *active* permite activar y desactivar el servicio de replicación (requisito 2.1.6). Los atributos indicados con negritas y subrayados indican la clave primaria de cada relación. El atributo *ntable y ttable* se refieren a la tabla origen a replicar y tabla destino respectivamente[8]. La tabla *tbl_log* (requisito 2.1.7) guarda los datos de conexión utilizados al momento de ejecutar el comando sql remoto (*nsql)*, si el mismo fue satisfactorio o no (*napply*); en caso de error, se guardan los códigos de error que retorna el motor Firebird/Interbase® (*nsqlerror, ngdscode*; requisito 2.1.8), guarda el número de transacción (*ntr*) el cual se usará en caso de recuperación de un servidor esclavo caído.

Se debe implementar un desencadenador (*trigger*) por cada tabla a replicar, cuyo pseudocódigo para la tabla "*A"* es el siguiente:

```
inicio

config = tbl_config_server ⋈ tbl_config_table

R = σ active = 's' y ntable = 'A' (config)
para cada tupla en R hacer:
  si insertando entonces sql = ninsert_fmt
  si borrando entonces sql = ndelete_fmt
  si cambiando entonces sql = nupdate_fmt
  reemplazar en sql los valores de los atributos de la tabla 'A'
  ntr = numero de transacción actual
 ejecutar p_apply(ntr,sql) en servidor remoto
  si ejecucion ok entonces
     apply='si'
  sino
     apply='no'
     guardo códigos FB de error
  fin si
  grabo tupla en tbl_log
fin para
fin
```

8 Como ejemplo, una tabla denominada "A" podría replicarse con otra denominación "B" en un servidor esclavo.

Se deben tener los datos de configuración de la tabla en cuestión, junto con los servidores esclavos que la replican; el código SQL a ejecutar se apoya en 3 máscaras de formato del comando para hacer *insert* (*ninsert_fmt*), *update* (*nupdate_fmt*) y *delete* (*ndelete_fmt*) respectivamente[9] y se realiza una ejecución remota del procedimiento almacenado *p_apply* que recibe como argumento el código SQL a ejecutar y el número actual de transacción. Por último, se guarda el resultado de la ejecución en la tabla de log *tbl_log*.

2.4 Implementación en base de datos esclava

La tabla *tbl_tr(tr)* contiene una sola tupla que guarda el último número de transacción aplicado en la base de datos (requisito 2.1.7). Por cada tabla replicada se implementó un desencadenador (*trigger*) de *before insert, update, delete* para no permitir cambios en la tabla, a menos que se trate del usuario de replicación o bien el administrador del sistema[10]. Por último, el procedimiento *p_appy(ntr,sql)* ejecuta el comando *sql* y actualiza *tbl_tr* con *ntr* (número de transacción enviado desde base de datos maestra). De esta forma, es posible "poner al día" un servidor esclavo que estuvo fuera de servicio; a partir de los datos guardados en las relaciones *tbl_log, tbl_tr*.

2.5 Diseño y desarrollo de la aplicación.

Teniendo en cuenta los requisitos 1.1, 1.2, 1.3; junto con la disponibilidad del controlador JDBC Jaybird[11] para Firebird/Interbase®, se optó por un desarrollo en Java SE con una interfaz gráfica desarrollada en Swing[12]. Acorde con los requisitos 1.3, 1.4, el sistema utilizará archivos de propiedades para guardar en los mismos los mensajes en distintos idiomas, así como también los datos de configuración de la replicación.

Optamos por una interfaz gráfica que permita seleccionar el servidor maestro, luego seleccionar los servidores esclavos, las tablas a replicar por cada servidor esclavo; guardar todos los datos asociados a la replicación en archivos de propiedades y permitir al usuario instalar la replicación; ya sea generando un *script* de replicación o bien conectándose a los distintos servidores y ejecutando los comandos DDL[13] que correspondan.

La aplicación es orientada a objetos, posee abstracciones simples y fáciles de comprender [7] tales como las siguientes clases:

Clase	Descripción
MainView	Interfaz gráfica

9 Se debe a que no fue posible generalizar una regla de generación de código, ya que el código sql de insert, update y delete cambia acorde con los distintos tipos de datos de los atributos de la tabla.

10 El modelo propuesto de replicación, todos los servidores esclavos son sólo de consulta, no se permiten cambios en tablas replicadas.

11 Disponible su descarga en http://www.firebirdsql.org/en/jdbc-driver/

12 Framework Java para el desarrollo de interfaz gráfica.

13 Data Definition Language, lenguaje de definición de datos para la creación, destrucción y modificación de objetos dentro de un SGBD.

FileIni	Manejo de archivos de propiedades
FBCon	Manejo de conexión Firebird/Interbase®, ejecución de consultas y comandos DDL
StoredProcedure	Permite ejecución de procedimientos almacenados
SqlTable	Representación de una tabla en una base de datos Firebird/Interbase®
SqlTableColumn	Representación de una columna de una tabla en una base de datos Firebird/Interbase®
SqlTrigger	Representación de un desencadenador (*trigger*) asociado a una tabla en una base de datos Firebird/Interbase®
SwingFactory	Creación de componentes swing utilizados en interfaz gráfica
ViewHtml	Ventana para la visualización genérica de contenido html

Tabla 1. Clases proyecto FBJReplicator

El controlador JDBC Jaybird ha permitido realizar sobre Firebird/Interbase® todas las operaciones que requirió esta aplicación. Solo se encontraron dificultades en cuestiones vinculadas con la administración de los servidores[14].

2.6 Implementación de la aplicación y bases de datos de ejemplo. Testing

La aplicación fue probada sobre el mismo conjunto de bases de datos y servidores que fueron utilizados en la implementación de forma manual de la arquitectura propuesta; generando la replicación en forma automática y verificando que los resultados obtenidos eran coincidentes con la implementación manual.

2.7 Empaquetado, publicación y distribución.

La aplicación desarrollada se denomina como proyecto *"FBJReplicator"* el cual se distribuye bajo licencia LGPL[15] . Se incluyen los programas fuentes, ejemplos, documentación, etc. en una estructura de directorios comprimidos en formato *zip* que los usuarios pueden descargar, descomprimir y compilar utilizando un compilador Java 1.7 o superior sobre la carpeta principal. El software está configurado para trabajar en idioma español e inglés. La documentación está escrita en español. La distribución y publicación de este trabajo se realiza a través del portal SourceForge. Este portal también posibilitó el trabajo en forma colaborativa entre los miembros de esta investigación y colaboradores, a través de su servidor SVN[16].

3 Resultados

Hasta el momento no se han obtenido resultados de implementaciones concretas en entornos de producción, teniendo en cuenta esta como la primera versión del

14 Implementación de "hot-backups", administración de usuarios en forma remota.
15 http://www.gnu.org/licenses/lgpl.html
16 Sistema de control de versiones Apache Subversión http://subversion.apache.org/

sistema[17], todas las pruebas realizadas han sido en entornos denominados no productivos. Se usaron redes locales y pruebas en un mismo computador con servidores esclavos virtualizados. Se han realizado las pruebas tanto en servidores Firebird/Interbase® 2.5 bajo plataforma Windows® como Linux. Los resultados obtenidos son alentadores, han permitido la depuración de varios errores y se han establecido tiempos razonables de replicación para un entorno distribuido de tipo maestro-esclavo.

Se requiere de administración o bien del desarrollo de mecanismos de auto-administración, en especial, cuando se registran caídas de servidores esclavos, puede observarse la lentitud en las actualizaciones de datos en tablas replicadas debido a la imposibilidad de confirmación de la transacción a través de 2PC. Esta situación puede mejorarse fácilmente con la incorporación de un mecanismo que permita desactivar una replicación luego de, por ejemplo, una cantidad *N* fallas seguidas; de esta forma, la arquitectura podría funcionar más rápido y luego "poner al día" el servidor que nuevamente entra en servicio.

4 Discusión

Con los resultados que se obtuvieron se puede decir que los mismos alientan a pruebas en entornos de producción, no obstante, es requisito la completa implementación del sistema que incluya: *administración* de usuarios de replicación, *backup* de servidores, *activación/desactivación* de replicación (manual y/o automática), *notificaciones*, *monitoreo, borrado/depuración* de log de transacciones, etc.

Es importante considerar las limitaciones de la implementación de este modelo de replicación. Por ejemplo, no es posible replicar utilizando comandos SQL que incluyan llamadas a funciones o expresiones que podrían arrojar resultados distintos en los distintos servidores, a lo largo del tiempo en que transcurre la transacción distribuida[18]. Ante la caída o fallo de un servidor esclavo, solo es posible "ponerlo al día", es decir, hacer un *redo* a partir de la información del log de transacciones, pero no es posible hacer un *undo* de los servidores[19]; pues el log de transacciones sólo almacena los comandos SQL ejecutados y no los datos almacenados antes o después de la ejecución del comando SQL.

Comparando con los productos disponibles para PostgreSQL, alientan a futuros desarrollos en este sentido para Firebird/Interbase®, por ejemplo, la creación de un

17 Al momento de escribir este artículo, el sistema sólo ha implementado la funcionalidad mínima de replicación para los tipos de datos básicos de las columnas de las tablas Firebird/Interbase® 2.5. Se requiere contar con otras herramientas adicionales para hacer pruebas productivas, como por ejemplo, la administración de usuarios y backups de los servidores.

18 Para citar un caso, la expresión CURRENT_TIMESTAMP no podría utilizarse dentro de un comando sql de replicación, en vez de ello, debería asignarse ese valor y pasar el valor a todos los servidores esclavos.

19 En este sentido, sólo es posible hacer un *rollback* de la transacción distribuida actual (por ejemplo, porque la misma no fuese confirmada -al menos- por uno de los servidores esclavos); pero una vez confirmada, ya no sería posible su *rollback*.

servidor de tipo middleware (*statement-based replication middleware*), replicación multimaster sincrónica o asincrónica, particionado de datos, ejecución paralela de queries[5], etc.

5 Conclusión

Es apreciable una falencia en Firebird/Interbase® en cuanto a documentación y productos evolucionados y bien probados que implementen reparto y distribución de datos; no existen para Firebird/Interbase® opciones tan variadas como las ofrecidas para PostgreSQL.

La aplicación propuesta se considera un aporte a las opciones de replicación disponibles para Firebird/Interbase®, alienta a su comunidad de usuarios a la implementación de sistemas distribuidos que permitan la evolución y desarrollo de este sistema, favorecidos por el tipo de licencia elegido y la posibilidad de continuar trabajando desde Sourceforge.

Esta aplicación requiere de la implementación de herramientas adicionales para ser usado en un sistema de producción real, no obstante, los resultados obtenidos con las opciones básicas ya implementadas son muy alentadoras e indican su factibilidad de implementación en un sistema productivo real.

Agradecimientos
Al Mg. Jorge Peri, al Ing. Luis H. Perna, al Lic. Pedro S. Asis, al Ing. Bono Juan, al Ing. Colussi Diego por facilitar los medios y recursos necesarios para el armado de este proyecto.

Referencias

1. Castaño Miguel, Adoración de; Velthuis Piattini, Mario; Martinez, Esperanza Marcos, Diseño de Bases de Datos Relacionales, Editorial Ra-Ma, Madrid, 2000, ISBN 84-7897-385-0, Pag 5.
2. Silberschatz, Abraham; Korth, Henry F.; Sudarshan S., Fundamentos de Bases de Datos, 4ta.Ed., Ed. Mc Graw Hill, Madrid, 2002, ISBN: 0-07-228363-7, Pag. 17.
3. Rahimi, Saeed K.; Haug, Frank S, Distributed Database Management Systems. A practical approach, John Wiley & Sons Inc. Publication, IEEE Computer Society, 2010, ISBN: 978-0-470-40745-5, Pag. 329-345.
4. Vinkenoog, Paul et al., Firebird 2.5 Language Reference Update, Oct 8 2011, disponible en http://www.firebirdsql.org/file/documentation/reference_manuals/reference_material/html/langrefupd25.html
5. PostgreSQL 9.1 Manual, Chapter 25. High Availability, Load Balancing, and Replication, disponible en http://www.postgresql.org/docs/9.1/static/high-availability.html
6. Blanchette, Jasmin, The Little Manual of API Design, Trolltech, a Nokia company, June 19, 2008, disponible en http://www4.in.tum.de/~blanchet/api-design.pdf
7. Jackson, Daniel, Software Abstractions, MIT Press, 2006.

Datos de Contacto:
Cherencio, Guillermo Rubén. Universidad Nacional de Lujan, Departamento de Ciencias Básicas. Rutas 5 y 7, Lujan, CP 6700, Provincia de Buenos Aires, República Argentina. E-mail: grchere@yahoo.com.
Perello, Mario. Universidad Nacional de Lujan, Departamento de Ciencias Básicas. Rutas 5 y 7, Lujan, CP 6700, Provincia de Buenos Aires, República Argentina. E-mail: mperello04@gmail.com.
Romero, Juan Carlos. Universidad Nacional de Lujan, Departamento de Ciencias Básicas. Rutas 5 y 7, Lujan, CP 6700, Provincia de Buenos Aires, República Argentina. E-mail: juancarlosjromer@gmail.com.

Segmentación de Mercado Usando Técnicas de Minería de Datos en Redes Sociales

Eduin Olarte[1], Marisa Panizzi[1], Rodolfo Bertone[2]
[1] Escuela de Sistemas. Universidad Argentina John F. Kennedy.
Bartolomé Mitre 1411, Ciudad Autónoma de Buenos Aires (C1037ABA), Argentina
[2] Instituto de Investigaciones en Informática LIDI. Facultad de Informática. UNLP – CIC
eduinolarte@gmail.com; marisapanizzi@outlook.com; rbertone@lidi.unlp.edu.ar

Resumen. Las redes sociales han ganado una gran popularidad durante la última década gracias al avance de nuevas tecnologías y al creciente interés de las personas por generar contenidos y compartirlos con sus contactos. Esto hace que los datos generados en las redes sociales crezcan exponencialmente con el tiempo.
Estos datos generados contienen información que se puede ser analizada con el fin de descubrir patrones que ayuden en múltiples disciplinas. El marketing es una de estas disciplinas que está estrechamente ligada a entender comportamientos, tendencias o gustos de las personas
El objetivo de este trabajo consiste en la aplicación de minería de datos (MD) para lograr el descubrimiento de patrones en datos provenientes de redes sociales. A partir de la obtención de patrones se busca realizar distintos tipos de segmentaciones que ayuden a los profesionales de marketing a orientar sus campañas.

Palabras clave: Explotación de información, Minería de datos, Segmentación de mercado, Redes Sociales, Aprendizaje Automático.

1 Introducción

El marketing es un concepto que surge para dar respuesta a la pregunta que se hacen a menudo las organizaciones: ¿a quiénes intentamos vender? El marketing es un proceso social por el cual los grupos e individuos satisfacen sus necesidades al crear e intercambiar bienes y servicios [1]. También es un proceso que comprende la identificación de necesidades, relaciones y deseos del mercado que se tiene como objetivo.

El mercado es demasiado amplio y está formado por clientes diferentes con diversas necesidades. Cada empresa busca centrarse en los individuos, de acuerdo a su capacidad, gusto y segmentos etarios. Con este fin, las empresas realizan estrategias de mercado. La estrategia de dividir el mercado en grupos homogéneos se conoce como segmentación. La segmentación de mercado consiste en dividir un mercado en grupos más pequeños de necesidades, características o comportamientos distintos que podrían requerir productos separados o mezclas de marketing [2].

Tradicionalmente, los datos para realizar las segmentaciones provienen de procesos de recolección tales como encuestas, historial de las ventas, etc. Con el advenimiento de las redes sociales se genera una nueva fuente de información, las redes sociales [3]. Según mediciones internacionales se estima que por minuto se generan más de cuatrocientos mil Tuits los cuales representan una gran cantidad de datos para ser analizados [4]. En la actualidad las redes sociales, proveen privilegios a los desarrolladores para acceder a la información fácilmente por medio de interfaces Web.

Este gran conjunto de datos puede ser analizado para identificar patrones y relaciones. Este análisis puede realizarse con minería de datos. La minería de datos es un campo multidisciplinario que combina estadísticas, aprendizaje automático, inteligencia artificial y tecnología de bases de datos (Figura 1) para establecer relaciones y encontrar patrones en grandes conjuntos de datos [5]. La identificación de patrones y relaciones permite lograr segmentaciones de distintos tipos y dar solución a las empresas.

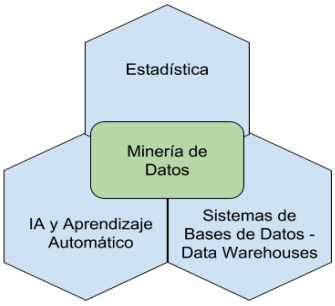


Fig. 1. Disciplinas científicas que comprenden a la Minería de Datos.

La minería de datos es la búsqueda de información nueva, valiosa y no trivial en grandes volúmenes de datos. Es un esfuerzo cooperativo de humanos y computadoras. Los mejores resultados se obtienen al equilibrar el conocimiento de los expertos humanos al describir problemas y objetivos con las capacidades de búsqueda de las computadoras.

En la práctica, los dos objetivos principales de la minería de datos son la predicción y la descripción. La predicción implica el uso de algunas variables o campos en el conjunto de datos para predecir valores desconocidos o futuros de otras variables de interés. La descripción, por otro lado, se centra en encontrar patrones que describen los datos que pueden ser interpretados por los humanos.

Los objetivos de la predicción y la descripción se logran mediante el uso de técnicas de extracción de datos. Basado en el tipo de patrones que se buscan, las técnicas de minería de datos pueden ser clasificación, clusterización, asociación y/o análisis de tendencias [6].

2 Análisis de Metodologías para proyectos de MD y herramientas de trabajo

En esta sección, se resume el análisis comparativo de las metodologías para proyectos de minería de datos. Para su análisis se han considerado KDD "Descubrimiento de conocimiento en bases de datos" [7], SEMMA (Sample, Explore, Modify, Model, and Assess [8] y CRISP-DM (Cross Industry Standard Process for Data Mining) [9]. Posteriormente se sintetiza el análisis comparativo de las herramientas de trabajo para los procesos de minería de datos.

2.1 Estudio comparativo de las metodologías para proyectos de minería de datos.

El proceso de minería de datos a menudo se caracteriza como un proceso iterativo de múltiples etapas que incluye la selección de datos, la limpieza de datos, la aplicación de algoritmos de minería de datos, la evaluación, entre otros. Se realizó un estudio comparativo de las metodologías con el propósito de identificar si las mismas soportan etapas equivalentes (tabla 1).

Tabla 1. Estudio comparativo de metodologías para proyectos de minería de datos.

Etapas	KDD	SEMMA	CRISP-DM
Objetivos	Pre KDD		Entendimiento de negocio
			Entendimiento de los datos
Pre-procesamiento	Selección	Muestra	Preparación de datos
	Preprocesamiento	Exploración	
	Transformación	Modificación	
Modelo	Minería	Modelado	Modelado
Resultados	Interpretación / Evaluación	Valoración	Evaluación
Despliegue	Pos KDD		Instalación

La metodología KDD cuenta con nueve etapas para el modelo de proceso, CRISP-DM cuenta con seis y SEMMA con cinco etapas. Al examinar los tres modelos de procesos de minería de datos, puede observar que en cierto punto son equivalentes entre sí. Comparándolos, se puede ver:

- El paso del proceso KDD "Desarrollo y comprensión del dominio de la aplicación" se puede identificar con la fase "Comprensión comercial" del proceso CRISP-DM.
- Los pasos del proceso KDD "Creación de un conjunto de datos objetivo" y "Limpieza y pre procesamiento de datos" pueden identificarse con las etapas "Muestra" y "Explorar" de SEMMA, respectivamente, y / o pueden identificarse con la fase "Conocimiento de datos" de CRISP-DM.

- La etapa de proceso KDD "Transformación de datos" se puede identificar con la etapa "Preparación de datos" de CRISP-DM y la etapa "Modificar" del proceso SEMMA, respectivamente.
- Las tres etapas de KDD "Elección de la tarea de minería de datos adecuada", "Elección del algoritmo de minería de datos adecuado" y / o "Algoritmo de utilización de minería de datos" se pueden identificar con la fase de "Modelado" de CRISP-DM y / o "Modelo" etapa del proceso SEMMA respectivamente.
- El paso del proceso KDD "Interpretación de patrones minados" se puede identificar con la fase de "Evaluación" del proceso CRISP-DM y / o la etapa "Evaluación" del proceso SEMMA, respectivamente.
- El paso KDD "Uso del conocimiento descubierto" se puede identificar con la fase de "Despliegue" del proceso CRISP-DM.

En este trabajo, se ha seleccionado como metodología de trabajo KDD ya que en la actualidad es la metodología más utilizada por expertos, además de que se presenta como el campo relacionado al Data Mining. Esto se debe a que es considerado como más completo y preciso, al proporcionar pasos especialmente dedicados a la interpretación y evaluación de los resultados obtenidos con posterioridad a la minería. Por el contrario, CRISP-DM y SEMMA no implementan un estadío de interpretación que para el propósito de este trabajo es muy importante ya que busca detectar patrones de relación útiles, entre usuarios e interpretar qué impacto puede alcanzar una campaña de marketing orientada a esos usuarios

2.2 Estudio comparativo de herramientas de trabajo

En este análisis, las herramientas se agruparon en Software Licenciado y Software Libre (Tabla 2).

Tabla 2. Herramientas para minería de datos.

Herramientas de Minería de Datos	**Software Libre**	Weka
		RapidMiner
		KMine
		R
		Python y sus librerías de Minería de Datos
	Software Licenciado	SAS Data Mining
		SPSS Modeler
		Oracle Data Mining
		Microsoft SQL Management Studio

Para el estudio comparativo (Tabla 3), se consideraron las características más importantes de cada herramienta con el propósito de identificar si las mismas soportan características equivalentes.

Tabla 1. Estudio comparativo de metodologías para minería de datos

Características	Weka	RapidMiner	R	Python	Kmine	SQL
Clusterización	Sí	Sí	Sí	Sí	Sí	Sí
Asociación mining	Sí	Sí	Sí	Sí	Sí	Sí
Regresión lineal	Sí	Sí	Sí	Sí	Sí	No
Regresión lógica	Sí	Sí	Sí	Sí	Sí	No
Clasificadores Bayesianos	Sí	Sí	Sí	Sí	Sí	No
Árboles de decisión	Sí	Sí	Sí	Sí	Sí	No
Análisis de series de tiempo	Sí	No	Sí	Sí	Sí	No
Análisis de texto	Sí	Sí	Sí	Sí	Sí	Sí
Procesamiento de BigData	Sí	No	Sí	Sí	No	No
Flujos Virtuales	Sí	Sí	No	No	Sí	No

Para el desarrollo de este trabajo se usará Python como lenguaje para aprovechar su dinamismo, su gran variedad de librerías disponibles relacionadas con los campos del aprendizaje automático y que están disponibles como código abierto. También es utilizado ampliamente en ciencias e ingeniería y dispone el soporte de la comunidad que ha estado trabajando en proyectos de Minería de Datos para despejar dudas acerca del desarrollo. Para el propósito de este trabajo Python dispone de la librería Pytweet[1], que permite el acceso a la Api[2] de Twitter y disponer de los múltiples puntos de acceso que la red social ha dispuesto para los desarrolladores.

3. Desarrollo del Caso de Estudio

Con el fin de garantizar una adecuada adquisición de conocimientos, se han realizado una serie de entrevistas con expertos en marketing para tener un dominio general y otro específico, de los aspectos que rodean a la problemática. Las sesiones con los expertos se han dividido de tal forma que ellos participen en pasos claves del desarrollo de este trabajo. Por esto se han desarrollado tres ciclos de entrevistas como parte de las fases de planteamiento de problema, revisión y selección de datos y validación de la solución. Para el relevamiento de información basado en entrevistas se ha utilizado el modelo propuesto por García Martinez R. et al. [11].

La técnica utilizada fue la de Entrevista no Estructurada. De los resultados recolectados surgen las siguientes situaciones:

1. Un problema que puede enfrentar una compañía es que al recolectar datos por medio de encuestas, entrevistas u observaciones, se obtienen datos limitados y estructurados que proveen una visión limitada del mercado. Por otra parte teniendo en cuenta el dinamismo del mercado, estos datos pueden quedar desactualizados rápidamente provocando que la campaña lanzada con posterioridad sea menos acertada.
2. Al generar, en una nueva estrategia de marketing digital por medio de las redes sociales es posible encontrar usuarios que no interesan directamente, pero representan una gran cantidad de casos de análisis. Ante estos casos, la asertividad de la campaña se reduce considerablemente. Estos usuarios pueden ser:
    - Usuarios tipo Bots o de creación automática.
    - Usuarios con poca actividad.
    - Usuarios que viven en un área o país a la que no apunta la campaña.
    - Usuarios con pocos seguidores.
3. Al presentarse un universo de usuarios de interés es necesario analizar cuáles de ellos generan más impacto y entender esas características que los relacionan con otros seguidores. Esto requiere analizar grandes volúmenes de datos e insume

1 "Pytweet: Una biblioteca pythonic que proporciona una interfaz simple para la API de Twitter. Los valores se normalizan en tipos de Python.

2 "Application Programming Interface". En la programación de computadoras, una interfaz de programación de aplicaciones es un conjunto de definiciones de subrutinas, protocolos y herramientas para crear aplicaciones de software.

gran cantidad de tiempo que puede producir desactualización del plan de marketing.

En este trabajo se buscar dar respuesta a las siguientes preguntas de investigación:

- ¿Es posible obtener datos de redes sociales en tiempo real?
- ¿Es posible filtrar el contenido de aquellos usuarios que no interesan directamente a la campaña de Marketing?
- ¿Es posible realizar segmentaciones de mercado de los seguidores de la cuenta en estudio usando metodologías y técnicas de Minería de Datos?

De acuerdo al objetivo planteado, no se llevará a cabo la instancia de implementación, es decir que no se contempla la instalación del desarrollo realizado. Sin embargo, se realiza el desarrollo conceptual de los modelos paso a paso y se muestran los resultados mediante gráficos y figuras que le permitan al profesional de marketing medir su utilidad y aplicabilidad.

El análisis realizado está limitado a una sola red social, ya que debido a la complejidad de la conexión a las distintas Apis y a las limitaciones de tiempo para realizar consultas, los tiempos del proceso de obtención de datos serian excesivos. Por esta razón también se contempla analizar una sola cuenta ya que obtener datos de otras cuentas necesitaría una cantidad de recursos que no están disponibles. En caso de contar con dichos recursos, el modelo podría escalarse.

Finalmente, el propósito del ejemplo es obtener los Tweets publicados en el ámbito del Capital Federal donde la variable "CapitalFederal" contiene un array con las coordenadas que encierran a la ciudad.

Como entorno de desarrollo se utilizó Anaconda[3] aprovechando sus ventajas para el manejo de paquetes y librerías. Además brinda un tablero de control que permite manejar e instalar los paquetes y aplicaciones para trabajar con el análisis de datos.

Se eligió a Twitter debido a que ofrece ciertas ventajas que la hacen muy interesante desde el punto de vista de análisis de datos. Este trabajo se centra en el análisis de una cuenta con gran número de seguidores de esta red social (Cuenta en Estudio).

La aplicación utilizada para extracción de datos es API REST. Ésta proporciona acceso de lectura sobre los datos públicos en Twitter. No requiere disponer datos de ingreso de las cuentas a las que se pretende acceder, por lo que se pueden extraer datos como los tuits de un usuario, información de perfiles usuario y datos de los seguidores entre otras opciones de una forma muy sencilla. Después un primer análisis en Twitter, se ha detectado como los perfiles de usuario en la red social no contienen información relevante como la edad, el sexo, características físicas. El acceso a la localización de los usuarios es limitado ya que pueden existir tuits localizados o no.

Por lo tanto, el análisis en Twitter enfoca el punto de vista de la actividad de los seguidores, la calidad de estos y sus intereses, datos existentes para cada usuario y accesibles públicamente mediante la API.

[3] Es una distribución de código abierto de los lenguajes de programación Python y R para procesamiento de datos a gran escala, análisis predictivo y computación científica, que tiene como objetivo simplificar la administración e implementación de paquetes.

Un factor a tener en cuenta es el límite en la tasa de peticiones. La tasa límite de la API se define en intervalos (o ventanas) de 15 minutos en las que según el método utilizado se pueden realizar un número determinado de peticiones.

En la fase de selección, inicialmente se analizan los datos disponibles relacionados a la cuenta de Twitter seleccionada. Posteriormente obtendrán los datos desde el Api de Twitter seleccionando aquellos datos que se consideren relevantes para el objetivo de este trabajo.

Se genera un Script en Python el cual incluirá los pasos para descargar los datos de estudio de la cuenta seleccionada. Estos pasos son: 1) autenticación y conexión con la Api de Twitter; 2) la obtención de datos de perfil de cada usuario mediante el método GET users/lookup; 3) la descarga y almacenamiento de los últimos 200 tuits de cada uno de los followers, a partir de los IDs obtenidos en la función 2.

Se tomó para el estudio una cuenta de una automotriz ampliamente conocida en Argentina que se unió a Twitter en Octubre del 2011, desde entonces ha publicado más de cuatro mil tuits y llega a más de cincuenta mil seguidores. Haciendo uso de la Api se obtienen datos más detallados. El acceso a estos datos depende de la obtención previa de las credenciales descritas anteriormente para lograr una autorización del tipo OAuth[8].

En la fase de minería de datos, se plantean escenarios asociados a cada tipo de segmentación (Segmentación Geográfica, Segmentación Demográfica, Segmentación Socioeconómica y Segmentación Conductual), los cuales buscan resolver la problemática desde el punto de vista de cada uno de ellos, analizando los atributos que las caracterizan y aplicando los modelos desarrollados con los datos obtenidos de los estadíos previos. Esto permite comprobar si el escenario de segmentación de mercado es válido y es posible realizar segmentaciones a partir de los datos obtenidos de Twitter.

En este artículo, se presenta el Escenario Segmentación Geográfica. El mismo consiste en realizar una división por países, regiones, ciudades, pueblos o barrios de la información analizada. Twitter permite acceder a la locación del usuario, pero en algunas situaciones con inconvenientes de valores incorrectos o inexistentes.

Para dar solución a este problema previamente se implementó un algoritmo para inferir la ubicación de usuario. Aún en estos casos, existen valores incorrectos que no pueden ser interpretados como una locación. Para ello, se implementa una de las tareas de minería de datos, la clasificación basada en reglas, la cual busca empareja o asociar datos a grupos predefinidos (aprendizaje supervisado).

Estos datos predefinidos son el conjunto de entrenamiento, el cual es preparado por el humano (Experiencia). La implementación del modelo requiere la construcción del conjunto de datos de entrenamiento. Para este fin se ha descargado la base datos del Correo Argentino. Una vez generado el conjunto de entrenamiento, el clasificador lo utiliza como se presenta en la Figura 3, donde determina según la regla que el valor ”cap. fed” se refiere a la *Capital Federal, Buenos Aires, Argentina.*

El clasificador basado en reglas hace uso de un conjunto de reglas IF-THEN para la clasificación. No es necesario generar en primera instancia un árbol de decisión. En este algoritmo, cada regla para una clase dada cubre muchas de las tuplas de esa clase. Según la estrategia general, las reglas se aprenden de a una por vez. Para cada vez que se aprenden las reglas, se elimina una tupla cubierta por la regla y el proceso continúa

para el resto de las tuplas. El algoritmo representado en la Figura 4, corresponde a la creación de regla basándose en el conjunto de entrenamiento.

```
Algoritmo: Cubrimiento secuencial
Input:  data #Datos de entrenamiento
Output:  Un conjunto de reglas IF-THEN.
Method:
rule_set={ }; // el listado de reglas iniciales es vacío
for clase in data
  rule  =  Learn_Rule(clase['locacion'],  clase['time_zone'],clase['classification_locatio
);
  #Remueve reglas repetidas generadas por entrenamiento previo
  Remove_Rules_Covered(rule_set, rule)
  # Agrega una nueva regla
  Rule_set.append(rule)
return rule_set;
```

Fig. 3. Algoritmo de generación de reglas

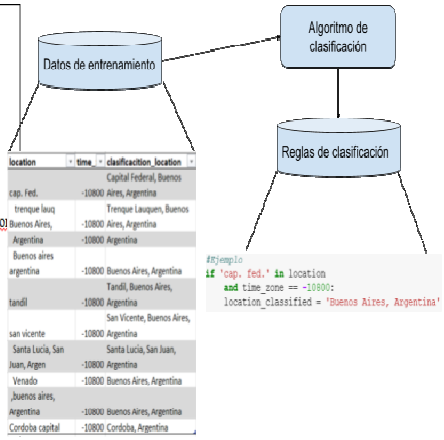


Fig.4. Clasificador de locaciones

Fase de evaluación del modelo: si bien el tratamiento realizado puede brindar un acercamiento a la locación del usuario, pueden existir casos en los que la locación no se obtenga por medio del clasificador implementado. Para estos casos, el campo se deja vacío y no será contemplado para el escenario de segmentación geográfica. La Figura 5 muestra los resultados de la ejecución del modelo. Se obtuvieron que al menos el 70% de los datos fueron clasificados exitosamente mientras el 30% no podrán ser tenidos en cuenta para realizar la segmentación geográfica.



Fig. 5. Resultados de la ejecución del modelo de clasificación

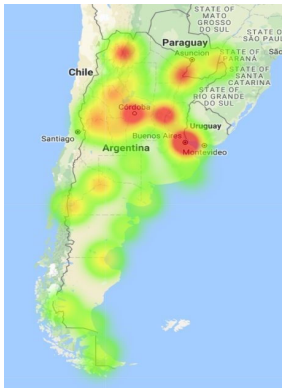


Fig. 6. Concentración de seguidores

Finalmente con las locaciones clasificadas se consiguió detectar la ubicación del 70% de los usuarios de interés que siguen a la cuenta. Con esta información se generó la segmentación geográfica de los seguidores y permitió crear un mapa de calor, el cual indica la densidad de concentración de los seguidores en regiones determinadas (Figura 6).

Otra segmentación geográfica que permite generar la obtención de estos datos fue la división por región, donde se observó que el 80% de los seguidores de importancia de la cuenta están ubicados en la región de la Pampa (Figura 7).Por otra parte, y como lo muestra la figura 8, es posible realizar segmentaciones por provincia.

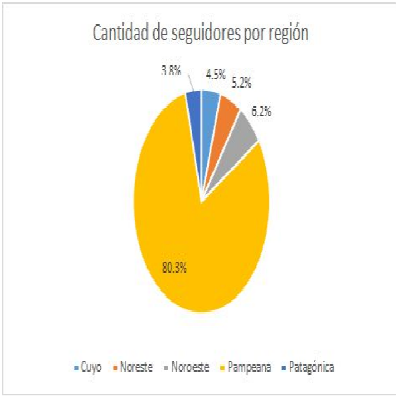


Fig 7. Cantidad de seguidores por Región.

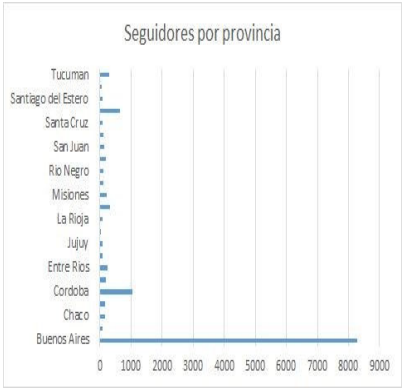


Fig. 8. Seguidores por provincia.

4 Conclusiones y futuros trabajos

En este proyecto se ha marcado como objetivo fundamental lograr segmentaciones de mercado a partir de datos obtenidos de redes sociales mediante la utilización de metodologías y técnicas de Minería de Datos.

La elección de Twitter como fuente de los datos resultó satisfactoria ya que provee una potente Api que permite consumir los datos, aunque existieron desventajas a nivel de datos, debido a que Twitter comparte una cantidad limitada de datos de usuarios. Esto no permitió realizar segmentaciones con algunos atributos como por ejemplo el Edad, Género u Ocupación.

Fue necesario elegir un entorno de desarrollo para conectarse a la interfaz expuesta por la red social (Api), se eligió Python para este propósito, lo cual resultó sumamente satisfactorio ya que este lenguaje en la actualidad tiene mucho soporte de la comunidad y cuenta con gran variedad de librerías de código abierto para múltiples propósitos.

Una de estas librerías resultó ser precisa para abstraer el desarrollo de la compleja implementación de consumir una Api y tener que implementar una paginación propia.

La posibilidad de consumir los datos directamente desde los servidores de Twitter brindó respuesta a la primera pregunta problema.

Se aplicó el proceso metodológico propuesto por KDD, por medio de sus procesos permitió solucionar los problemas encontrados. El pre procesamiento generó nuevos datos en base a los obtenidos y utilizarlos como atributos para realizar la segmentación de mercado.

Como trabajos: a) explotar más técnicas de NLP (Neuro Linguistic Programming), por ejemplo investigando sobre el análisis de sentimientos; b) crear una aplicación que permita realizar analytics sobre cualquier cuenta y ofrecer los resultados de la analítica en la misma interfaz.

Referencias

1. Kotler, Philip; Gary Armstrong, John Saunders, Veronica Wong (2002). «Capítulo 1: ¿Qué es Marketing?». Principles of Marketing (3ª edición europea edición). Essex (Inglaterra): Prentice Hall. ISBN 0-273-64662-1.
2. Lamb, C. W., Hair, J. F., & McDaniel, C. D. (2004). Marketing. South-Western.
   May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148--1158. Springer, Heidelberg (2006)
3. Chen, Z., Kalashnikov, D. V., & Mehrotra, S. (2009). Exploiting context analysis for combining multiple entity resolution systems. In Proceedings of the 35th SIGMOD international conference on Management of data - SIGMOD '09. https://doi.org/10.1145/1559845.1559869
4. Digital in 2017: Global Overview - We Are Social. (n.d.). Retrieved October 4, 2017, from https://wearesocial.com/special-reports/digital-in-2017-global-overview
5. Ashton Acton,PhD. General Editor. Issues in General Science and Scientific Theory and Method: 2011 Edition. ISBN: 978-1-4649-6346-9.
7. Chen, M.-S., Han, J., & Yu, P. S. Watson IBM Research (1996). Data Mining: an Overview from Database Perspective. http://hanj.cs.illinois.edu/pdf/survey97.pdf
8. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. 1996. From data mining to knowledge discovery in databases. AI Magazine, 17(3): 37-54.
9. SAS Institute Inc. 1998. SAS Institute White Paper, From Data to Business Advantage: Data Mining, The SEMMA Methodology and the SAS® System, Cary, NC: SAS Institute Inc.
10. Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 2000. CRISP-DM 1.0 Step by step BI guide.
11. García Martínez, R., & Britos, P. V. (2004). Ingeniería de Sistemas Expertos. Buenos Aires.

Text pre-processing tool to increase the exactness of experimental results in summarization solutions

Augusto Villa Monte[1], Julieta Corvi[2], Laura Lanzarini[1], Cristina Puente[3], Alfredo Simón Cuevas[4], and José A. Olivas[5]

[1] Institute of Research in Computer Science LIDI
[2] Faculty of Computer Science, National University of La Plata
La Plata, Buenos Aires, Argentina
{avillamonte,laural}@lidi.info.unlp.edu.ar, julieta.corvi@gmail.com
[3] Advanced Technical Faculty of Engineering ICAI
Comillas Pontifical University, Madrid, Spain
cristina.puente@icai.comillas.edu
[4] Faculty of Computer Engineering, Technological University of Havana
"José Antonio Echeverría", La Habana, Cuba
asimon@ceis.cujae.edu.cu
[5] Department of Information Technologies and Systems
University of Castilla-La Mancha, Ciudad Real, Spain
joseangel.olivas@uclm.es

Abstract. For years, and nowadays even more because of the ease of access to information, countless scientific documents that cover all branches of human knowledge are generated. These documents, consisting mostly of text, are stored in digital libraries that are increasingly consenting access and manipulation. This has allowed these repositories of documents to be used for research work of great interest, particularly those related to evaluation of automatic summaries through experimentation. In this area of computer science, the experimental results of many of the published works are obtained using document collections, some known and others not so much, but without specifying all the special considerations to achieve said results. This produces an unfair competition in the realization of experiments when comparing results and does not allow to be objective in the obtained conclusions. This paper presents a text document manipulation tool to increase the exactness of results when obtaining, evaluating and comparing automatic summaries from different corpora. This work has been motivated by the need to have a tool that allows to process documents, split their content properly and make sure that each text snippet does not lose its contextual information. Applying the model proposed to a set of free-access scientific papers has been successful.

Keywords: Automatic Summarization, Extractive Approaches, Web Scraping, Document Representation, Summaries Evaluation.

1 Introduction

For years, technological advances have allowed recording all types of information and easily storing it. As the use of the Internet becomes popular and storage costs are reduced, news, articles, books, tweets, e-mails, audios, images, videos, among others, are created, duplicated, stored, shared and accessed constantly. All this is possible thanks to textual data. To give an example, both the e-mails that are sent daily and the Web search queries that are made continually are nearly all text-based. Even a video uploaded to YouTube would remain largely inaccessible were it not for the title, description, and other metadata in text format. It is clear that not everything is text but text is everywhere [1]. Textual data surrounds our lives and has been growing continuously in recent years. Today, text represents one of the most valuable resources and the use of automated tools for its processing is essential.

In the scientific field, a myriad of articles is produced, published and hosted on repositories that can be easily accessed. These documents encompass all topic areas and represent the human knowledge. Separating what is essential from what is not, has proven to be a difficult task to be performed manually when the volume of information is immense. In this context and when it comes to text, obtaining summaries by computer reduces this huge volume of non-structured information to its most important content to facilitate its manipulation in an automatic way. This allows obtaining the core contents of a document in less time than of what it would take to do it manually. This is known as automatic text summarization.

Among the existing approaches, the summary by extraction is easy to develop since the program does not have to generate a new text that provides a degree of generalization to construct sentential meanings. The extractive approach is a relatively low-cost solution in relation to the additional linguistic knowledge resources specifically required (such as ontologies and dictionaries) to construct a summary by abstraction [2]. An extractive summary is a set of text portions (from single words to whole paragraphs) literally copied from the input that constitute the summary [3]. The extractive approach selects parts of a document without requiring complex semantic analysis [4] . However, to this end it is indispensable to assign a score to each part. This scoring allows to order all parts of a document from highest to lowest in a ranking list where the first positions are more relevant [5].

The specific values are obtained by metrics whose calculations depend on the formulation of certain equations. The metrics range from identifying certain expressions (e.g. “this article describes” or “finally”) or keywords within the text to more complex calculations. In many of the works on extractive summaries, documents are modeled as n-dimensional vectors of numerical features obtained by calculating n metrics. This is based on the classic vector space model proposed by Salton, Wong and Yang in 1975 [6]. Then, these feature vectors are used to obtain an automatic summary by applying more sophisticated algorithms to them [7]. However, in few works the way in which those features were calculated is developed in depth. Designing a program that selects representative and

significative phrases from documents automatically requires precise instructions [8].

In most cases, known collections of documents (such as DUC [9]) are used in the experiments but special considerations during the preprocessing of the documents are not always detailed. This does not allow to be objective when comparing results and obtaining conclusions. This paper presents a text document manipulation tool to increase the exactness of results when obtaining, evaluating and comparing automatic summaries from different corpora. This tool ensures the correctness of calculations and allows to recalculate metrics changing the "view" of the document without too much effort. This work has been motivated by the need to have a tool that allows to process documents, split their content properly and make sure that each text snippet does not lose its contextual information (necessary for the calculations of the metrics). With this tool, the documents are adequately stored in a database and can be reconstructed according to the needs of the experiment in question. This allows the integration of automatic summary generation systems that operate with significant differences on documents, something quite useful in this field.

The remaining of this paper is organized as follows. In Section 2 the main approaches for text summarization are detailed. Section 3 describes the proposed tool and how to use it. Finally, in Section 4, conclusions and some future lines of work are presented.

2 Related work

Both the Royal Spanish Academy and the Cambridge Dictionary agree in that a summary is a series of short, clear and precise statements that give the essential and main ideas about something. The automatic generation of text summaries is the process through which a "short version" of one or more documents is created using a computer (with no human intervention) where the information is kept.

Even though it started being researched many years ago, automatic text summarization is still a relevant topic that receives ongoing scientific contributions [10, 11]. There are two main types of automatic summaries: extractive and abstractive [12]. Extractive summaries are formed by "phrases" from the document that were appropriately selected. Abstractive summaries are formed by the "ideas" developed in the document, without using the phrases exactly as they appear in the original document. The differences between the two approaches can be clearly seen in [4].

In literature there is a vast number of related works with extractive summaries whose common main goal is to reduce document size while preserving the information of the source document [13]. Even though they have different approaches on the issue, a set of metrics is commonly used. Each metric analyzes a given characteristic of the document and allows to apply sorting criteria to its content. This order is obtained by assigning a score to each part of the document and by ordering the values from highest to lowest.

In 1958, at IBM, Hans Peter Luhn developed a simple summarization algorithm which used the distribution of word frequency to weigh sentences [14]. At the same time, Baxendale used the position of the sentences [15]. Both of them made the earliest proposals for automatic indexing. A few years later, Harold Parkins Edmundson and Roland Eugene Wyllys, based on Luhn's work, proposed using the presence of certain words and the overlapping of words and titles [16]. Many years later, Joel Larocca Neto et al. used the term frequency and the Inverse Sentence Frequency (an adaptation of the TF-IDF measure used in Information Retrieval) [17]. The following year, Yihong Gong and Xin Liu proposed the use of Latent semantic analysis (LSA) for text summarization without the use of lexical resources such as WordNet [18]. In 2004, Rada Mihalcea introduced a graph-based ranking model for sentence extraction applied to text summarization [19]. The same year, Fatma Jaoua Kallel et al. presented a summarization method based on the ratio of keywords in the sentences and in the document [20]. In 2007, Lucy Vanderwendea et al. operationalized the idea of using word frequency averaged for sentence selection [21]. These are just some of the most cited references in literature. A broader list of methods can be consulted in [7, 12].

In recent years, various metric-based techniques have been developed to extract the important contents from text documents to represent their summaries (e.g. [22] in whose work its authors assessed the performance of sentence scoring techniques individually and applying different combination strategies). The common factor in all these works is the use of metrics to model the documents as feature vectors. To do this, three steps are followed: Selecting a collection, pre-processing the documents and calculating the metrics. Subsequently, the same pre-processed documents are used to evaluate the results obtained. But it happens that several decisions made during the preprocessing of the documents have an influence on the calculation of the metrics. When using any test collections for our research works we do not have the same "view" of the documents that other authors have. It is not correct to compare summaries generated by methods whose input vectors are calculated in different ways. For example, it is not the same to calculate the metrics by sentence than by paragraph or by taking into account only nouns instead of all the words.

This work presents a tool that aims to solve these problems. The proposal consists of a set of routines to correctly preprocess the documents and a database to store them properly. In short, it is intended to increase the exactness of results when obtaining, evaluating and comparing automatic summaries obtained from different corpora using varied methods. The implementation of the proposed system will be detailed in the following section.

3 Description of Proposal

3.1 Database Model

In this paper, we propose a database design to store text documents in a structured manner allowing to record all the content of a document and

its details. This design was conceived for scientific documents taking into account the needs of extractive summarization. These types of documents use a predetermined structure that typically starts with the title of the article, the authors and other information (institutions, e-mails), followed by a summary, a set of keywords and each of the sections that compose the document.

Documents are fundamentally a sequence of characters in a string encoded with a chosen character set. Whereas plain text files only contain the characters of a text, other formats, such as XML, can express the content using a markup strategy. It consists of a tag-based structure that identifies specific parts within a document. The Center for Digital Research in the Humanities defines the XML format as an encoding standard that assists in the creation, retrieval and storage of documents. This allows processing the document and successfully storing it to the proposed database.

Normally, in these types of documents, the first sections provide general information in the introduction and then go on to detail more specific aspects in relation to the topic being discussed. Finally, they discuss the results obtained and the conclusions drawn and list bibliographic references. All this information must be stored. Also, the document has a title, a summary and a list of keywords. Then, its content is organized into sections, sub-sections, sub-sub-sections, etc., each of these with its own title. The content of each section consists of a set of paragraphs formed by sentences. Each sentence contains a sequence of words. Each word must be registered along with its location within the document. With the data design proposed here, the original document can be re-built at any time using all information stored in the database. Figure 1 shows the database design proposed for storing these types of documents.

The database consists of fifteen tables that store all the information needed to locate each word within a document. The model takes into account the journal, the issue and the articles published in each issue ("journal", "edition," and "document" tables). The text in the document is broken down into the tables called "section", "paragraph," and "sentence." The table called "section" is recursive, which allows recording section embedding, typical organization in scientific documents. The rest of the information (all titles, the keywords list and the text of sections) is related to the table called "word." Each word in the text is reduced to its stem by applying a stemming algorithm and it is stored in the table "stem" relating to the corresponding word in the table "word." This is the most significant term reduction task in Text Mining, since the number of words derived from the same stem is very high.

3.2 Assessment and Data Use

In this work, we used a subset of free access articles published in a journal on biomedical, environmental, social and political health issues from Public Library of Science (PLOS). These documents meet all requirements and can also be downloaded for free in XML format through the Internet. Having the documents in this format facilitates the identification of each of the parts of the document.

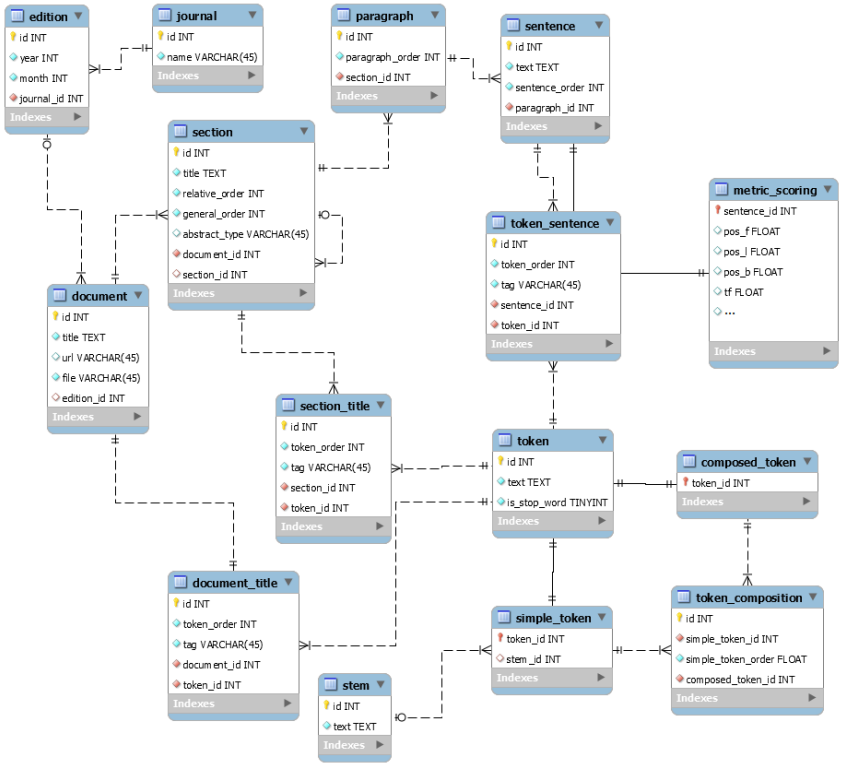


Fig. 1. Database model proposed for storing text content from scientific articles.

Files in this format can be processed with special libraries that create a tree with all the elements of the document. Each node in a tree is an object that represents a part of the document. This can be used to extract specific information from the document to store it in the database. In addition, these objects can be manipulated by programming and any change made to them will be reflected in the later visualization of the document. This can be interesting if, when identifying the relevant parts of a document, they were highlighted automatically to differentiate them from the rest of the document just like the human being does with a marker.

After retrieving the XML files, they were processed as required to upload them to the database, following these steps for each article:

- The title was extracted, together with the summary created by the author(s).
- Each of the sections was processed, and their titles identified. Certain full sections, such as Bibliography and Acknowledgments, were disregarded, as well as all figures, tables and equations included in the article.

– To obtain the sentences, the paragraphs in each section were segmented using the full stop as delimiter, except when it was used as decimal separator and as part of an abbreviation, among other uses.
– The keyword list, the titles, and the sentences were tokenized using white spaces and punctuation marks. Each token was added to the "token_sentence" table after verifying that the token existed.
– When a token does not exist previously, on the one hand it is added to the "token" table, marking it as a stopword if it is, and on the other hand, in "simple_token" or "composed_token", as appropriate. Every simple token has its corresponding root stored in the "stem" table.

The process of loading all this information to the database from the XML file was carefully carried out so as to avoid altering the order of the sections and their structure. The Data Base Management System used was MySQL. Python programming language and Natural Language Toolkit (NLTK) were used for automatic file download from the Internet and their processing [23]. In order to do this a script was developed using urllib and BeautifulSoup libraries. The Porter stemming algorithm and english stopword list of NLTK were used. Porter stemmer is one of the best known algorithms to perform this task [24]. The stoplist is a list of words with empty meaning called stopwords, which contains pronouns, prepositions, articles and a few verbs. In addition to providing stemming and stoplists, the NLTK package provides tokenization, tagging, parsing, and very useful semantic reasoning functions. All of the text pre-processing tasks that are mentioned in this paragraph are common in this area of research.

During this step, a document is transformed and several decisions have to be taken: Which words to consider stopwords? Which algorithm of stemming to use? What is considered a token? What are the delimiters to segment the text? And many more questions. The final result is a view of the document from which, for example, a series of metrics will be calculated. For this reason, it is essential to have control over the impact of the pre-processing performed on the documents.

3.3 Usefulness of the Proposed Model

From a set of documents, each of the scoring metrics used for automatic summary generation must be calculated for each sentence. This requires, among other things, processing variables of text type several times using the string split method. Even if before calculating metrics we tried different preprocessing actions, it would take much longer.

Once the documents are stored in the proposed database each of the metrics can be calculated through a single SQL query for all sentences in the document. But if it is done in this way the query will be complex and not easy to prepare. A better solution is to combine SQL queries with Python code. The query will retrieve the corresponding identifiers of the tokens associated with each sentence of the document, and with Python the frequencies necessary to calculate each

metric will be easily computed. The interesting thing is that, like their stems, the tokens can be retrieved, and even the stopwords or certain types of words (such as nouns, verbs, etc.) can be filtered. In addition, it is possible to recalculate the metrics by paragraphs without requiring many modifications.

Once the metrics are calculated, each document will be represented by a matrix of as many rows as sentences the document has, and as many columns as metrics are calculated. These matrices will be stored in the table "metric_scoring" shown in Figure 1 and can be obtained by filtering the tuples of each document.

As has already been mentioned, creating extractive summaries requires selecting certain parts of the document that contribute to the desired summary. Just like humans mark on paper the text that they consider important, an automatic system should assign each sentence in the original document to one of two possible classes: "Summary," if the sentence belongs to the summary, or "Non-Summary," if it does not. Instead of drastically deciding if a sentence belongs to the summary or not, the list of sentences ordered by metric from highest to lowest can be split into two groups using a threshold value: the class "Summary" can be assigned to the best positioned N, and the class "Non-Summary" to all others. The value of N is a parameter specified by the user that allows to select the N most relevant sentences and to control the size of the resulting summary.

Once the sentences that will form the automatic summary have been selected, they should be evaluated by comparing them with an expected summary. For this task, the identifiers of the tokens of the generated and expected summary can be obtained through an SQL query. Then, a small program in Python calculates the intersection between both sets to get the exact value of Rouge. ROUGE, developed by Chin-Yew Lin, provides different measurements frequently used in literature to the assessment of automatic summaries quality [25].

Since the specific size of the summary is not known a priori, summaries are usually evaluated by varying the value of N and averaging the results. Also, since not all the words in the expected summary are used in the document, the rouge values obtained between documents for the different sizes are not comparable to each other. For this reason, before averaging it is very important to divide each value by the maximum value of rouge obtained with the full document as summary.

Another useful aspect of this model is the flexibility to reconstruct a document according to the input format of other automatic systems. Sometimes, experiments are carried out using online summarization systems, whose input must be built in a certain way. This tool was used in two previous research papers related to automatic summaries [26, 27].

4 Conclusions and future works

Automatically obtaining summaries from text documents continues to be a subject of study that constantly receives scientific contributions. The extractive summarization approach is one of the most commonly used methods in literature,

since it can create a summary by setting aside semantic aspects that the abstractive approach usually requires.

In this article, a data model for storing scientific documents with a predefined structure was presented. By using it, a set of metrics can be calculated for each document using SQL queries and Python code. Thus, the sentences in a document are sorted following specific criteria, which allows generating various extractive summaries of a controlled size. It should be noted that open source tools were used for every aspect of this work.

This proposal is a practical contribution to summarization area that allows to easily calculate metrics, diminishing developing time and some possible ambiguity of interpretation. It saves calculation time, clearly expresses how each metric is calculated, and facilitates experimentation in this research area. Also, the document can be easily reconstructed according to the needs of the experiment to be carried out, allowing the integration of automatic summary generation systems that operate over the documents with significant differences.

In the future, we intend to introduce changes in the model to equip it with semantic. It is expected that incorporating a semantic analysis, the calculation of the metrics based on frequency or intersection of words can be improved. Additionally, an application will be developed to allow users to view the document with the parts that were selected by each of the summaries.

References

1. Schreibman, S., Siemens, R., Unsworth, J.: A New Companion to Digital Humanities. Blackwell Companions to Literature and Culture. Wiley (2016)
2. Mani, I.: Automatic Summarization. Natural language processing. J. Benjamins Publishing Company (2001)
3. Mani, I.: Summarization evaluation: An overview (2001)
4. U., H., I., M.: The challenges of automatic summarization. Computer **33**(11) (2000) 29–36
5. Edmundson, H.P., Wyllys, R.E.: Automatic abstracting and indexing—survey and recommendations. Commun. ACM **4**(5) (1961) 226–234
6. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Commun. ACM **18**(11) (1975) 613–620
7. Nenkova, A., McKeown, K. In: A Survey of Text Summarization Techniques. Springer US, Boston, MA (2012) 43–76
8. Cremmins, E.T.: The art of abstracting. Professional writing series. ISI Press (1982)
9. National Institute of Standards and Technology (NIST): Document Understanding Conferences (DUC). http://www-nlpir.nist.gov/projects/duc/index.html (2002)
10. Spärck Jones, K.: Automatic summarising: The state of the art. Inf. Process. Manage. **43**(6) (2007) 1449–1481
11. Gambhir, M., Gupta, V.: Recent automatic text summarization techniques: a survey. Artificial Intelligence Review **47**(1) (2017) 1–66
12. Torres Moreno, J.M.: Automatic Text Summarization. Cognitive science and knowledge management series. Wiley (2014)
13. Gupta, V., Lehal, G.S.: A survey of text summarization extractive techniques. Journal of emerging technologies in web intelligence **2**(3) (2010) 258–268

14. Luhn, H.P.: The automatic creation of literature abstracts. IBM J. Res. Dev. **2**(2) (1958) 159–165
15. Baxendale, P.B.: Machine-made index for technical literature: An experiment. IBM J. Res. Dev. **2**(4) (1958) 354–361
16. Edmundson, H.P.: New methods in automatic extracting. J. ACM **16**(2) (1969) 264–285
17. Larocca Neto, J., Santos, A.D., Kaestner, C.A., Freitas, A.A. In: Generating Text Summaries through the Relative Importance of Topics. Springer Berlin Heidelberg, Berlin, Heidelberg (2000) 300–309
18. Gong, Y., Liu, X.: Generic text summarization using relevance measure and latent semantic analysis. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '01, New York, NY, USA, ACM (2001) 19–25
19. Mihalcea, R.: Graph-based ranking algorithms for sentence extraction, applied to text summarization. In: Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions. ACLdemo '04, Stroudsburg, PA, USA, Association for Computational Linguistics (2004)
20. Jaoua Kallel, F., Jaoua, M., Belguith Hadrich, L., Ben Hamadou, A.: Summarization at laris laboratory. In: Proceedings of the Document Understanding Conference. DUC'04 (2004)
21. Vanderwende, L., Suzuki, H., Brockett, C., Nenkova, A.: Beyond sumbasic: Task-focused summarization with sentence simplification and lexical expansion. Inf. Process. Manage. **43**(6) (2007) 1606–1618
22. Oliveira, H., Ferreira, R., Lima, R., Lins, R.D., Freitas, F., Riss, M., Simske, S.J.: Assessing shallow sentence scoring techniques and combinations for single and multi-document summarization. Expert Systems with Applications **65**(Supplement C) (2016) 68 – 86
23. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python. 1st edn. O'Reilly Media, Inc. (2009)
24. Porter, M.F.: Readings in information retrieval. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1997) 313–316
25. Lin, C.Y.: Rouge: a package for automatic evaluation of summaries. In Marie-Francine Moens, S.S., ed.: Text Summarization Branches Out: Proceedings of the ACL-04 Workshop, Barcelona, Spain, Association for Computational Linguistics (2004) 74–81
26. Villa-Monte, A., Lanzarini, L., Rojas-Flores, L., Olivas, J.A.: Document summarization using a scoring-based representation. In: 2016 XLII Latin American Computing Conference (CLEI). (2016) 1–7
27. Puente, C., Villa-Monte, A., Lanzarini, L., Sobrino, A., Olivas, J.A.: Evaluation of causal sentences in automated summaries. In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). (2017) 1–6

Aceleración en la Recuperación de Información utilizando Algoritmos de Minería de Datos de R.

Osvaldo Spositto, Hugo Ryckeboer, Mauro J. Casuscelli,
Lorena Matteo, Julio Bossero.

Departamento de Ingeniería e Investigaciones Tecnológicas.
Universidad Nacional de La Matanza, Prov. Buenos Aires, Argentina
Florencio Varela 1902, San Justo, Prov. Buenos Aires, Argentina
spositto@unlam.edu.ar, hugor@unlam.edu,ar, mcasuscelli@alumno.unlam.edu.ar,
lmatteo@unlam.edu.ar, jbossero@unlam.edu.ar.

Resumen. Para acelerar la respuesta inicial en sistemas de recuperación de información sobre depósitos documentales privados de mediano tamaño se estudia la posibilidad de segmentar el mismo y elaborar la respuesta examinando sólo un segmento. Se analiza la pérdida de calidad que ello provoca. Las herramientas para fraccionar y elegir segmento provienen de la algoritmia de la minería de datos y se eligió el lenguaje R por tener ya incorporado los algoritmos básicos y ser un lenguaje que completo permite escribir el código que los vincula.

Palabras clave: Recuperación de Información, K-Means, Redes Neuronales, LSI, Distancia Euclidea.

1 Introducción

La recuperación de información es una técnica de la que se disfruta cuando se realizan búsquedas en Internet y es pretensioso intentar introducir mejoras a los buscadores más famosos. No obstante eso, existen depósitos documentales (corpus) privados que no se desea exponer al público y sobre los cuales se tiene interés en tener un sistema de recuperación. En la medida que tales depósitos aumentan de tamaño el tiempo de respuesta sube ya que es proporcional a la cantidad de documentos.

En todo sistema que interactúa con el hombre, hay que tener en cuenta la psicología de éste, el cual quiere respuestas con sensación de instantáneas y contra esta característica conspira el crecimiento del corpus. Una solución es aumentar la potencia de cómputo, pero esto no está al alcance de todos.

En este trabajo se analiza la posibilidad de fraccionar el corpus de modo tal de reducir el tiempo sin gran desmedro en la calidad de la primera respuesta que entrega el sistema frente a un requerimiento. Se propone que los segmentos contengan documentos afines de modo tal que muchas consultas queden resueltas por examen de un solo segmento aunque esa respuesta adolezca de algunos documentos.

También la propuesta tiene en cuenta que las consultas la realiza una persona y que habiendo varios documentos válidos en la respuesta inicial el usuario estará ocupado dando tiempo al sistema de perfeccionarla para cuando solicite las siguientes páginas.

Tratándose de poblaciones grandes, tanto los documentos como las consultas, ambos imposibles de describir con un patrón regular, evaluar esta propuesta será inevitablemente de un modo estadístico. Las herramientas para realizar esta tarea fueron sacadas de la minería de datos (MD), la cual justamente ha crecido para elaborar conclusiones sobre universos irregulares.

La tarea aquí comenzada se presta para futuras investigaciones, y además la técnica expuesta en un modo abstracto es aplicable a otras situaciones de apareo de objetos.

En la sección dos se describe someramente los principios de las tecnologías involucradas, limitado a lo efectivamente utilizado. En la sección tres se describe las tareas afines de los investigadores y algunas pocas ideas sobre lo realizado. En la cuartea, se detalla la idea y su concreción en código. Finalmente en la última sección se detallan las experiencias numéricas y el modo de juzgarlas.

2 Marco Teórico

El presente trabajo intenta poner la minería de datos al servicio de las necesidades de la recuperación de la información. La explicación de los algoritmos se reduce a lo necesario para facilitar la comprensión de la sección 4.

2.1 Recuperación de Información

Por recuperación de información (RI) se conoce una disciplina que ayuda a ubicar dentro de un repositorio de documentos los que mejor puedan resolver las necesidades intelectual del usuario del sistema de recuperación de información (SRI) [1].

Aunque el nombre de la disciplina quedó establecido así sería más adecuado verlo como un sistema que provee una lista ordenada de sugerencias de lectura [9], esperando que el usuario por examen de los mismos encuentre el material, no necesariamente sólo información, buscado. A diferencia de otros técnicas no utiliza, o al menos no primariamente, los metadatos del documento [2].

Como ya se mencionó, el conjunto de documentos sobre la cual se hará la selección se denomina el corpus y los sistemas que brindan el servicio, los buscadores. La idea que guía esta actividad es que las consultas y los documentos esperados comparten un mismo vocabulario. Afinando esta idea y teniendo en cuenta las inflexiones que sufren las palabras por necesidades gramaticales se ha pasado rápidamente a que más que palabras concretas conviene atenerse a los lexemas.

Si cada lexema se toma como una dimensión del espacio del habla, habrá lexemas con distinta frecuencia lo que hace que cada documento queda representado por un vector en este gigantesco espacio y allí también se representa la consulta.

Con una conveniente medida de distancia se puede lograr un ordenamiento de los documentos del corpus en función de la pregunta formulada y esto constituye la respuesta teórica al requerimiento formulado.

En la práctica no se entrega de una vez la lista total, imposible de manejar

intelectualmente, sino trozos, por ejemplo 10 documentos por vez, comenzando por los más promisorios. Después de examinar sus títulos y abrir los más promisorios bajo la óptica del investigador pasa a otra hoja si no encontró algo apropiado o por el contrario reformula su consulta.

Distintos sistemas difieren por el modo detallado con el cual construyen los vectores representativos, se los llama modelos. Los coeficientes son todos positivos y lo más sencillo es hacerlo booleano, unos si el lexema aparece, ceros si no.

Mejor que ello contar las ocurrencias, a las cuales se les aplican diversas correcciones. Esto se conoce como el modelo vectorial [1]. Los vectores son ralos ya que según la temática tratada aparecen unos lexemas y otros no.

Finalmente recurriendo a una descomposición en valores singulares (DVS) se logra evitar los errores que introducen la polisemia y la sinonimia. Los vectores resultantes son densos pero la experiencia indica que pueden reducirse sus dimensiones a unos pocos centenares. Se lo conoce como el modelo LSI (Latent semantic indexing).

Las consultas se deben volcar en el mismo modelo que se haya aplicado al corpus, como si fueran minúsculos documentos y enfrentar su vector representativo con los vectores de cada documento para luego ordenarlos, o al menos obtener el trozo inicial de lo que sería un vector ordenado. El tiempo de proceso es proporcional a la cantidad de documentos del corpus. El objetivo de esta investigación supone la existencia de algunos de estos modelos y es indiferente respecto de la calidad de los mismos.

2.2 La Minería de Datos

La minería de datos es una disciplina ya bien establecida y desarrollada en numerosos libros [6]. El objetivo principal es extraer nuevos conocimientos a partir de datos. Existen diversos tipos de métodos para extraer el conocimiento, estos métodos se agrupan de acuerdo al tipo de tarea que realizan. Las principales tareas son: clasificación, regresión, agrupación y asociación [6]. Tal vez sea más adecuado decir que organizan a través de estas tareas la información disponible para facilitar el conocimiento de la situación por parte de los usuarios de estos sistemas, las decisiones que deben tomar y su eventual delegación en sistemas automatizados. A continuación se explican de manera resumida dos de ellas que son las que fueron usadas en esta investigación.

Agrupación. También conocida como segmentación (en inglés se la conoce como clustering), es una técnica que permite analizar y examinar datos que no se encuentran etiquetados, formando conjuntos de grupos a partir de su similitud [6]. Los que comparten un mismo grupo recibirán una misma etiqueta, distinta de la de otros grupos. Los objetos a clasificar poseen propiedades sobre las cuales se puede definir un criterio de similitud o distancia. Especificar tales criterios se simplifica si las características son numéricas, con valores en un conjunto conceptualmente continuo. Esta situación la tenemos en los modelos de representación de documentos, salvo en el booleano, prácticamente en desuso. Las etiquetas son arbitrarias, optando muchos sistemas en aplicar números naturales consecutivos, carentes de todo significado adicional. Para nuestra aplicación la función distancia deberá armonizar con aquella que usa el ordenador de documentos en la recuperación.

Clasificación. Este tipo de tarea predice la categoría a la que pertenece un objeto dado. Se debe basar sobre el conocimiento de las categorías de otros objetos ya clasificados. De algún modo intenta ubicarlos en el grupo que contiene otros cercanos a él. Los métodos de clasificación extraen características de los objetos que ya están ubicadas en categorías durante un pre-proceso para agilizar las posteriores clasificaciones.

2.3 Algoritmos Utilizados

A continuación se da explica los dos algoritmos elegidos para realizar este proyecto.

Algoritmo K-Means. (En español debiera llamarse K-Medias), presentado por MacQueen [13] en 1967, es uno de los algoritmos desarrollados para resolver el problema del agrupamiento. La idea del algoritmo es proporcionar una clasificación de información de acuerdo con los propios datos, basada en análisis y comparaciones entre sus valores numéricos. Así, el algoritmo proporcionará una clasificación automática sin la necesidad de supervisión humana, es decir, sin pre-clasificación existente. Debido a esta característica, se considera como un algoritmo del tipo No Supervisado [6].

Es un algoritmo iterativo, parte de *K*, valor propuesto por el usuario, puntos en el espacio multidimensional de las características, que llamaremos centroides. La forma de elegir los centroides iniciales varía según distintas implantaciones que tiene el método. De allí en más cada iteración realiza dos pasos:

a) Por cada objeto a particionar se calcula cual es el centroide más cercano y se lo etiqueta como perteneciente a él.
b) Actualización centroides: se actualiza la posición del centroide de cada etiqueta tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.

Es usual que esta actividad converja y los cambios que sufren los centroides y acorde a ello los cambios de etiqueta sean cada vez menores. La teoría del método demuestra que la suma de los cuadrados de las distancias de los objetos etiquetadas a sus respectos centroides disminuye en cada paso.

Visto como problema matemático la solución no es única y puede ser aconsejable reiniciarla con nueva elección de centroides iniciales.

El comportamiento del algoritmo está influenciado por:

- El número de centroides (*K*) elegidos.
- La elección de los centroides iniciales.
- El orden en que las muestras son presentadas, en el caso de inicialización autónoma.
- Las propiedades geométricas de los datos.

Redes neuronales artificiales. Las Redes Neuronales Artificiales (RNAs o ANNs, en inglés, Artificial Neuronal Networks), son modelos computacionales que surgieron como intento de conseguir formalizaciones matemáticas acerca de la estructura y el comportamiento del cerebro humano. Simulan un aprendizaje a través de la experiencia. Los algoritmos desarrollados alrededor de esa idea resultaron útiles para resolver muchas situaciones de las cuales se posee un conocimiento insuficiente para

plantear una solución rigurosa. Evaluados estadísticamente logran un gran porcentaje de aciertos.

Los elementos básicos de un sistema neuronal biológico son las neuronas, agrupadas en redes compuestas por millones de ellas y organizadas a través de una estructura de capas [6]. En un sistema neuronal artificial puede establecerse una estructura jerárquica similar, posiblemente más regular que las biológicas. Las neuronas de una capa reciben estímulos solamente de las neuronas de la capa previa, si la hubiera y si no del exterior. A su vez su salida es enviada con distinto grado de intensidad a las neuronas de la capa siguiente, si las hubiera, de forma tal que una RNA puede concebirse como una colección de procesadores elementales (neuronas artificiales), conectados entre sí o bien a entradas externas y con una salida que permite propagar la señal por múltiples caminos.

Modelo de McCulloch-Pitts. Propuesto en 1943 y de salida binaria, la cual calcula la suma ponderada de sus entradas producidas por otras unidades, y da como salida un uno (1) si aquella se encuentra por encima de un umbral, o un cero (0) si está por debajo.

La figura 1 ilustra esquemáticamente una neurona según el modelo de McCulloch-Pitts, En ella se supone *j* entradas que llegan atenuadas por un coeficiente w_{ij} a una *i*-ésima neurona. Su estímulo neto es la suma de tales entradas, la función escalón tiene un umbral, si la suma lo supera da 1 si no, da 0.

Los valores de los coeficientes de esta y demás neuronas se ajustan para que tenga la red el comportamiento deseado.

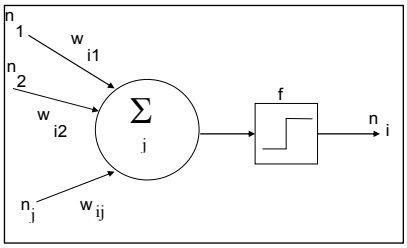


Figura 1. Modelo de neurona de McCulloch y Pitts

Este modelo ha evolucionado. Para facilitar el entrenamiento es bueno sustituir el escalón por una función analítica pero con alta pendiente en lo que sería el umbral. De esa forma enfrentando la red con casos de resultado conocido se modifican en sucesivas rondas los coeficientes, esperando que luego, frente a casos nuevos entregue el resultado correcto. Todas las salidas internas pasan a ser números reales, pasando a entero la salida final si la aplicación lo exige. En la aplicación aquí desarrollada requerimos salida real sin redondeo.

3 Antecedentes y Trabajos Relacionados

Este trabajo continúa con la línea de investigación de los proyectos PROINCE[1] C151 y C177, cuya temática se orientó: primero al estudio del tema y posteriormente a la realización de un prototipo de Sistema de recuperación de la Información, aplicando la metodología conocida como *Indexación Semántica Latente* (ISL o LSI[2]). Éste utiliza un proceso numérico llamado *Descomposición en Valores Singulares* para identificar patrones en las relaciones entre los términos contenidos en una colección de textos no estructurados y solucionar el problema de la sinonimia[3] y a la polisemia[4]. [12]

Varios autores de este trabajo, participaron en otro proyecto relacionado con la Minería de Datos: Análisis Comparativo de Modelos de Clasificación de Minería de Datos (Data Mining). Su aplicación en la predicción de perfiles de alumnos en riesgo de deserción. Proyecto PROINCE C176, en los años. 2015-2016. En la actualidad, los mismos, están llevando a cabo, desde el año pasado, una investigación cuyo título es: Uso de Minería de Datos para acelerar la recuperación de documentos.

En trabajos anteriores [9], Clusterdoc, es un sistema de recuperación y recomendación de documentos que está dirigido a usuarios con necesidades de búsqueda de información, que a través de algoritmos de agrupamiento divide el conjunto de datos en pequeños grupos con características comunes, lo cual permite minimizar el espacio de búsqueda y proporcionar información adaptada a los intereses del usuario.

Respecto a las RNA en [10], se presenta un trabajo muy completo, que nos introduce en este algoritmo en cuanto a sus definiciones, principios y tipología, en una aplicación concreta en el campo de la recuperación de la información. Se concluye aquí, que existen varias aplicaciones que explotan las características de las RNA y las aplican en nuestro campo de estudio, estableciendo que se encuentran todavía limitaciones muy grandes. El principal problema, citan, "…consiste en el volumen de procesamiento de información necesario…"y que esto se debe en parte "…a que la mayoría de las aplicaciones aquí citadas simulan el funcionamiento de una red masivamente paralela mediante un ordenador secuencial con arquitectura Von Neumann. Estas simulaciones no explotan la principal característica de las redes, el procesamiento paralelo…". Este trabajo concluye diciendo que a pesar de estas limitaciones las técnicas basadas en RNA aplicadas a la recuperación de la información, constituyen un campo de investigación muy prometedor.

Por último tenemos el trabajo de Augusto Cortez Vásquez y otros [11], que aborda el problema de una aplicación usando el algoritmo supervisado: máquinas de soporte vectorial (MVS) en el área de recuperación de información. El objetivo de esta propuesta es crear un modelo que permita etiquetar un texto con una categoría predefinida dado un conjunto de documentos D y un conjunto de categorías C, se trata de encontrar una función que haga corresponder a un documento *d* tomado de D, una

[1] Programa de Incentivos a Docentes Investigadores SPU-ME
[2] Por sus siglas en inglés, Latent Semantic Indexing.
[3] La sinonimia es una relación semántica de identidad o semejanza de significados entre determinadas expresiones o palabras.
[4] Una palabra polisémica es aquella que tiene dos o más significados que se relacionan entre sí.

categoría determinada *c* en C. Algo similar a una parte de nuestro trabajo que etiqueta cada documento con un número de grupo o clúster. En este proyecto se utilizó el análisis lexicográfico para identificar los lexemas. Constructos[5] aportes de expresiones regulares. Se realizó una comparación de subcadenas, para determinar el grado de semejanza entre dos textos a mayor subcadena en común mayor es el grado de semejanza. Por último se diseñó de la función kernel que fue utilizada, la cual determinó la eficacia de la MVS construida.

4 Uso de minería de datos en la recuperación de documentos.

El objetivo de todo sistema de recuperación de documentos es sugerir una lista de documentos ordenados de acuerdo a la probabilidad estimada de ser adecuados al requerimiento formulado. Sólo el examen de los documentos por parte del requeridor confirma el mayor o menor acierto del sistema. A esto se agrega la impaciencia propia del modo acelerado en que se vive que quiere la respuesta como si fuera instantánea.

Tal como se señaló al describir la tecnología subyacente, conforme los corpus aumentan de tamaño ese tiempo aumenta por ser este proporcional al tamaño. Para reducir ese tiempo se puede recurrir a un aumento de potencia de cómputo los que encaran ese camino recurren especialmente al paralelismo y en especial a las placas de video.

En esta investigación se analiza una línea alternativa, fraccionar el corpus. Esto requiere dos algoritmos preparatorios:

a) uno que particione el corpus utilizando una noción de vecindad o similitud y
b) el entrenamiento de un algoritmo de clasificación que direccione la consulta hacia la parte más promisoria.

Ambos servicios los estudia y provee la minería de datos.

Luego por cada consulta se debe ejecutar dos pasos:

c) aplicar el algoritmo que direcciona la consulta hacia una de las partes, para
d) enfrentar la consulta con cada documento de esa parte para determinar su grado de adecuación y posterior posición en la lista de documentos sugeridos.

Si la parte elegida efectivamente contiene un alto porcentaje de los documentos que hubieran encabezado la lista de haber hecho el proceso sobre la totalidad el usuario no sentiría demasiado la baja de la exhaustividad. Evidentemente habrá consultas cuya respuesta completa esté repartida entre varias partes, pero mientras haya en la página inicial suficientes documentos representativos de la respuesta ideal para que el usuario los examine hay tiempo de procesar otras partes y mostrarle cuando solicite la segunda página lo que hubiera faltado en la primera.

Se puede destacar algunas armonías que debe haber entre los cuatro procesos aquí señalados: Los procesos (b) y (c) se deben realizar sobre elementos ubicados en un mismo espacio conceptual. El entrenamiento del algoritmo de clasificación trabaja sobre representaciones vectoriales de los documentos. El clasificador debe recibir la consulta expresada en el mismo espacio de representación lo que aconseja someterlo a las mismas transformaciones que sufren los documentos. Por otra parte es necesario que el proceso (a) use la misma fórmula de distancia que (d). Así, si la consulta está

5 Un constructo es una construcción teórica que se desarrolla para resolver un cierto problema científico

cerca de un elemento de una partición, estará en términos comparativos cerca de todos. Se puede destacar que los procesos relacionados con clasificación podrían no utilizar la totalidad de los vectores que describen a los documentos buscando un compromiso entre velocidad y precisión.

Una manera de introducir uniformidad en el espacio de los documentos es escalar sus vectores descriptivos para tener módulo unitario. En esas condiciones las dos medidas intuitivas de distancia, una basada en el coseno del ángulo entre las direcciones y la euclídea son equivalentes desde el punto de vista práctico, como surge de la siguiente deducción aplicada a dos versores a y *b* de dimensión *d*

Partiendo del cuadrado de la distancia euclídea se llega al doble de la distancia medida a partir del coseno del ángulo entre dos versores:

$$\sum_{k=1}^{d}(a_k - b_k)^2 \;=\; \sum_{k=1}^{d}(a_k^2 - 2a_k b_k + b_k^2) \;=\; 2 - 2\sum_{k=1}^{d} a_k b_k \;=\; 2\,(1 - \sum_{k=1}^{d} a_k b_k)$$

la suma de los cuadrados en el primer paso intermedio da 1 por ser versores y cuando se quiere usar el coseno del ángulo como distancia debe ser complementado a 1 para que direcciones paralelas tengan distancia nula.

5.1 Evaluación de los resultados obtenidos

Métricas para evaluar . Es necesario introducir alguna métrica que permita apreciar la calidad del resultado obtenido y para ello hay varias propuestas. Conviene tener presente que la lista de documentos exhibidos al procesar una partición es una sublista de la lista que provee el corpus completo manteniendo el orden de ésta. Pues un documento precede a otro por su mayor afinidad sin tener en cuenta cuales son los elementos que lo acompañan.

Una primera, pensando en la aplicación a búsqueda documental analiza el escenario de uso del sistema, el usuario ingresa la consulta y obtiene una página con 10 documentos sugeridos. Mientras los examina el sistema puede seguir procesando una o más particiones adicionales y enriquecer una segunda página. Una primera estadística elegida es: “De haber procesado el corpus completo cuantos de los documentos hubieran provenido de la partición consultada, número entero que estará entre 0 y10”. Planteado en sentido inverso podría ser el décimo elemento que muestra cuán lejos estaría en la lista resultado si se hubiera procesado el corpus completo.

Experimentos realizados Se hicieron 4 experimentos con vectores de características de 5 elementos generados al azar en el rango [0, 1, 0). Los pseudo-corpora tienen 120 vectores, que fueron particionados en 4 partes y sobre ellos se procesaron 100 consultas con vectores de las mismas características. Tanto en los corpora como en las consultas los vectores se normalizaron a módulo 1.0.

Después se hizo un experimento adicional con un pseudo-corpus con 1200 vectores de largo 7 y 700 consultas.

Primera evaluación. Se la describe en forma tabular:

Tabla 1: Primer criterio de evaluación aplicado a 4 experimentos con vectores de largo 5

Exp.	0	1	2	3	4	5	6	7	8	9	10	≥ 7

0	0	1	1	4	10	11	12	15	12	18	16	61
1	0	0	0	5	10	12	10	15	18	11	19	63
2	0	0	4	4	7	12	18	18	14	10	13	55
3	0	3	1	2	6	7	14	12	18	17	20	67

Se observa que en la primera tabla que alrededor del 60% de los casos se tienen 7 o más de los documentos que hubieran entrado en la primera página de haber procesado el corpus completo.

La experiencia se repitió sobre el pseudo-corpus de 1.200 vectores de 7 elementos haciendo 700 consultas sobre el mismo, los resultados fueron porcentualmente similares:

Tabla 2: Primer criterio de evaluación aplicado a un experimento con vectores de largo 7

Exp.	0	1	2	3	4	5	6	7	8	9	10	≥ 7
4	0	3	6	31	49	75	67	89	80	123	177	469

Segunda evaluación. Se la describe en forma tabular pero incompleta ya que el décimo elemento exhibido al computar sólo la partición elegida puede estar muy alejado en una evaluación total, los valores posibles comienzan en 10

Tabla 3: Segundo criterio de evaluación aplicado a un experimento con vectores de largo 7

Exp.	10	11	12	13	14	15	16	17	18	19	20	21	22	≤ 14
0	20	13	8	8	6	6	5	4	3	7	3	3	2	55
1	19	4	9	8	6	8	4	4	4	2	1	3	7	46
2	13	6	6	10	9	5	2	8	4	2	6	4	1	44
3	20	13	8	8	6	6	5	4	3	7	3	3	2	55
4	177	89	67	55	45	32	32	24	24	22	16	25	11	433

Se observa que aproximadamente la mitad tiene su décimo elemento no más atrás de la posición 14 y en el experimento grande supera el 60%

6 Conclusiones y futuros trabajos

Los números obtenidos en las simulaciones son promisorios lo que incentiva seguir investigando para obtener porcentajes aún mejores. Al contemplar las tablas de resultados hay que tener presente que las recuperaciones se habrían obtenido en el 25% del tiempo de proceso que hubiera insumido un proceso del corpus completo.

Se destacan algunas ideas que debieran contribuir a obtener una mejora evaluación:

- En lugar de particionar recubrir el corpus con K partes, lo que resolvería el problema de documentos cercanos a la frontera de dos o más partes
- Particionar en más partes y fusionar dos o más entre los más promisorios.
- Probar con otros algoritmos de particionado prefiriendo aquellos que logren partes más equilibradas.
- Superponer dos particionados de distinta semilla y unir las partes que uno u otro hubieran recomendado.

Agradecimientos. A Cecilia Gargano por su contribución en algunos cálculos

iniciales.

Bibliografía

[1] Salton, G.: Automatic Information Organization and Retrieval. McGraw−Hill, N.Y. (1968).

[2] Seco Naveiras, D.: Técnicas de indexación y recuperación de documentos utilizando referencias geográficas y textuales. Universidade da Coruña. Dep. de Computación. (2009), <http://ruc.udc.es/dspace/handle/2183/ 7172>. Citado el: 20/06/2018.

[3] Salton, G.; McGill, M.J.: Introduction to Modern Information Retrieval, New York: McGraw-Hill, (1983).

[4] Manning, C., & Schutze, H. Chapter 11.: Probabilistic Context Free Grammars. En: Foundations of Statistical Natural Language Processing. (1999).

[5] Zazo Rodríguez Á. F. y otros.: Diseño de un motor de recuperación de la información para uso experimental y educativo. Facultad de Documentación Universidad de Salamanca (2000). <http://bid.ub.edu/04figue2.htm>. Citado el: 20/06/2018.

[6] Hernández Orallo, J., Ramírez Quintana, M.J. Ferri Ramírez, C.: Introducción a la Minería de Datos. Pearson, ISBN: 84 205 4091 9. (2005).

[7] Bedregal Lizárraga, C.: Agrupamiento de Datos utilizando técnicas MAM-SOM. Universidad Católica San Pablo. (2008). :
http://personales.dcc.uchile.cl/~cbedrega/publications/Tesis.pdf

[8] Vallejo Huanga, D.: Clustering de documentos con restricciones de tamaño. Universitario en Gestión de la Información. (2015). https://riunet.upv.es/bitstream/handle/10251/69089/ Vallejo-Clustering de Documentos con RestriccionesdeTamaño.pdf?sequence=23

[9] Giugn, M.: Clusterdoc, un sistema de recuperación y recomendación de documentos basado en algoritmos de agrupamiento. Telematique, vol 9 - nro 2. (2010). http://www.redalyc.org/pdf/ 784/78415900002.pdf

[10] de Moya Anegón, F.: La aplicación de Redes Neuronales Artificiales (RNA): a la recuperación de la información. Revistes Catalanes Obert. (1998). http://www.raco.cat/index.php/Bibliodoc/article/view/56630

[11] Cortez Vasquez, A.: Categorización de Textos mediante Máquinas de Soporte Vectorial. Revista de Investigación de sistemas e Informática. Universidad Nacional Mayor de San Marcos Facultad de Ingenieria de Sistemas e Informática. (2003). ISSN 1816-3823. http://revistasinvestigacion.unmsm.edu.pe/index.php/sistem/article/viewFile/5711/4942

[12] Venegas, R.: Análisis Semántico Latente: una panorámica de su desarrollo. Pontificia Universidad Católica de Valparaíso. Chile. Revista signos [online]. (2003), vol.36, n.53, ISSN 0718-0934. http://dx.doi.org/10.4067/S0718-09342003005300008.

[13] MacQueen, J. B.: Some Methods for classification and Analysis of Multivariate Observations. En: Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297 (1967)"

Análisis Preliminar del Rendimiento de Algoritmos para el Procesos de Descubrimiento de Reglas de Pertenencia a Grupos

Gabriel Ciciliani [1], Sebastián Martins [2], Hernán Merlino [1,2]

[1] Programa de Magíster en Ingeniería en Sistemas de información. Escuela de Posgrado. Facultad Regional Buenos aires. Universidad Tecnológica Nacional, Argentina
[2] Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información Grupo de Investigación en Sistemas de Información. Universidad Nacional de Lanús. Argentina
gabriel.ciciliani@gmail.com, smartins089@gmail.com, hmerlino@gmail.com

Resumen. En el campo de la explotación de información, el proceso de descubrimiento de reglas de pertenencia a grupos se caracteriza por la utilización combinada de un proceso de descubrimiento de grupos (clustering) y uno de inducción de reglas. Dada la variedad de algoritmos de clustering e inducción de reglas disponibles en la actualidad, es de interés poder conocer a priori qué pareja de algoritmos es más conveniente para un set de datos, en a base sus características. En este artículo se propone un diseño experimental que permita validar el rendimiento de los algoritmos, en base a métricas internas, para distintos tipos de sets de datos, con características específicas, de forma tal que permita vincular dichas características con la pareja de algoritmos que mejor rendimiento ofrece. En adición, se presentan resultados preliminares obtenidos.

Palabras Clave. Minería de datos, clustering, inducción de reglas, descubrimiento de reglas de pertenencia a grupos, estudio de algoritmos

1 Introducción

La **Explotación de Información** es una sub-disciplina de los Sistemas de Información, que brinda a la Inteligencia de Negocio las herramientas para la transformación de información en conocimiento (García-Martínez et al., 2015) y se define como la búsqueda de patrones interesantes y de reglas importantes, previamente desconocidas, en grandes cantidades de información almacenada en distintos medios (Martins, 2016). Un **proceso de Explotación de Información** consiste en un grupo de tareas relacionadas que se realizan con el objetivo de obtener información útil y significativa a partir de grandes cantidades de datos. Para esto, dichos procesos se valen de la utilización de algoritmos de Minería de Datos (García-Martínez et al., 2015; Martins, 2016).

Un procedimiento recurrente en la explotación de información es el **proceso de descubrimiento de reglas de pertenencia a grupos** (García-Martínez et al., 2015) consiste en tomar el conjunto de datos a estudiar y aplicar un algoritmo de agrupamiento o "clustering" para separarlo en distintos grupos (también llamados clases o clusters) y aplicar luego algoritmos de inducción de reglas para explicitar las

combinaciones de atributos, con sus respectivos rangos de valores, que definen la pertenencia a cada grupo descubierto (Kaski, 1997; Hall y Holmes, 2003).
Tanto para el procedimiento de *clustering* como para la de inducción de reglas existen varios algoritmos (Xu y Tian 2015; Sehgal y Garg 2014; Panchuk, 2015). Esto se explica por la necesidad de encontrar, en la diversidad de dominios de negocio, los algoritmos de explotación de información que mejor identifican los patrones ocultos en dicha masa de información.
Este trabajo se estructura en 5 secciones: la presente sección (1) ofrece una introducción, la sección (2) presenta una descripción de la problemática a abordar, la sección (3) consiste en la formulación del diseño experimental, la sección (4) muestra los resultados preliminares, finalizando con las conclusiones en la sección (5).

2 Descripción del Problema

Numerosos trabajos demuestran que la performance de los algoritmos varía notablemente tanto con las características del set de datos a analizar como por los valores de los ejemplos de dicho set (Kogan, 2007; López-Nocera, 2012; Panchuk, 2015; Sehgal y Garg 2014; Smith, Woo, Ciesielski y Ibrahim 2002). Podemos decir entonces que uno de los desafíos de la ingeniería de explotación de información es encontrar los algoritmos que mejor describen el set de datos a analizar.

Varios investigadores se han volcado al estudio de la relación entre las características de un set de datos y la performance de algoritmos relevantes en el descubrimiento de grupos (Xu y Tian 2015), inducción de reglas (Smith et al., 2002) y el descubrimiento de reglas de pertenencia a grupos (Kogan, 2007; López-Nocera, 2012; Panchuk, 2015)

Particularmente para este último proceso, los trabajos citados proponen una clasificación de los dominios a los cuales pertenecen los datos analizados y demuestran empíricamente que pareja de algoritmos es más eficaz para cada tipo de dominio. Dicha clasificación se basa en parámetros del mismo y del set de datos a analizar, como la cantidad de atributos, la cantidad de clases (grupos), la cantidad de reglas por clase, etc. En este contexto, puede apreciarse que parte de las variables utilizadas para el estudio están estrechamente vinculadas con los resultados del proceso de descubrimiento de reglas de pertenencia a grupos en sí.

Partiendo de los estudios realizados en (Kogan, 2007; López-Nocera, 2012; Panchuk, 2015) y en base a las líneas de trabajo futuras definidas, se desprende la necesidad de extender el estudio en busca de un método que permita identificar a priori qué combinación de algoritmos de clustering e inducción de reglas es el más adecuado para un set de datos dado, sin recurrir a características que dependen en sí mismas de los resultados del proceso de descubrimiento de reglas de pertenencia a grupos.

En síntesis, el presente trabajo pretende sentar las bases del estudio que permita predecir la mejor conformación de un proceso de descubrimiento de reglas de pertenencia a grupos únicamente en base a características específicas del set de datos a analizar.

3 Formulación del Diseño Experimental

Para abordar el problema definido en la sección anterior, se realizará un diseño experimental que permita comprender el comportamiento de los algoritmos de acuerdo a las características del set de datos, generando una base de conocimiento que permita posteriormente definir las parejas más eficientes.
Esto implica, en primera instancia, definir la forma mediante la cual se determinará la eficiencia del proceso. En (Feldman y Sanger, 2007) se identifican tres alternativas posibles:

- *Evaluación externa*: consiste en partir de un set de datos conocido y, aplicando diferentes parejas de algoritmos, encontrar cual es el que mejor describe la realidad. La limitación de esta estrategia es evidenciada por el objetivo del proceso de descubrimiento de grupos en sí: la búsqueda de patrones ocultos en el set de datos. No siempre se cuenta con una verdad conocida en base a la cual se pueden evaluar los distintos algoritmos.
- *Evaluación manual*: un humano experto analiza los resultados generados por cada pareja de algoritmos. La debilidad de esta estrategia radica en los tiempos/costes y la subjetividad de dichas revisiones manuales, especialmente si se considera que en muchos casos los resultados obtenidos son contra-intuitivos.
- *Evaluación Interna*: se caracteriza por definir métricas o índices que de alguna forma describan la *calidad* de los resultados obtenidos. Esta estrategia cuenta también con cierto grado de subjetividad: buenos valores de métricas no siempre significan alto valor de calidad en los resultados obtenidos. Además, existe una tendencia de ciertos algoritmos a generar buenos valores de métricas relacionadas (Van Craenendonck y Blockeel, 2015).

A pesar de dichas limitaciones, la evaluación interna no requiere conocimientos previos de los patrones que se intentan extraer de los datos, y el cálculo y evaluación de métricas puede automatizarse. De esta forma, se puede aplicar a cualquier problema de descubrimiento de reglas de pertenencia a grupos. Para mitigar la tendencia de ciertos algoritmos de *clustering* a puntuar mejor en ciertas métricas, se utilizaron cinco índices de diferente naturaleza, detallados en la sección 3.3. Como criterios adicionales, se incluyeron también métricas secundarias que permiten una segunda instancia de evaluación en caso de empate. Para la evaluación de los algoritmos de inducción de reglas, se aplica una estrategia similar, aunque utilizando una única métrica primaria y una única secundaria, descritas en la sección 3.5.
Como se describió en la sección anterior, el objetivo del experimento es poder asociar características de un set de datos con una pareja de algoritmos que retorne la mejor separación en grupos y las mejores reglas que describen esos grupos, en base a métricas internas. Para ello se siguen los siguientes pasos:

1) Generar un set de datos artificialmente con valores específicos para cada una de las características a estudiar
2) Verificar que se cumplan los valores establecidos para cada característica

3) Someterlo a un proceso de descubrimiento de reglas de pertenencia a grupos utilizando diferentes parejas de algoritmos
4) Calcular métricas internas tanto para la fase de separación en grupos como para la de inducción de reglas
5) Determinar la pareja ganadora en base a las métricas obtenidas
6) Analizar patrones en el comportamiento de las parejas de algoritmos asociados con las características de los set de datos generadas en la masa de datos experimentales.

La secuencia anterior se repite para cada set de datos, tipificado de la A a la Y (tabla 1) de acuerdo con la combinación de características (descriptas en la sección 3.1). Una vez obtenidas las relaciones entre las características del set de datos y los algoritmos que mejores resultados generan, podría ser posible determinar qué pareja de algoritmos utilizar con solo conocer ciertas características del set, sin necesidad de incurrir en un análisis comparativo de algoritmos. A continuación se describe cada fase del experimento en detalle.

3.1 Generación y validación del set de datos

Los sets de datos para este estudio son generados artificialmente y manipulados de forma que posean las características requeridas. La estructura es del tipo matricial, donde cada ejemplo del set está constituido por una cantidad fija de valores numéricos reales. La figura 1 presenta la estructura del subproceso de generación del set de datos.

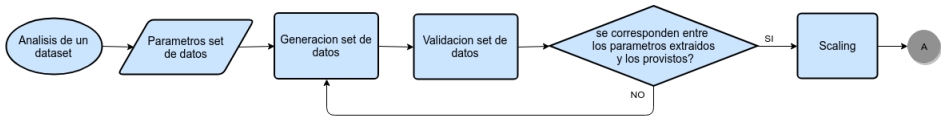


Fig 1. Diagrama de flujo subproceso de generación del set de datos

En cuanto a la distribución de los diferentes atributos, se optó por intercalar atributos "normales" con atributos "uniformes" considerando que, en la práctica, los datos no son aleatorios y responden generalmente a distribuciones conocidas. Las características del set de datos a considerar en este estudio son:

a) **Cantidad de atributos (CA):** cantidad de columnas o valores que presenta cada ejemplo del set de datos.
b) **Cantidad de ejemplos (CE):** número de filas que presenta el set de datos.
c) **Porcentaje de ejemplos con tendencia lineal (PL):** para determinar si a un ejemplo se lo considera dentro de un grupo lineal se calculará la distancia del mismo a la recta de ajuste obtenida por el método de descomposición en valores singulares (Golub, Reinsch 1970). La distancia umbral está definida como un porcentaje del módulo de la recta contenida entre los planos correspondientes a los máximos valores de cada coordenada.
d) Porcentaje de ejemplos repetidos (PR): cantidad de ejemplos que tienen exactamente los mismos valores en todos sus atributos.
e) Cantidad de grupos de ejemplos repetidos (GR): ligado directamente al punto anterior, esta característica indica cuántos grupos de ejemplos repetidos existen.

f) Porcentaje de *outliers* (PO): se consideran atípicos o *outliers* a aquellos ejemplos dentro del 20% más lejano al punto medio de la nube de puntos cuya distancia sea mayor a la distancia del ejemplo inmediatamente anterior a dicho 20%, multiplicado por un coeficiente.

Variando la composición de cada set en base a las características mencionadas se obtienen los diferentes tipos de sets a estudiar (A-Y) . La tabla 1 ilustra las combinaciones realizadas.

Tabla 1. Combinaciones de atributos utilizadas para la generación de datos.

Tipo	CA	CE	PL	PR	GR	PO
A	8	1000	N	N	0	N
B	8	1000	B	N	0	N
C	16	1000	B	N	0	N
D	24	1000	B	N	0	N
E	8	1000	M	N	0	N
F	8	1000	A	N	0	N
G	8	1000	N	B	2	N
H	8	1000	N	M	2	N
I	8	1000	N	A	2	N
J	8	1000	N	M	3	N
K	8	1000	N	A	3	N
L	8	1000	N	M	4	N
M	8	1000	N	A	4	N

Tipo	CA	CE	PL	PR	GR	PO
N	8	1000	N	N	0	B
O	8	1000	N	N	0	M
P	8	1000	N	N	0	A
Q	8	1000	B	B	2	B
R	8	1000	B	B	2	M
S	8	1000	B	B	2	A
T	8	1000	B	A	2	B
U	8	1000	B	A	2	M
V	8	1000	B	A	2	A
W	8	1000	A	B	2	B
X	8	1000	A	B	2	M
Y	8	1000	A	B	2	A

A = Alto, M = Medio, B = Bajo, N = nulo

Característica	B	M	A
PL	10%	40%	80%
PR	10%	20%	40%
PO	6%	12%	18%

Antes de someter al set de datos generado a los diferentes algoritmos, es necesario validar que las características solicitadas estén presentes utilizando el mismo método con el que luego se pretende caracterizar a los sets reales.
Si una o más características, deducidas por la rutina de análisis, difieren de las especificadas al generar el set de datos, se genera un set nuevo hasta que no haya discrepancias o se exceda una cantidad límite de intentos. Cabe aclarar que la fase de análisis contempla un margen de error de +/- 5% para las métricas expresadas como porcentajes de la cantidad de ejemplos (PL, PR y PO).
Este proceso de validación es el mismo que se utilizará posteriormente para predecir qué pareja de algoritmos es la más conveniente en base a las características del set de datos obtenidas.

3.2 Descubrimiento de grupos

Una vez generado y validado el set de datos, se lo somete a diferentes algoritmos de descubrimientos de grupo o *clustering*. La figura 2 ilustra el procedimiento aplicado.

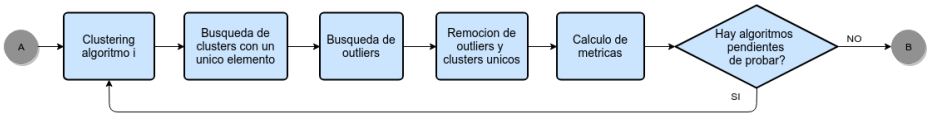


Fig 2. Diagrama de flujo subproceso de descubrimiento de grupos

Los algoritmos y variantes utilizados en este trabajo son:

1) K-Means en tres variantes de inicialización de centroides distintas:
   a) Definición de centros aleatoria,
   b) Variante k-means++ (Arthur, Vassilvitskii 2007)
   c) Determinación de centros mediante Análisis de componentes principales (Alrabea, Senthilkumar, Al-Shalabi, Bader 2013);
2) DBSCAN
3) Birch
4) Meanshift.

Para los algoritmos que no poseen una estrategia automática de configuración de sus parámetros, se utiliza un estrategia del tipo *grid-search* para determinar los parámetros de operación óptimos del estimador.

Dado que no todos los algoritmos operan de la misma forma, pueden darse situaciones donde el algoritmo *ignore* ciertos ejemplos por no poder asociarlos a un *cluster* o donde algunos de los *clusters* generados posea un único elemento. Estas dos situaciones se consideran "indeseadas" ya que se espera que ningún dato sea omitido en el análisis y un cluster de un único elemento no constituye un grupo en sí. Es por esto que tanto los grupos únicos como los ejemplos no clasificados son ignorados tanto en el cálculo de métricas como en la etapa de inducción de reglas, previo almacenamiento de la cantidad de ejemplos omitidos y el motivo, para ser considerado en la selección del algoritmo ganador, de ser necesario.

3.3 Cálculo de métricas sobre los grupos obtenidos

Los grupos descubiertos por cada algoritmo y variante mencionados son puntuados utilizando las métricas siguientes **métricas primarias**: a) Índice Davies-Bouldin, b) Índice Dunn, c) Índice Calinski-Harabasz, d) Índice silhouette y e) Suma de cuadrados. Como **métricas secundarias** se utilizan: tiempo de cómputo, cantidad de clusters de un único elemento, y puntos descartados por el algoritmo. La figura 3 resume el proceso de determinación del algoritmo ganador.

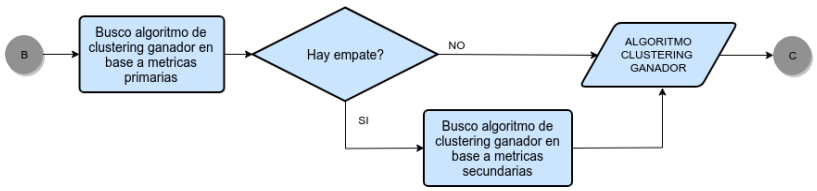


Fig 3. Diagrama de flujo subproceso determinación de algoritmo ganador (clustering)

Para obtener el algoritmo de descubrimiento de grupos ganador, se busca aquel que mejores valores tuvo para la mayor cantidad de métricas primarias. El valor de la métrica de ser al menos un 5% mayor (o menor dependiendo de la métrica) a la actual ganadora para tomar su lugar. Si existe un empate entre dos o más algoritmos, se utilizan las métricas secundarias.
Las métricas de ejemplos ignorados y grupos de un único elemento se utilizan solo si algún algoritmo finalista incurrió en dichos comportamientos. En caso de que exista un empate nuevamente, se elige uno de los algoritmos ganadores al azar para proceder a la etapa de inducción de reglas, dejando registro de los finalistas.

3.4 Inducción de reglas de pertenencia a grupos

En esta etapa se utilizan las etiquetas generadas por el algoritmo ganador para inducir reglas que describen las características de cada grupo. La figura 4 resume el subproceso de inducción de reglas de pertenencia a grupos. Los algoritmos de inducción de reglas utilizados para este trabajo son: CART y CN2. Para la determinación del valor mínimo de ejemplos por hoja óptimo, se utiliza también una estrategia del tipo Grid-search, combinado con validación cruzada o *cross-validation*, para evitar sesgos producidos por la utilización del mismo set de datos tanto en el entrenamiento como para en la evaluación.

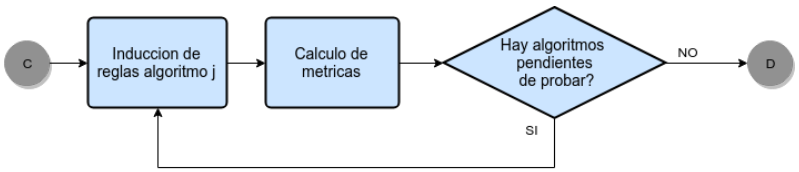


Fig 4. Diagrama de flujo subproceso inducción de reglas

3.5 Cálculo de métricas sobre las reglas generadas

Para determinar el algoritmo de inducción de reglas ganador se utiliza la curva ROC (*Receiver Operating Characteristic* o Característica Operativa del Receptor). Esta, a diferencia de otras métricas generalmente utilizadas, permite contemplar los falsos positivos, factor vital en contextos de desbalanceo de datos. Investigaciones recientes (Jurgovsky 2018) presentan estrategias para evaluar la performance de un clasificador basadas en el área bajo la curva (AUC, por sus siglas en ingleś) ROC. En este trabajo, una vez calculados los puntos de la curva ROC, se computa el área bajo la misma utilizando la regla del trapecio. A mayor valor de área, mejor será la capacidad de las reglas generadas de clasificar los ejemplos correctamente (Bradley, A 1997). La figura 5 ilustra el flujo del subproceso determinación de algoritmo de inducción de reglas ganador.
El área bajo la curva ROC es, entonces, la única métrica primaria utilizada para la evaluación de los algoritmos de inducción. El tiempo de cómputo será utilizado como métrica secundaria en caso de empate.

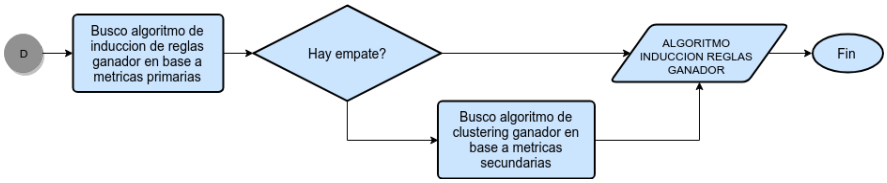


Fig 5. Diagrama de flujo subproceso determinación de algoritmo ganador (inducción de reglas)

4. Resultados preliminares

En la tabla 2 podemos observar los resultados obtenidos en base al primer lote de ejecuciones (417) del proceso realizadas hasta el momento (de un total de 1300), para cada uno de los 25 tipos de set de datos. Las filas resaltadas en gris oscuro corresponden a aquellos tipos que respondieron mejor al algoritmo ganador en más de un 70% de los casos.

En general puede observarse una baja cantidad de algoritmos ganadores determinados por tiempo de cómputo, aunque para el tipo A (ver tabla 1), ocurre aproximadamente en el 22% de los casos, siendo éste el porcentaje más alto del lote de datos. Asimismo, para los tipos resaltados, se observa que en su mayoría los algoritmos se decidieron en base a la cantidad de métricas sobresalientes. Puede observarse también un porcentaje de empates de un dígito o menos para todos los sets de datos. Ni los clusters de un único elemento, ni los ejemplos ignorados fueron un factor relevante para elegir el algoritmo ganador, como puede inferirse de los valores de la columna %CUE y %IE.

Los resultados de las corridas para los algoritmos de inducción arrojó que en el 100% de los casos, el algoritmo ganador fue CART y se destacó por tiempo de cómputo y no por presentar un área bajo la curva ROC mayor a CN2. Solo en el 2.8% de los casos, contemplando todos los tipos de sets de datos, CN2 fue elegido como ganador, en todos los casos por mostrar un mejor valor de área.

5 Conclusiones

De los resultados preliminares puede concluirse que para los tipos de datos resaltados en la tabla 2, existe una fuerte tendencia de los algoritmos ganadores a realizar una separación de los ejemplos en grupos de mejor *calidad* desde la perspectiva de las métricas internas utilizadas. Esto sugiere que el utilizar dicho algoritmo en un set de datos, con características similares a las tipificadas, tiene altas probabilidades de generar mejores grupos, comparado con los generados por los algoritmos restantes.

En cuanto al algoritmo utilizado para inducir las reglas de pertenencia a grupos, CART supera a CN2 en más del 97 de los casos, pero únicamente en performance tornandolo el algoritmo a elegir, sin importar el tipo de datos o el algoritmo de clustering utilizado en la etapa anterior.

Como futuro trabajo se prevé ampliar el análisis sobre un conjunto de datos mayor en búsqueda de patrones adicionales que vinculen características del set de datos, con las métricas y los algoritmos. Asimismo se pretende validar empíricamente los patrones obtenidos utilizando sets de datos reales. Se proponen como futuras líneas de

investigación la inclusión de nuevas características del set de datos a considerar y la ampliación de algoritmos tanto de clustering como de inducción de reglas.

Agradecimientos

Las investigaciones que se reportan en este artículo han sido financiadas parcialmente por los Proyectos 80020160400001LA y 80020160500002LA de la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Lanús.

Tabla 2. Resultados de las ejecuciones para los algoritmos de descubrimiento de grupos.

Tipo	1er Alg	% casos 1	2do Alg	% casos 2	% mp	% tiempo	% CUE	% EI	% emp
A	kmeans_++	78,5	kmeans_random	10,8	77,8	22,2	0	0	0
B	kmeans_++	69,8	kmeans_random	16,3	80,4	19,6	0	0	0
C	kmeans_++	48,6	kmeans_random	34,9	98,1	1,9	0	0	0
D	kmeans_++	44,6	kmeans_random	33	97,9	2,1	0	0	0
E	meanshift	58,7	birch	28,3	100	0	0	0	0
F	dbscan	59,9	kmeans_++	18,4	100	0	0	0	1,2
G	kmeans_++	70,9	kmeans_random	13	83,3	16,7	0	0	0,5
H	kmeans_++	53,6	dbscan	31,8	88,5	11,5	0	0	0,5
I	dbscan	68,4	kmeans_++	25	95,9	4,1	0	0	0
J	kmeans_++	60,7	dbscan	20,1	89,2	10,8	0	0	2,8
K	dbscan	87,9	kmeans_++	9,7	100	0	0	0	0,5
L	kmeans_++	65	dbscan	13,7	89,8	10,2	0	0	3,8
M	dbscan	90,4	kmeans_++	7,6	100	0	0	0	3,8
N	meanshift	92,2	kmeans_++	3,8	100	0	0	0	0
O	meanshift	92	kmeans_++	3,3	100	0	0	0	0
P	meanshift	86,1	kmeans_++	5,7	100	0	0	0	0
Q	meanshift	90,3	birch	5,7	100	0	0	0	0
R	meanshift	86,7	kmeans_++	5,5	100	0	0	0	0
S	meanshift	82,3	kmeans_++	8,6	100	0	0	0	0
T	dbscan	100	N/A	0	100	0	0	0	0
U	dbscan	99,8	meanshift	0,2	100	0	0	0	0
V	dbscan	99,5	meanshift	0,5	100	0	0	0	0
W	meanshift	53,3	birch	21,8	100	0	0	0	0
X	meanshift	56,5	birch	22	100	0	0	0	0
Y	meanshift	57,1	kmeans_++	20,6	100	0	0	0	0

***Referencias**: **1er Alg**: Algoritmo ganador; **% casos 1**: Porcentaje del set de datos que respondió mejor al algoritmo ganador; **2do Alg**: Segundo algoritmo en cantidad de set de datos mejor agrupados; **% casos 2**: Porcentaje del set de datos correspondiente al segundo algoritmo; **% mp**: Porcentaje de casos en los que el 1er algoritmo ganó en base a las métricas primarias; **% tiempo**: Porcentaje de casos en los que el 1er algoritmo ganó por tiempo de cómputo; **% CUE**: Porcentaje de casos en los que el 1er algoritmo ganó por haber generado menor cantidad de clusters de un único elemento; **% EI**: Porcentaje de casos en los que el 1er algoritmo ganó por haber ignorado una menor cantidad de ejemplos; **% emp**: Porcentaje del total de sets de datos analizado donde existió un empate entre dos o más algoritmos.*

Referencias

Alrabea, A., Senthilkumar, A. V., Al-Shalabi, H., & Bader, A. (2013). Enhancing k-means algorithm with initial cluster centers derived from data partitioning along the data axis with PCA. Journal of Advances in Computer Networks, 1(2), 137-142.

Arthur, D., Vassilvitskii S. (2007). k-means++: The advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition, 30(7), 1145-1159.

Cogliati, M., Britos, P. y García Martínez, R. (2006). Patterns in Temporal Series of Meteorological Variables Using SOM & TDIDT.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874. http://people.inf.elte.hu/kiss/13dwhdm/roc.pdf

Feldman, Ronen; Sanger, James (2007). The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge Univ. Press. ISBN 0521836573. OCLC 915286380

García-Martínez, R., Britos, P., Martins, S., & Baldizzoni, E. (2015). Explotación de Información. Ingeniería de Proyectos. Editorial Nueva Librería ISBN, 978-987-1871-34-6.

Golub, G. H.; Reinsch, C. (1970). "Singular value decomposition and least squares solutions". Numerische Mathematik. 14 (5): 403–420. doi:10.1007/BF02163027. MR 1553974.

Grosser, H., Britos, P. y García Martínez, R. (2005). Detecting Fraud in Mobile Telephony Using Neural Networks

Hall, M. y Holmes, G. (2003) Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. IEEE Transactions on Knowledge and Data Engineering, Vol. 15, Tomo 6, páginas 1437-1447.

Jurgovsky, Johannes, et al. "Sequence classification for credit-card fraud detection." Expert Systems with Applications 100 (2018): 234-245.

Kaski, S. (1997). Data exploration using self-organizing maps.

Kogan A., (2007). Integración de Algoritmos de Inducción y Agrupamiento. Estudio del Comportamiento.

López-Nocera M. (2012). Descubrimiento De Conocimiento Mediante La Integración De Algoritmos De Explotación De La Información.

Martins, S., Pesado, P., & García-Martínez, R. (2016). Intelligent Systems in Modeling Phase of Information Mining Development Process. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 3-15). Springer International Publishing.

Panchuk J. (2015) Comportamiento De Integración De Algoritmos Para Descubrimiento De Reglas De Pertenencia A Grupos. http://sistemas.unla.edu.ar/sistemas/gisi/TFLS/Panchuk-TFL.pdf

Sehgal, G., & Garg, D. K. (2014). Comparison of Various Clustering Algorithms. International Journal of Computer Science and Information Technologies, 5(3), 3074-307.

Smith, K. A., Woo, F., Ciesielski, V., & Ibrahim, R. (2002). Matching data mining algorithm suitability to data characteristics using a self-organizing map. In Hybrid information systems (pp. 169-179).

Van Craenendonck, T., & Blockeel, H. (2015). Using internal validity measures to compare clustering algorithms. In Benelearn 2015 Poster presentations (online) (pp. 1-8).

Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. Annals of Data Science, 2(2), 165-193.

Clasificación de autores para un proceso de recomendación integrado a un metabuscador científico

Cantero, A.[1], Kuna, H.[1], Rey, M.[1]

[1] Departamento de Informática. Facultad de Ciencias Exactas, Químicas y Naturales.
Universidad Nacional de Misiones
{canteroalejandro, hdkuna}@gmail.com

Resumen. La búsqueda de producción científica en la web se ha convertido en un desafío, tanto por volumen, variedad y velocidad de actualización. Se requiere de herramientas que ayuden al usuario a obtener resultados relevantes ante la ejecución de una consulta. Dentro de estas herramientas, este equipo ha desarrollado un metabuscador específico para el área de ciencias de la computación. En su evolución se pretende incorporar recomendaciones de autores para cada consulta de sus usuarios. La generación de tales recomendaciones requiere de un método que cuente con la capacidad de clasificar a los autores para definir su inclusión y posición en un listado de sugerencias para el usuario final. En este trabajo se presenta un método que cumple con este objetivo, habiendo sido evaluado y habiendo obtenido resultados que permiten plantear su inclusión en un posterior desarrollo del sistema recomendador.

Palabras clave: indicadores bibliométricos, datos científicos, autores científicos, esquema de clasificación, sistemas de recomendación.

1 Introducción

La recuperación desde la web de la producción científica es, probablemente, uno de los desafíos más importantes de la actividad científica en la actualidad. El gran volumen de datos disponibles, su heterogeneidad y su velocidad constante de actualización hacen que sea necesario contar con herramientas que faciliten el acceso a información relevante ante una consulta particular. Como respuesta a esta situación, este equipo de trabajo en particular ha desarrollado un Sistema de Recuperación de Información (SRI), específicamente un metabuscador de dominio específico, para la recuperación de publicaciones y datos científicos pertenecientes a las ciencias de la computación.

Este tipo de herramientas no son estancas, sino que constantemente están en evolución para proveer resultados de mayor relevancia y calidad a sus usuarios sin que esto perjudique su rendimiento general. Ejemplos de esto son los algoritmos de ranking [1], así como la generación automática de expansiones de consulta [2].

En este contexto, una de las mejoras propuestas para el SRI [3], es la creación de un método que permita la recomendación de autores relacionados con la consulta de un usuario, a fin de reducir los tiempos de espera mientras el metabuscador ejecuta las consultas sobre fuentes externas, generando asi una respuesta complementaria de potencial interés para el usuario.

Con este objetivo se identificaron tres tareas principales a resolver previamente: la recopilación sistemática y automatizada de los datos primarios [4], la necesidad de elaborar una estructura que sea capaz de contener de forma organizada los metadatos recolectados [5], y la clasificación de autores. Esta última, tiene por objetivo describir y sintetizar la trayectoria que tiene cada autor respecto a un área de investigación, evaluando aspectos como la actividad, la vigencia, la especificidad dentro de una temática específica y su impacto en la comunidad científica entre otras cuestiones.

El objetivo del presente trabajo es definir un método de clasificación para un sistema de recomendación de autores científicos, considerando métricas e índices utilizados en el ámbito académico para la evaluación de los autores y su producción.

El resto del artículo se estructura de la siguiente manera, en la sección 2 se mencionan los antecedentes de la línea de investigación, en la sección 3 se presenta el método de clasificación, en la sección 4 se detalla la evaluación y validación del método propuesto, en la sección 5 se mencionan las conclusiones a las que se ha arribado y en la sección 6 se lista la bibliografía consultada.

2 Antecedentes

2.1 Un Meta-buscador para las ciencias de la computación

Un producto software debe evolucionar con el tiempo, incorporar nuevas funcionalidades y optimizar sus procesos a fin de brindar una mejor experiencia a sus usuarios. En este sentido, en la actualidad el metabuscador desarrollado para la recuperación de documentos científicos del área de Ciencias de la Computación, pretende integrar diferentes procesos de explotación de información que sean de utilidad para la extracción de conocimiento a partir de los datos que se procesan a partir de las búsquedas de los usuarios [6]. El objetivo de la incorporación de estos procesos es optimizar el funcionamiento general del sistema, tanto en lo que respecta a sus procesos internos como a aquellos componentes que tienen interacción con el usuario, en una búsqueda de mejorar su experiencia con la herramienta. En lo que respecta al presente trabajo, la integración de un sistema de recomendación de autores se concibe como una solución que aportaría al usuario un conjunto de datos acotado complementario a los resultados de la búsqueda.

2.2 Sistemas de Recomendación

Los Sistemas de Recomendación (SR) son técnicas y herramientas software que sugieren ítems que podrían ser relevantes para un usuario particular. Existen tres categorías generales para los métodos que son utilizados actualmente [7] , estos están clasificados de acuerdo a la forma en que son obtenidas las recomendaciones.

Los basados en contenido recomiendan ítems en función de valoraciones de ítems similares que el usuario haya realizado previamente [8]. Los SR colaborativos elaboran sus recomendaciones en base a perfiles de usuario con similitudes que hayan valorado a un elemento previamente [9]. Y los métodos híbridos se utilizan para proveer recomendaciones usando conceptos de ambos enfoques previos, superando así algunas de sus limitaciones.

2.3 Clasificación de autores para un recomendador

Un criterio típico de recomendación es la devolución de un conjunto de documentos en el cual todos sus elementos tienen un atributo con un valor que coincide total o parcialmente con el interés de un usuario. Sin embargo, únicamente retornar un conjunto desordenado de autores que mostraron alguna coincidencia con los intereses del usuario podría no ser un criterio efectivo. Por ello se consideró el desarrollo de un método capaz de organizar y clasificar a los autores considerando los aspectos más relevantes para la comunidad científica.

En una tarea de relevamiento bibliográfico se encontraron estudios enfocados en la evaluación del rendimiento de los autores en un nivel institucional [10] y sobre los factores que afectan su desempeño [11, 12]. Se observó que plataformas científicas como Microsoft Academic Search [13] (MAS) pone a disponibilidad de sus usuarios una sección de ranking de autores clasificados por "TOP" el cual permite discriminar por área de investigación, índice o puntaje y una ventana de tiempo. El equipo de investigación detrás de la plataforma AMiner[14] ha trabajado en el desarrollo de métricas que permiten evaluar a los autores por dimensiones o aspectos relevantes y en base a ello permite a sus usuarios una amplia gama de filtros para que estos puedan encontrar algún autor experto o influyente dentro de un área de investigación. Algunos autores han estudiado y analizado la trayectoria de los autores desde una dimensión con el fin de determinar el grado de utilidad de las métricas o del estado del arte de un área específica. Otros optaron por desarrollar modelos predictivos para detección de potenciales autores emergentes haciendo uso de reconocidas y nuevas métricas bibliométricos. Sin embargo, a pesar de los trabajos hallados no se encontró una taxonomía estandarizada de la trayectoria de los científicos y por ello se propone una clasificación de autores que considere los aspectos más relevantes de la trayectoria según las dimensiones más relevantes para la comunidad y que sea capaz de adaptarse a las necesidades de un proceso de recomendación.

3 Desarrollo del método de clasificación de autores

3.1 Cuestiones relativas a los datos

La sistematización de la recuperación de datos primarios para el método de clasificación de autores se estructuró a partir del trabajo realizado previamente sobre el metabuscador. Se utilizaron los procedimientos de Extracción, Transformación y Carga (ETL por su sigla en inglés) de los datos de las entidades involucradas en los procesos del metabuscador [4]. Estos datos fueron almacenados siguiendo los lineamientos planteados para la definición de los perfiles de las entidades con las que opera el SRI. Puntualmente, para este trabajo son de especial importancia los perfiles de los autores generados a partir de las fuentes de consulta habitual del metabuscador [5].

3.2 Métricas y aspectos evaluables de autores

Como se describió en la sección 2.2 para el logro de los objetivos del SR es necesario definir qué criterios se utilizarán para la evaluación de los autores. En [15] se describen

todos los posibles aspectos evaluables de un autor científico detallando para cada caso las técnicas disponibles y más utilizadas para su cuantificación. Considerando [14, 15] se elaboró la Tabla 1 que resume los aspectos que resultan más relevantes para la evaluación de la trayectoria.

Los índices bibliométricos son conocidos y utilizados por toda la comunidad científica en general. Dado que cada indicador puede presentar ventajas y desventajas en su utilización, se volvió frecuente en la comunidad, la adopción de más de un indicador en la evaluación de los autores [16]. Siguiendo esta línea de pensamiento, en

Tabla 1. Aspectos evaluables de la trayectoria de los científicos.

Aspecto	Definición	Técnica
Actividad	Evaluación de la producción científica, su circulación, dispersión y el impacto que ha producido en la comunidad.	El recuento de las publicaciones considerando el espacio de tiempo transcurrido entre las mismas.
Diversidad	Evaluación del grado de especificidad dentro de un campo, debido que un autor puede estar involucrado en diferentes campos de investigación.	Determinar la especificidad de un autor respecto a un área temática obteniendo una proporción entre la cantidad de artículos dedicados a un área específica con respecto al total de su producción.
Vigencia	Relativo a la vida media de un artículo considerando su citación a través del tiempo.	El recuento de las citaciones obtenidas en las publicaciones utilizando ventanas de tiempo.

este trabajo se propone el uso de una combinación de índices estratégicamente organizados que permitan evaluar más de un aspecto relevante. De esta manera, con el fin de desarrollar un conjunto de métricas primarias y derivadas de los índices bibliométricos más conocidos para la evaluación de autores se propone utilizar los siguientes índices:

Índice de Impacto Promedio: a fin de generar un índice que permita reflejar un impacto unificado de la producción de un autor se procedió al análisis de los índices H, G e i10 [17–19] de una muestra de 73[1] autores. De este análisis se observó la correlación que existe entre estos índices en reflejar la trayectoria de los autores, esto se debe a que todos ellos se basan en la cantidad de citas que han obtenido las publicaciones del autor.

Observando esta situación se propone un índice que promedie los anteriores obteniendo uno simplificado para la evaluación del autor definida por la ecuación (1).

$$IIP(a_n) = \frac{(indiceH(a_n) + indiceG(a_n) + indiceI10(a_n)}{3} \quad (1)$$

Índice de Longevidad: este índice se propone a fin de representar la vida académica del autor. Se define este índice (2) como la diferencia entre el último año de publicación y el año de su primera publicación.

[1] Muestra compuesta por índices bibliométricos de autores dedicados al campo de minería de datos. Fuente: elaboración propia con datos de MAS, query: "*datamining*".

$$IL(a_n) = (\text{año } del \text{ último } paper) - (\text{año } del \text{ } primer \text{ } paper) \quad \textbf{(2)}$$

Tasa de Publicaciones por Ventana de Tiempo: a fin de reflejar el desempeño de un autor dentro de una ventana de tiempo, se define este índice (3) como la relación entre el número de trabajos de un autor publicados entre los años de inicio y fin que determinan la ventana de tiempo.

$$\text{Sean } i \in (añoInicio,..,añoFin) \; ; \; p \in publicaciones \; ; \quad \textbf{(3)}$$

$$P(a) = \{p_i \text{ es publicación del autor } a\} \; ; \; TPpVT(a_n) = \frac{\text{count}(\text{P}(a_n))}{añoFin - añoInicio}$$

Índice de Diversidad Relativo: el índice de diversidad (ID) definido en [14] intenta reflejar el grado de participación que tiene un autor en relación a diversos campos de investigación dentro de una ciencia o área de conocimiento. En este índice los valores más altos corresponden a aquellos autores que tienen un alto grado de participación en varios campos. Con el fin de evaluar la diversidad de un autor según un contexto determinado, se propone un índice que calcule el grado de diversidad de un autor con respecto a otros autores (4). Para ello se calcula la distancia que existe entre el mayor índice de diversidad de una colección con respecto a un determinado autor, de esta manera el más cercano y el más lejano estarán próximos a 1 y 0 respectivamente.

$$IDR(a_n) = \frac{ID(a_n)}{max(ID(a_1), ID(a_2), \ldots, ID(a_z))} \quad \textbf{(4)}$$

3.3 Clasificación de Autores

Dado que no se encontró en la literatura una clasificación estandarizada de los autores en función de su trayectoria científica, se propuso la definición de un conjunto de clases que se consideraron representativas de las diferentes variedades de autores. A estas clases se vincularon indicadores que permiten evaluar distintos aspectos del autor en función de la especificidad de cada clase definida. Y se estableció para cada índice un umbral que permita determinar si un autor cumple o no todas las condiciones para pertenecer a una clase.

Se propone un esquema de clasificación compuesto por 7 (siete) clases que sintetizan la trayectoria de los autores y los aspectos considerados de interés dentro de la comunidad científica, organizados en las siguientes agrupaciones:

Autores Nuevos: Son aquellos autores que su tiempo de actividad en la labor científica es reducida y que han tenido un impacto considerable con respecto a sus colegas con igual tiempo de actividad. Restricciones aplicadas:

(a) Índice de longevidad entre 1 y 5 años.

(b) La Cantidad de Citas deberá ser igual o superior al percentil 30 de la Cantidad de Citas de todos los autores que cumplen con la condición (a).

Autores Top: Son aquellos autores que han mostrado un impacto por encima de la media de sus colegas, destacándose la comunidad científica. Restricciones:

(c) IIP mayor o igual al 150% del valor del desvío estándar obtenidos a partir de la serie de IIP de todos los autores.

Autores Expertos: Son aquellos que han trabajado en proyectos de un campo determinado y son referentes en los mismos debido al impacto de sus trabajos. Restricciones:

(d) El IDR del autor deberá estar por debajo del 45%. Excluyendo del cálculo los valores que superen el Límite Superior considerados valores anómalos.
(e) El IIP deberá ser superior al percentil 40 del IIP de todos los autores excluyendo los valores que superen el Límite Superior por ser considerados anómalos.

Autores de Moda: Son aquellos autores que han realizado trabajos nuevos o innovadores en un campo de conocimiento que le han provisto de un impacto considerable en la comunidad en los últimos años. Restricciones:

(f) El autor no podrá ser moda si cumple con las restricciones de un autor Experto.
(g) El *TPpVT* de los últimos 5 años deberá ser mayor al valor de toda su carrera.
(h) El IIP del autor deberá ser mayor al primer cuartil (1Q) del IIP de todos los autores que cumplen con la condición (g).

Autores con Trayectoria: Son aquellos autores que han desarrollado una carrera como investigadores y han realizado contribuciones con impacto dentro de las ciencias de la comunidad científica. Restricciones:

(i) IIP deberá ser mayor al segundo cuartil (2Q) del IIP de todos los autores.

Autores One Hit Wonder: Son aquellos autores que han logrado un impacto en la comunidad por en un momento acotado, pero no se ha observado continuidad en su labor científica, al menos desde lo que respecta a publicaciones de su autoría. Restricciones:

- El Índice de longevidad deberá ser igual a 1 año.
- El IIP deberá ser mayor a la mediana o 2Q del IIP de todos los autores con longevidad menor o igual a 5 años.

Autores Sin Clasificación: Estos autores no corresponden con ninguna de las agrupaciones anteriores, razón por la cual se decidió agruparlas en una clase que permitan estudios posteriores.

3.4 Algoritmo propuesto para la clasificación de autores

Aunque preliminarmente se puede llegar a la conclusión de que el proceso de clasificación puede consistir en la aplicación directa de los axiomas sobre un conjunto de autores, realizarlo sin una estrategia que permita el cálculo de varios índices y la comprobación de todas las condiciones para cada una de las clases se incurriría en un uso ineficiente del proceso de cómputo. Si a esto se añade que existen clases con mayor especificidad en la evaluación de los autores, criterio reflejado en el número de índices utilizados y los umbrales de aceptación, ocurre que algunas clases son más específicas que otras que evalúan aspectos más generales. Esta situación hace que un autor puede estar en los umbrales de aceptación de más de una clase ya que puede cumplir con las condiciones necesarias, sin embargo, ante esta situación se priorizará evaluar y clasificar al autor en función de aquellas clases más específicas que coincidan con su perfil.

Con motivo de dar solución a estas cuestiones se propone un algoritmo de clasificación de autores a partir de un árbol de decisión booleano ad-hoc que dado una secuencia específica en la evaluación de los datos disponibles permita una clasificación eficiente.

```
Variables: autor_i: i-ésimo autor de una colección.
autores_sin_clasificar: vector cuyos elementos corresponden a
autores no clasificados.
IIP_nuevos: vector con los IIP de autores con menos de 5 años de
actividad académica.
IIP_todos: vector con los IIP de todos los autores.
IIP_expertos_LLS: vector con los IIP de los autores por debajo
del límite superior respecto a todos los autores.
IIP_moda: vector con los IIP de los autores que tienen un
promedio de publicaciones mayores en los últimos 5 años.
CC_nuevos: vector con la cantidad de citas de los autores con menos
de 5 años de actividad académica.
Por cada autor_i en autores_sin_clasificar:
  si calcLongevidad(autor_i) <= 1
   and IIP > calc2Cuartil(IIP_nuevos)
     clasificarComoOneHitWonder(autor_i)
  sino calcLogevidad(autor_i) <= 5
   and calcPercentil30(CC_nuevos)>= cantidadDeCitas(autor_i)
     clasificarComoNuevo(autor_i)
  sino (calcIndiceDeDiversidad(autor_i)
   and calcIIP(autor_i) < calcLimiteSuperior(IIP_todos)
   and calcDiversidadRelativa(autor_i) < 0.45
   and calcIPP(autor_i) >= calcPercentil40(IIP_expertos_LLS))
     clasificarComoExperto(autor_i)
  sino calcPromPubs5anios(autor_i) >= calcPromPubsTotalAños(autor_i)
   and calc1Cuartil(IIP_moda) < calcIPP(autor_i)
     clasificarComoModa(autor_i)
  sino calcIPP(autor_i) >= 1.5 * calcDesvioEstandar(IIP_todos)
    clasificarComoTop(autor_i)
  sino calcIPP(autor_i) > calc2Cuartil(IIP_todos)
    clasificarComoTrayectoria(autor_i)
  sino
    clasificarComoSinClasificacion(autor_i)
fin
```

4 Experimentación y Resultados

En la literatura del área se pudo observar que los métodos de clasificación son evaluados en líneas generales según dos dimensiones: el esquema de clasificación en sí mismo y las decisiones de clasificación [20].

En [21] se plantea que cuando un esquema de clasificación tiene el menor grado de incertidumbre posible, más efectivas serán sus decisiones de clasificación. Esto se logra cuando se tiene el mayor conocimiento acerca del dominio en el que se está desarrollando. Por tal motivo se realizó una evaluación del esquema analizando el grado de incertidumbre para cada dimensión:

- Objeto: Las propiedades de los autores para generar la clasificación están basados en índices bibliométricos de amplio uso y pueden ser generados a partir del conocimiento de las publicaciones y cantidad de citas de cualquier autor.
- Clase: Debido que la clasificación es propuesta en este trabajo, se tiene conocimiento de todas las restricciones definidas por cada clase, por lo que resulta claro y preciso.
- Esquema: En esta dimensión se presenta un grado de incertidumbre medio, debido a que no existe un estándar único establecido que permita una definición ampliamente aceptada por la comunidad. El grado de precisión en la definición de las clases se sustenta en un análisis subjetivo relacionado a la búsqueda de representar en las clases el interés que se podría tener por un autor en el contexto de aplicación del método de clasificación.

Para la evaluación de las decisiones de clasificación se siguió el enfoque propuesto en [21] que propone evaluar la bondad de ajuste de cada clase al evaluar autores. En este caso, el proceso de evaluación consistió en la ejecución de una serie de algoritmos de clasificación automática sobre un conjunto de datos de autores previamente clasificados utilizando el método planteado en el presente trabajo. El objetivo en esta tarea consistió en verificar si un método automático podría ser entrenado en base a la clasificación generada y luego evaluar la efectividad de los resultados. En caso de obtener valores de efectividad elevados, se podría postular que las clases y el algoritmo de clasificación ad-hoc desarrollados son consistentes.

Siguiendo este enfoque se conformó una base de datos de autores extraídos de las plataformas MAS y AMiner mediante un proceso automatizado de ETL desarrollado para el metabuscador. El método de recolección de los datos se inició utilizando como punto de partida los autores de los artículos científicos que han sido recuperados por las consultas de los usuarios, los datos recuperados están conformados por los índices bibliométricos, publicaciones y coautores disponibles en la plataforma de origen. Posteriormente se procedió a la evaluación de la calidad de los datos recuperados con el fin de determinar el grado de correctitud, completitud y consistencia mediante técnicas de muestreo aleatorio y el "estándar de oro". De esta manera el dataset quedo conformado por un total de 2164 perfiles de autores siguiendo la estructura definida en [5]. Posteriormente, se seleccionó una serie de métodos de clasificación, algunos de la familia TDIDT[2] como son: C4.5, Random Forest, Simple CART y Rule Induction (RI), en forma conjunta con una red neuronal de tipo perceptrón multicapa y un clasificador bayesiano implementados en la herramienta WEKA. Los parámetros de los métodos empleados fueron establecidos con los valores por defecto que plantea la herramienta, siempre que el dataset no requiriera algún tipo de ajuste por las características del mismo. Los resultados obtenidos del proceso de entrenamiento-testeo (proporción 70/30 del dataset) de los métodos mencionados se detallan en la tabla 2, en el cual se observan resultados con una eficiencia entre el 81.82% al 95.22% de efectividad en la clasificación y una tasa de error por debajo del 20% para el peor de los casos. En los casos de baja efectividad, se observó que la clase que presenta el mayor problema de precisión es *Moda* debido que el algoritmo confunde a la clasificación de los autores con la clase *Trayectoria*, esto se debe probablemente a lo descrito previamente relativo al solapamiento entre las clases más específicas y las generales.

[2] TDIDT sigla para "Top Down Induction of Decision Trees".

Tabla 2. Resultados algoritmos de clasificación automática.

Criterios de Evaluación	C4.5	RF	Simple CART	RI	RNA-MC	Bayes.
Accurancy	95.22%	95.22%	94.3%	92.45%	81.97%	81.82%
Clasification Eror	4.78%	4.78%	5.7%	7.55%	18.03%	18.18%
Weighted Mean Precision	95.28%	95.23%	94.56%	78.74%	74.95%	83.28%
Weighted Mean Recall	94.38%	94.08%	88.92%	79.63%	74.82%	76.61%

Para el recall o exhaustividad se observa que las clases *Moda* y *OHW* presentan el menor porcentaje de efectividad con respecto a las clases Top y Nuevo respectivamente, esto se debe probablemente a que estas clases son similares en ciertos aspectos, sin embargo, las restricciones establecidas por clase definen su pertenencia a una u otra.

5 Conclusiones y Líneas futuras de investigación

Se ha generado un método de clasificación para autores científicos en el marco de un SRI del área de ciencias de la computación. El método presentado incluye la definición de una serie de clases que representan diferentes estadios en la carrera de un investigador, siendo contemplados diversos factores como el impacto generado, su especialización y su trayectoria. El desarrollo presentado ha sido validado en términos de su definición y pertinencia, siendo que sobre un dataset en el que fue aplicada la clasificación se probaron algoritmos de clasificación automática, obteniendo resultados de alta efectividad y baja tasa de errores.

El método se ajusta a las necesidades de un SR de autores para un metabuscador de tipo científico del área de ciencias de la computación. Su implementación permitirá presentar perfiles de autores de relevancia para los usuarios a partir de sus consultas.

Como líneas futuras se proponen, inicialmente, la implementación del sistema de recomendación para el metabuscador. Además, la transición del método desarrollado a uno que integre técnicas de clasificación automática a fin de generalizar y optimizar los procesos involucrados, evaluando la factibilidad de incluir en el análisis una mayor cantidad de indicadores.

6 Bibliografía

1. Kuna, H.D., Rey, M., Martini, E., Solonezen, L., Sueldo, R.: Generación de un algoritmo de ranking para documentos científicos del área de las Ciencias de la Computación. Presented at the XVIII Congreso Argentino de Ciencias de la Computación , Mar del Plata, Argentina October (2013).
2. Rey, M., Kuna, H.D., Martini, E., Podkowa, L., Pautsch, J.G.A., Zamudio, E.: Generación de un método de expansión de consultas basado en ontologías para un sistema de recuperación de información. Presented at the XX Congreso Argentino de Ciencias de la Computación , La Matanza, Argentina (2014).
3. Rey, M., Kuna, H.D., Martini, E., Canteros, A., Cantero, A., Rambo, A., Biale, C.O.: Propuesta de esquemas de perfiles para la recuperación de datos científicos para un sistema

de recuperación de información del área de Ciencias de la Computación. Presented at the XXII Congreso Argentino de Ciencias de la Computación , Junín, Argentina (2016).
4. Rey, M., Kuna, H.D., Rambo, A., Canteros, A., Cantero, A., Martini, E., Corrales, N., Rauber, F.: Propuesta de procesos complementarios para un sistema de recuperación de información. Presented at the XIX Workshop de Investigadores en Ciencias de la Computación , Buenos Aires August 23 (2017).
5. Kuna, H., Rey, M., Zamudio, E., Olivas, J.A., Rambo, A., Cantero, A., Canteros, A., Martini, E., Biale, C.O.: An entity profile schema for data integration in an academic metasearch engine. In: International Conference on Artificial Intelligence. , Las Vegas, Nevada (2017).
6. Kuna, H.D., Rey, M., Zamudio, E., Canteros, A., Cantero, A., Rambo, A.R., Martini, E., Pautsch, J.G.A., Biale, C.O., Krujoski, S., Rauber, F.: Diseño y construcción de procesos de explotación de información para el área de ciencias de la computación. Presented at the XX Workshop de Investigadores en Ciencias de la Computación, WICC 2018 (2018).
7. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering. 17, 734–749 (2005).
8. Mooney, R.J., Roy, L.: Content-based Book Recommending Using Learning for Text Categorization. In: Proceedings of the Fifth ACM Conference on Digital Libraries. pp. 195–204. ACM, New York, NY, USA (2000).
9. Rich, E.: Stereotypes and User Modeling. In: Kobsa, A. and Wahlster, W. (eds.) User Models in Dialog Systems. pp. 35–51. Springer Berlin Heidelberg (1989).
10. Carpenter, M., Gibb, F., Harris, M., Irvine, J., Martin, B., Narin, F.: Bibliometric profiles for British academic institutions: An experiment to develop research output indicators. Scientometrics. 14, 213–233 (1988).
11. McGrail, M.R., Rickard, C.M., Jones, R.: Publish or perish: a systematic review of interventions to increase academic publication rates. Higher Education Research & Development. 25, 19–35 (2006).
12. Costas, R., van Leeuwen, T.N., Bordons, M.: A bibliometric classificatory approach for the study and assessment of research performance at the individual level: The effects of age on productivity and impact. J. Am. Soc. Inf. Sci. 61, 1564–1581 (2010).
13. Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J. (Paul), Wang, K.: An Overview of Microsoft Academic Service (MAS) and Applications. In: Proceedings of the 24th International Conference on World Wide Web. pp. 243–246. ACM, New York, USA (2015).
14. Tang, J.: AMiner: Mining Deep Knowledge from Big Scholar Data. In: Proceedings of the 25th International Conference Companion on World Wide Web. pp. 373–373. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2016).
15. Velasco, B., Bouza, J.M.E., Pinilla, J.M., Román, J.A.S.: La utilización de los indicadores bibliométricos para evaluar la actividad investigadora. Aula abierta. 40, 75–84 (2012).
16. Costas, R., Bordons, M.: Is g-index better than h-index? An exploratory study at the individual level. Scientometrics. 77, 267–288 (2008).
17. Hirsch, J.E.: An index to quantify an individual's scientific research output. PNAS. 102, 16569–16572 (2005).
18. Egghe, L.: Theory and practise of the g-index. Scientometrics. 69, 131–152 (2006).
19. Aguillo, I.F.: Is Google Scholar useful for bibliometrics? A webometric analysis. Scientometrics. 91, 343–351 (2012).
20. Ranganathan, S.R.: Prolegomena to library classification. Madras Library Association, Madras (1937).
21. Bedford, D.: Evaluating classification schema and classification decisions. Bulletin of the American Society for Information Science and Technology. 39, 13–21 (2013).

Incremento de Semántica en Fases Iniciales de Transformación de Modelos de Datos

Viviana Ferraggine[1], Federico Casanova[2] y Pablo Rinaldi[2]

1 INTIA, Facultad de Ciencias Exactas, U.N.C.P.B.A., Paraje Arroyo Seco
Tandil, Buenos Aires, Argentina
2 PLADEMA, C.I.C. - U.N.C.P.B.A., Paraje Arroyo Seco
Tandil, Buenos Aires, Argentina
{vferra, fcasano, prinaldi} @exa.unicen.edu.ar

Abstract. El Diagrama de Entidades y Relaciones Extendido es una pieza clave en el modelado de datos, a partir del cual mediante un proceso de transformación genera el Esquema Físico de Datos basado en un Modelo Objeto-Relacional. En el presente trabajo se identifican y se proponen soluciones a varios problemas encontrados en dicho proceso, originados por la falta de herramientas y metodologías que permitan expresar la semántica de determinadas situaciones. Varias de estas soluciones son aun objeto de especificación, en este trabajo se detallan las fases iniciales del proceso de transformación sobre las cuales ya se ha cubierto completamente dicha especificación y se describen las transformaciones necesarias en las siguientes fases.

Keywords Modelo de Entidades y Relaciones Extendido, Esquema Lógico Estandar, Esquemas Físico de Datos, SQL.

1 Introducción

Las fases de modelado conceptual, lógico y físico de datos son fundamentales en el proceso de diseño de bases de datos relacional u objeto-relacional. Para lograr un diseño completo y exacto, se necesitan metodologías que permitan representar y trasladar todos los aspectos relevantes del Universo de Discurso (UdeD) al esquema físico final [2], [3], [6], [10].

El modelado conceptual de datos permite a los diseñadores reconocer y representar los objetos del negocio y los vínculos semánticos entre ellos, y sobre esa base poder proporcionar un mapeo directo entre el mundo real percibido y su representación, sin distorsión o ambigüedad [16].

Tanto el Modelo de Entidades y Relaciones original (MER) [5], como su extensión más difundida, el Modelo de Entidades y Relaciones Extendido (MERExt), simbolizan conceptos de datos en términos de entidades y relaciones [11], [23], [24]. El minimalismo de su notación ha contribuido a su éxito, proporcionando

mecanismos de abstracción suficientes para especificar ciertas reglas del negocio y restricciones inherentes, y los lenguajes gráficos o visuales como el Diagrama de Entidades y Relaciones Extendido (*DERExt*) facilitan la discusión, comunicación y validación por parte de los expertos en el dominio [7], [8]. Dichos *DERExt* pasarán luego a formar parte de la documentación de diseño y desarrollo del proyecto, resultando en piezas claves a futuro para el entendimiento de la semántica del modelo de datos plasmado en el esquema físico.

Muchos autores han escrito acerca de las reglas de transformación del *DERExt* al Esquema Relacional (Esq R) proponiendo algoritmos con diferentes variantes [4], [9], [11], [14], [23], [24]. En particular, algunos reconocen la etapa de diseño lógico estándar, cuya definición resulta la más apropiada para la etapa de transformación del *DERExt* a un Esquema Lógico Estándar (*ELE*) inicial, basado en un Modelo Objeto-Relacional [9]. Estos algoritmos resultan mejorables ya que una variedad de conceptos plasmados en el *DERExt* se pierde al obtener el *ELE*, y por su incapacidad de permitir, de ser necesario, un proceso inverso de reconstrucción sin ambigüedades de un *DERExt* a partir de un *ELE* o de un Esquema Físico de Datos (*EFD*).

Como mencionamos, el problema surge cuando se intentan aplicar las reglas de transformación y mantener la semántica plasmada en el *DERExt* en la siguiente fase de la metodología, esto es debido a que en su gran mayoría utilizan el modelo relacional o el modelo binario para transformar el *DERExt* en un *EFD*. Las reglas que se encuentran en los textos por lo general sólo especifican transformaciones de elementos simples del *DERExt*, esto conlleva a una pérdida importante de semántica en dicho proceso, dejando en manos de quien implementa el esquema físico en la base de datos la definición de mecanismos para controlar y comprobar los aspectos que no pudieron ser representados.

Si bien se han propuesto innovaciones en la notación del *ELE* [17], basándose en reglas de construcción del *DERExt* y extendiendo el algoritmo para su conversión en un *ELE* enriquecido [18], [19], aún hay aspectos y construcciones que no han sido suficientemente explorados ni analizados.

En este trabajo se propone una forma de representar en el *ELE* todos los aspectos incluidos en el *DERExt*, enriqueciendo la notación de las reglas conocidas para la construcción del mismo, y extendiendo el algoritmo de transformación a fin de resolver las cuestiones relativas a ambigüedades y falta de ortogonalidad y cumplir plenamente con las cualidades de completitud que se pretenden de la metodología. Se tomaron como base las reglas de transformación más difundidas [4], [11], [15], [21], [22], [23], [24]. Esto permite que dicho *ELE* pueda ser traducido automáticamente a un *EFD*, utilizando los recursos del lenguaje SQL declarativo. Esta traducción se realiza pasando por una etapa intermedia denominada Esquema Lógico Estándar Tardío (ELET) que permite incorporar ya decisiones de implementación respecto del SQL a utilizar, sin perder el nivel de abstracción en la etapa final.

2 Propuesta de Mejoras al Proceso de Transformación del Modelo Conceptual de Datos a Esquema Físico de Datos

El proceso de transformación convencional desde un *DERExt* hasta un *EFD* implementado en un estándar SQL declarativo se puede dividir en 3 etapas [23]:

1. Transformación del *DERExt* al *ELE*: es el proceso de transformación de los conceptos semánticos modelados, basada en los elementos constitutivos del *DERExt*, entidades, relaciones y atributos con la información de todos sus componentes.
2. Transformación del *ELE* al Esquema Lógico Estándar Tardío (*ELET*): es el proceso de Diseño Lógico Tardío, donde se identifican, caracterizan y especifican las decisiones relacionadas con la forma de implementación de atributos multivaluados, atributos compuestos, relaciones 1:1, jerarquías y otras construcciones que dependen en gran medida de cuestiones físicas de implementación en una determinada clase de base de datos.
3. Transformación del *ELET* al *EFD*: es el proceso en el cual se toman el *ELET* y se lo transforma en SQL estándar.

A continuación, se enuncian las propuestas de mejoras sobre las diferentes transiciones en las transformaciones descriptas. Dichas propuestas se basan principalmente en adicionar en cada una de las transformaciones la traza de las decisiones de diseño que se van aplicando, exceptuando aquéllas que pueden ser deducidas.

2.2 Transformación del *DERExt* al ELE

En aras de facilitar la representación de una mayor variedad de situaciones del *DERExt*, en esta sección se propone un conjunto de reglas para una notación que refleje todos los conceptos semánticos modelados, basada en los elementos constitutivos del diagrama, sus extensiones y en las reglas de transformación incluidas en los algoritmos más difundidos [4], [11], [24].

El algoritmo propuesto combina tres aspectos clave: la formación sistemática de nombres de componentes del *DERExt* y conceptos, el orden en el que aparecen en cada esquema de componente y una especificación precisa de todos los conceptos relevantes del modelo.

Se adicionaron elementos que, aun no siendo estrictamente relacionales, permiten que el usuario experto pueda obtener un *ELE* estrictamente relacional, o bien que opte por uno post-relacional (con características de objetos) [1], [4], [13], [16]. En lo que sigue, se presentarán los algoritmos de conversión en un *ELE* basado en las transformaciones del *DERExt*.

Formalmente un *ELE* es un conjunto de componentes C donde C = {C1, C2, ..., Cm} representa a cada uno de los componentes típicos del modelo y además restricciones de unicidad, de no nulidad, dependencias funcionales, restricciones de integridad referencial, restricciones de cardinalidad y otras restricciones de integridad. Un esquema de componente se expresa $esq(C_i)$ y representa el conjunto de conceptos que definen el componente C_i [17], [18].

Se define como concepto a cada una de las partes constitutivas de *esq(C$_i$)*. Se propone conservar un orden prefijado para cada concepto resultante de la transformación; se utilizarán estratégicamente paréntesis '( )', y llaves '{ }' para definir conjuntos de conceptos del mismo tipo; punto y coma ';' como separador de conceptos diferentes, y comas simples ',' como separadores de conceptos de un conjunto o bien de las partes de un concepto. Cuando no haya existencia de un determinado concepto, su lugar deberá ser respetado colocando el ';' correspondiente. El único concepto que debe estar presente sin excepciones es el del identificador de cada componente, los restantes podrían estar ausentes. Cuando un concepto pueda admitir nulos, se indicará precedido por '¬'.

2.2.1. Reglas de Transformación para Entidades (*DERExt* – ELE)

En la aplicación de las reglas para obtener *esq(C)* siendo C un tipo de entidad, las características que deben preservarse son el nombre del tipo de entidad y su contexto: regular (R), subtipo (SB), supertipo (SP), supertipo de una jerarquía de más de un nivel (SPB) o débil (D). Los conceptos correspondientes son: lista de todos sus atributos con sus diferentes características y multiplicidades, clave primaria y claves alternativas. En el proceso estándar, los conceptos y el nombre se preservan aplicando los algoritmos convencionales de conversión, sin embargo, el contexto se infiere en función de particularidades de la clave primaria, incluyendo las dependencias en las que se la involucra, lo que genera algunas de las problemáticas ya mencionadas.

Se propone incorporar un prefijo al esquema lógico de las entidades, que se corresponda con la clasificación de contextos de las entidades ya mencionada (R, SB, SP, SPB ó D). Esto tiene el doble propósito de, por un lado, evitar las ambigüedades que surgen en los procesos de reingeniería (por ejemplo, la distinción entre entidades débiles y atributos multivaluados o relaciones unarias y jerarquías), y por otro, contribuir a la posible implementación de algoritmos automatizados de transformación de *ELE* a *EFD* sin pérdida de semántica.

Para la transformación de los diferentes tipos de entidades es necesario realizar al menos dos acciones: creación del tipo de componente entidad y transformación de su lista de conceptos. En un proceso automatizado los pasos a seguir deben guardar el siguiente orden de Transformación de Tipos de Entidades: Paso 1 Regulares (R), Paso 2 Débiles (D), Paso 3 Supertipos (SP), Paso 4 Subtipos (SB)

2.2.2. Reglas de Transformación para Relaciones (*DERExt* – ELE)

Las reglas destinadas a definir los *esq(C)* correspondientes a tipos de relaciones deben permitir representar directa o indirectamente los siguientes conceptos: nombre de relación; identificador principal e identificador o identificadores alternativos, si los hubiese; lista de los atributos propios del tipo de relación, con sus diferentes características y multiplicidades; el orden: unaria, binaria, ternaria o de orden superior; y sus correspondientes cardinalidades mínimas y máximas respecto de las entidades involucradas. Bajo el supuesto de una interpretación semántica de cardinalidades siguiendo el estilo Chen (look across) [5], [7], [8].

Para resolver el problema que implica no poder conservar todos los aspectos definitorios de ciertos tipos de relaciones, se propone obtener un *esq(C)* para cada uno de ellos, prefijándolo adecuadamente de acuerdo a la cardinalidad máxima de las

entidades que intervienen. De esta manera, para las relaciones unarias o binarias con cardinalidades N:1, N:N o 1:1, el prefijo será NN, N1, o 11 respectivamente, siguiendo en parte la idea presentada en [20].

Las relaciones ternarias enlazan 3 componentes entidad, entonces siguiendo el mismo espíritu que para las relaciones unarias o binarias, se sugiere nombrarlas por la tripla de cardinalidades: NNN, NN1, N11, 111. La cardinalidad máxima en el prefijo será siempre N, aunque en esta propuesta se admite que además de indefinida (N) pueda ser conocida y especificada mediante un entero mayor que 1, semántica que se pierde cuando se realizan las transformaciones siguiendo los algoritmos clásicos.

El *DERExt* original para las cardinalidades mínimas acepta valores mínimos 0 o 1, proporcionando una visión restringida de las diferentes situaciones. Se propone incorporar una versión extendida que permite generalizar todos los casos de valores mínimos y máximos de cardinalidades, considerando $0 \leq$ mín $\leq$ máx<, donde mín podrá tomar los valores 0, 1 ó *a* y máx podrá tomar los valores N ó *b* (donde N es el indefinido estándar, y *a* y *b* representan una multiplicidad constante respectivamente, con *b* >= 1). Esta información podría ser fácilmente extendida y completada indicando valores de intervalos para las cardinalidades mínimas y máximas (no se considera en este trabajo).

Se utiliza la notación de puntos (<Componente>.<Atributo>) para indicar el componente del cual proviene el atributo, conservando así el origen de un atributo respecto de la entidad de la cual proviene.

Al aplicar las reglas para transformar *DERExt* en un conjunto de *esq(C)* se debe establecer un orden. Se incluyen en primer lugar las claves primarias de las entidades participantes debidamente prefijadas con el nombre de la entidad u objeto del cual proceden, separadas por ‘,’. Del segundo al quinto lugar se colocan los identificadores alternativos, atributos descriptores simples, compuestos y multivaluados respectivamente. El sexto lugar se reserva para representar los pares de cardinalidades (min, max) correspondientes a cada entidad relacionada, colocando ambos límites entre ‘( )’ y asociados posicionalmente a cada clave proveniente de las entidades relacionadas, las cuales se enumeran en el séptimo y último lugar (reservado para representar las claves extranjeras). Las transformaciones de los componentes del 1 al 5, siguen las reglas que se detallarán en la siguiente sección.

2.2.3 Reglas de Transformación para Atributos (*DERExt* – ELE)

En la aplicación de las reglas destinadas a definir los *esq(C)* correspondientes a atributos tanto en entidades como en relaciones, los atributos simples se convierten en atributos simples. Los atributos compuestos se convierten en una lista encerrada por ‘( )’ de sus partes componentes, separadas por ‘,’ y precedida por el nombre del compuesto. Los atributos multivaluados se encierran entre ’{ }’. Adicionalmente, un atributo multivaluado compuesto se debe transformar considerando primeramente su composición y luego su carácter multivaluado. Un atributo compuesto con partes multivaluadas resultaría de transformar primero las partes del compuesto incluidos los multivaluados, y luego su transformación como compuesto. Cuando un atributo puede ser nulo dada la opcionalidad indicada en el diagrama, se lo deriva siguiendo las reglas anteriores, y se precede por ‘¬’.

2.3 Caso de Ejemplo

Para ejemplificar parte de lo descripto se utiliza el diagrama de la Fig. 1, que muestra la relación (0,1) : (0,N) *Pertenece* entre las entidades *EMPLEADO* y *DEPARTAMENTO* con dos atributos en la relación, que indican a partir de qué fecha inicia su actividad en el departamento, si lo tiene asignado, y una posible fecha de finalización si ha dejado de pertenecer al mismo. La semántica de ambos atributos está ligada a la cardinalidad de la relación (0,1), que como vimos se pierde con los algoritmos de transformación estándar.

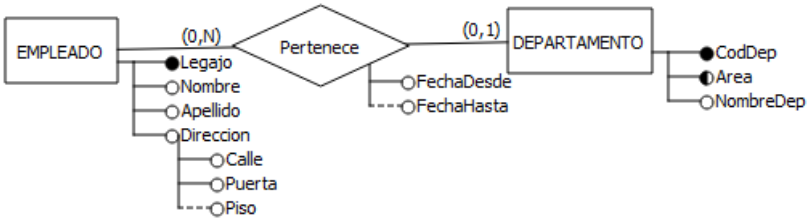


Fig. 1. Ejemplo de una situación muy simple y real, relativa a empleados y departamentos

2.3.1. Transformación *DERExt*-ELE según Proceso Estándar

Las reglas de transformación de *DERExt* a ELE, presentadas en los textos de la temática, se obtienen las relaciones donde se indican con subrayado simple las claves, subrayado doble las claves alternativas, su correspondiente Restricción de Integridad Referencial (RIR) indicada formalmente y restricciones de no nulidad (NN) (atributos que no pueden ser nulos) [4], [11], [15], [21], [22], [23], [24].

EMPLEADO (Legajo, Nombre, Apellido, Calle, Puerta, Piso, CodDep,
FechaDesde, FechaHasta)
DEPARTAMENTO(CodDep, Area, NombreDep)
RIR: EMPLEADO(CodDep) << DEPARTAMENTO(CodDep)
NN: EMPLEADO(Legajo, Nombre, Apellido, Calle, Puerta)
DEPARTAMENTO(CodDep, Area, NombreDep)

En esta transformación se perdieron dos conceptos semánticos importantes, la relación que existe entre los atributos de la relación EMPLEADO FechaDesde y FechaHasta con la cardinalidad de la relación Pertenece (0,1) y el descriptor de atributo compuesto Dirección.

2.3.2 Transformación *DERExt*-ELE segun propuesta

R_EMPLEADO[Legajo;;Nombre,Apellido, Direccion(Calle,Puerta,¬Piso);] (1)

R_DEPARTAMENTO[CodDep;AreaNombreDep;;] (2)

N1_Pertenece[EMPLEADO.Legajo;;FechaDesde,¬FechaHasta;;;(0,N),(0,1);
EMPLEADO.Legajo, DEPARTAMENTO.¬CodDep] (3)

En (1) y (2) se transformaron las entidades regulares (R_) Empleado y Departamento como dos componentes con la lista de atributos separados por ";" en el siguiente orden: 1 Concepto clave primaria (simple o compuesto). 2 Concepto clave alternativa (simple o compuesto). 3 Concepto descriptor simple. 4 Concepto descriptor compuesto. 5 Concepto descriptor multivaluado.

En (3) se transformó la relación Pertenece y los conceptos tienen el mismo orden que para las entidades (de 1 a 5). El sexto concepto, como ya se mencionó, representa las cardinalidades de las entidades involucradas, cuya ubicación se corresponde a la de las claves extranjeras de las entidades involucradas, para el ejemplo (0,N),(0,1). El séptimo y último concepto es el que corresponde a las claves extranjeras, en el ejemplo EMPLEADO.Legajo y DEPARTAMENTO.¬CodDep que es a su vez un atributo simple opcional.

Cabe destacar que en este caso, de una relación 1:N, y a diferencia del proceso estándar, se genera un componente para dicha relación.

2.3 Transformación del *ELE* al ELET

Una vez generado el *ELE* el proceso continúa con la siguiente etapa de transformaciones en las que se completan las decisiones cercanas a la implementación en SQL tanto en componentes de tipo entidad, relación o atributo.

Estas decisiones serán reflejadas en el ELET, que o bien puede continuar con una forma sintáctica de especificación similar a la del ELE, o ser implementado en un lenguaje intermedio de mayor nivel tal como XML o JSON.

Las especificaciones de esta fase de transformación se encuentran actualmente en desarrollo, por lo que a continuación se describirán los procesos y consideraciones a tener en cuenta en esta etapa.

2.3.1. Reglas de Transformación para Componentes Entidades (ELE al ELET)

En los componentes de tipo entidad existen opciones de transformación para subtipo (SB), supertipo (SP) y supertipo de una jerarquía de más de un nivel (SPB). Es posible tener en cuenta cuatro opciones de implementación distintas:

Opción A: crear una tabla para el supertipo (con clave primaria, y atributos comunes a todos los subtipos) y otra tabla para cada subtipo (con clave igual a la del supertipo, y atributos específicos del subtipo).

Opción B: crear una tabla por cada subtipo, incluyendo en cada una, los atributos del supertipo, y adicionando los atributos específicos del subtipo.

Opción C: crear una única tabla que contenga la clave primaria y atributos del supertipo, y los atributos de cada subtipo (completando con nulos, los atributos faltantes), a su vez, indicando el o los discriminante/s de subtipo al que pertenece/n el ejemplar, según corresponda.

Opción D: crear una tabla correspondiente al supertipo, y otra tabla que contenga los atributos de cada uno de los subtipos, especificando el tipo al que pertenece, según corresponda.

La elección de qué opción tomar, corre por cuenta de quien lleve adelante el diseño físico de ese modelo de datos, quien deberá considerar los pro y contras de cada opción aplicada al dominio en cuestión.

2.3.2. Reglas de Transformación para Componentes Relaciones (ELE al ELET)

Para los componentes de tipo relación es necesario adicionar las acciones referenciales y las opciones de matching para las claves extranjeras.

En el caso de las acciones referenciales permiten definir la acción a realizar ante una modificación o borrado sobre las claves en la tabla referenciada y según el SQL estándar es posible optar para cada acción por No Action, Restrict, Cascade, Set Null o Set Default. Por otro lado, se tienen los tipos de matching que se pueden definir sobre una clave extranjera (compuesta por más de un atributo) que admite nulos y que establece en qué modo se cumple la integridad referencial, las opciones posibles son Simple, Partial o Full [12].

Otro concepto que resulta importante es el de rolename. Este es el nombre que se utiliza ante la repetición de nombres de columnas a la hora de realizar la transformación. Esta repetición no es aceptada por SQL, por lo que se requiere un nombre que la reemplace, en casos como en relaciones unarias donde se incluiría dos veces la clave primaria y con entidades que poseen múltiples relaciones N:1 (recordar que la entidad de cardinalidad N tiene tantos atributos claves extranjeras como relaciones).

2.3.3. Reglas de Transformación para Componentes Atributos (ELE al ELET)

Dentro de los atributos, encontramos diferentes formas de transformarlos, según el rol que ocupan dentro del componente entidad o relación (claves principales, claves secundarias, claves extranjeras o atributos normales). Las siguientes transformaciones son válidas para cada tipo sin importar su rol, dado que el mismo es otorgado mediante sentencias especiales definidas por SQL.

Los atributos simples son agregados a la definición de la tabla a la que pertenecen, adicionando sólo el tipo de dato en el ELET. Los atributos multivaluados pueden ser transformados de acuerdo a 3 formas:

Opción A: agregar una cierta cantidad (especificada por el usuario) de atributos simples a la declaración de la tabla.

Opción B: como un arreglo, donde en la declaración simplemente se especifica como si fuese un atributo simple, pero al tipo de dato se le agrega la indicación de que se trata de un arreglo de elementos de ese tipo.

Opción C: creación de una tabla, compuesta simplemente por el atributo multivaluado (esta vez especificado como un atributo simple), y por la clave principal de la tabla que especifica la entidad que tiene ese atributo multivaluado. De esta forma, en esta nueva tabla, tanto la clave extranjera (referenciando a la clave primaria de la tabla principal) como el atributo multivaluado, forman parte de la clave primaria.

Por último, para transformación de atributos compuestos, existen dos formas diferentes de obtener la transformación a SQL. La primera forma consiste en definir cada atributo que compone al atributo compuesto como un atributo simple, en la tabla principal. No existe diferencia alguna con la derivación de un atributo simple. La segunda forma es crear un nuevo tipo de datos. Este tipo de datos va a contener todos los campos que tenga el atributo compuesto, y en la tabla principal se creará como un atributo cuyo tipo de dato es el creado previamente.

2.4 Transformación del *ELET* al EF/ SQL

Este proceso parte de la especificación del *ELET* y genera el esquema físico SQL final. Para llevar a cabo este proceso se requiere establecer un conjunto de reglas de transformación, las cuales se encuentran actualmente en definición. Para el ejemplo de la sección 2.3.1 y 2.3.2 el SQL generado se detalla a continuación.

```
CREATE TYPE Direccion_type AS(Calle varchar(40) NOT NULL, Puerta integer NOT
NULL, Piso integer);

CREATE TABLE EMPLEADO ( Legajo integer NOT NULL, Nombre varchar(40) NOT
NULL, Apellido varchar(40) NOT NULL, Direccion Direccion_type, FechaDesde date,
FechaHasta date, CodDep integer,   CONSTRAINT pk_empleado PRIMARY KEY
(Legajo),  CONSTRAINT ck_pertenece (check (CodDep IS NULL AND FechaDesde IS
NULL AND FechaHasta IS NULL) OR (CodDep IS NOT NULL AND FechaDesde IS
NOT NULL AND FechaHasta IS NOT NULL) OR (CodDep IS NOT NULL AND
FechaDesde IS NOT NULL));

CREATE TABLE DEPARTAMENTO (CodDep integer NOT NULL, Area char(5)
UNIQUE, NombreDep varchar(40) NOT NULL, CONSTRAINT pk_departamento
PRIMARY KEY (CodDep));

ALTER TABLE EMPLEADO ADD CONSTRAINT
FK_EMPLEADO_DEPARTAMENTO FOREIGN KEY (CodDep)    REFERENCES
DEPARTAMENTO (CodDep);
```

3 Conclusiones y Futuros Trabajos

En este trabajo se ha especificado un *ELE* incluyendo todos los aspectos del modelo conceptual de datos, proponiendo sólo cambios menores en la notación de las reglas conocidas para la construcción del diagrama, extendiendo el algoritmo de transformación para resolver cuestiones relativas a ambigüedades y falta de ortogonalidad.

Se han incorporado cualidades de completitud que se pretenden de la metodología, para posibilitar que el usuario experto pueda optar por una transformación tanto hacia un modelo relacional puro como a uno post-relacional. Se han descripto los lineamientos que deben seguir las reglas de especificación que establezcan la correspondencia de un *ELE* con el ELET, y de éste con el EF/SQL.

Este trabajo es un avance respecto al objetivo más global de obtener en un proceso automatizado la transformación desde un modelo conceptual hasta EF/SQL, el cual incremente la cantidad de cuestiones semánticas que contempla disminuyendo a un mínimo las decisiones de implementación que quedarán bajo responsabilidad de quien la lleve adelante en la base de datos.

Debido a esta incorporación de mayor semántica, los artefactos intermedios de las fases de este proceso constituyen también una mejora sustancial en la documentación del modelo de datos, y en la comunicación de desarrollo y post desarrollo de un proyecto de implementación.

Referencias

1. An, Y., Hu, X., & Song, I. Y.: Maintaining mappings between conceptual models and relational schemas. Journal of Database Management (JDM), 21(3), 36--68, (2010).
2. Badia, A.: Entity-Relationship modeling revisited. ACM SIGMOD Record, 33(1), 77--82, (2004).
3. Badia, A., & Lemire, D.: A call to arms: revisiting database design. ACM SIGMOD Record, 40(3), 61--69, (2011).
4. Batini, C., Ceri, S., & Navathe, S. B.: Conceptual database design: an Entity-relationship approach (Vol. 116). Redwood City, CA: Benjamin/Cummings, (1992).
5. Chen, P. P. S.: The entity-relationship model—toward a unified view of data. ACM Transactions on Database Systems (TODS), 1(1), 9--36, (1976).
6. Codd, E. F.: The relational model for database management: version 2. Addison-Wesley Longman Publishing Co., Inc., (1990).
7. Cuadra, D., Nieto, C., Martínez, P., Castro, E., & Velasco, M.: Preserving relationship cardinality constraints in relational schemata. In Database Integrity: Challenges and Solutions (pp. 66-112). IGI Global, (2002).
8. Cuadra, D., Martínez, P., Castro, E. Al-Jumaily, H.: Guidelines for representing complex cardinality constraints in binary and ternary relationships, Software & Systems Modeling Journal, Vol.12, 4, 871-889, (2013).
9. de Miguel Castaño, A., Velthuis, M. G. P., & Martínez, E. M.: Diseño de bases de datos relacionales. Editorial Alfaomega, (2000).
10. Date, C. J.: An introduction to database systems. Pearson Education India, (2006).
11. Elmasri, R., & Navathe, S.: Fundamentals of Database Systems, 6th Ed. Adison-Wesley Pub. Comp., (2010).
12. Date, C. J., & Darwen, H.: A guide to the SQL Standard: a user's guide to the standard relational language SQL. Addison-Wesley. (1989).
13. Goelman, D., & Song, I. Y.: Entity-relationship modeling re-revisited. In International Conference on Conceptual Modeling (pp. 43-54). Springer, Berlin, Heidelberg, (2004).
14. Halpin, T., & Morgan, T.: Information modeling and relational databases. Morgan Kaufmann, (2010).
15. Hansen, G. W., & Hansen, J. V.: Database management and design. Upper Saddle River, New Jersey: Prentice Hall, (1996).
16. Parent, C., Spaccapietra, S., & Zimányi, E.:. Spatio-temporal conceptual models: data structures+ space+ time. In Proceedings of the 7th ACM international symposium on Advances in geographic information systems (pp. 26-33). ACM, (1999).
17. Pieris, D.: Modifying the Entity relationship modelling notation: towards high quality relational databases from better notated ER models. Preprint arXiv:1306.5690, (2013).
18. Pieris, D.: A novel ER model to relational model transformation algorithm for semantically clear high quality database design. arXiv preprint arXiv:1306.6734, (2013).
19. Pieris, D.: Extending the ER Model to relational Model novel transformation Algorithm: transforming relationship Types among Subtypes. arXiv preprint arXiv:1307.4519, (2013).
20. Pieris, D., & Rajapakse, J.: Logical database design with ontologically clear entity relationship models. In Information and Automation for Sustainability (ICIAfS), 2012 IEEE 6th International Conference on (pp. 209-214). IEEE, (2012)
21. Ramakrishnan, R., & Gehrke, J.: Database management systems. McGraw Hill, (2000).
22. Silberschatz, A., Korth, H. F., & Sudarshan, S.: Database system concepts, 5th Ed., New York: McGraw-Hill, (2005)
23. Teorey, T.J.: Database Modeling and Design. The Entity-Relationship Approach", Morgan Kaufmann, San Mateo, (CA), (1990).
24. Teorey, T. J., Lightstone, S. S., Nadeau, T., & Jagadish, H. V.: Database modeling and design: logical design. Elsevier, (2011).

Workshop Ingeniería de Software (WIS)

Coordinadores

Patricia Pesado (UNLP) - Elsa Estevez (UNS) - Alejandra Cechich (UNCOMA) - Horacio Kuna (UNaM)

Temas

Administración de Configuración de Software y Despliegue
Administración de Proyectos de Software
Arquitecturas y Diseño de Software
Calidad de Software
Dependabilidad y Confiabilidad del Software
Desarrollo Ágil de Software
Desarrollo de Software basado en Componentes
Entornos para el Desarrollo de Software
Evolución del Software
Estándares para el Desarrollo de Software
Herramientas de Software
Herramientas y Tecnologías de la Web 2.0
Ingeniería de Requerimientos de Software
Ingeniería de Software Distribuido
Ingeniería de Software Embebido y Tiempo Real
Ingeniería de Software Empírica
Ingeniería de Software para Sistemas Ubícuos y Móviles
Ingeniería de Software para la Web
Ingeniería Reversa
Mantenimiento de Software
Metodologías de Desarrollo de Software
Métricas de Software
Modelamiento de Software
Modelamiento de Procesos de Negocio
Orientación a Aspectos
Patrones y Frameworks
Plataformas como Servicios
Políticas para el Software y para las TICs
Privacidad
Procesos de Software
Reuso de Software
Sistemas de Información
Sistemas de Software basados en Cloud Computing
Sistemas de Software Sensibles al Contexto
Software como Servicio
Soporte Computacional para Tareas Colaborativas
Testeo de Software
Teoría y Métodos Formales
Verificación y Validación de Software
Workflows

White-Box Testing Framework for Object-Oriented Programming based on Message Sequence Specification

Juan Ignacio Rodríguez Silva[1], Martín Larrea[1,2,3]
[1]Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (DCIC-UNS)
[2]Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET)
[3]Laboratorio de I+D en Visualización y Computación Gráfica, (UNS-CIC Prov. de Buenos Aires)
nachorodriguez12@hotmail.com, mll@cs.uns.edu.ar

Abstract. Software is a crucial element in the functionality of devices and industry. Likewise, the operation of an enterprise or organization depends largely on the reliability of the software systems used for supporting the business process or particular tasks. The quality of software has become the most important factor in determining the success of products or enterprises. In order to accomplish a quality software product several methodologies, techniques, and frameworks have been developed, each of them tailored to specific areas or characteristics of the software under review. This paper presents a white-box testing framework for Object-Oriented Programming based on Message Sequence Specification. In the context of an object-oriented program, our framework can be used to test the correct order in which the methods of a class are invoked by its clients. The implementation of the framework is based on aspect-oriented programming.

Keywords: Software Verification Validation, White-Box Testing, Object-Oriented Programming, Message Sequence Specification, Aspect-Oriented Programming

1 Introduction

Verification and Validation (V&V) is the process of checking that a software system meets its specifications and fulfills its intended purpose. The software engineering community has acknowledged the importance of V&V process to ensure the quality of its software products. The V&V process, also known as software testing or just testing, is composed of V&V techniques. There are many different V&V techniques which are applicable at different stages of the development lifecycle. The two main categories of testing techniques are white-box and black-box. In the first one, the testing is driven by the knowledge and information provided by the implementation or source code. While in the second one the specification of the software, module, or function is used to test the object under review.

In 1994, Kirani and Tsai ([4]) presented a technique called Message Sequence Specification that, in the context of an object-oriented program, describes the correct order in which the methods of a class should be invoked by its clients. The method sequence specification associated with an object specifies all sequences of messages that the object can receive while still providing correct behavior. Daniels and Tsai [5] used the idea of Message Sequence Specification as a testing tool but without implementing a framework to support this technique.

We developed a framework for testing object-oriented programs based on Message Sequence Specification. Our framework can be used to test the correct order in which the methods of a class are invoked by its clients. The implementation of the framework was done using aspect-oriented programming, which allows to test the source code without the need to change any line of code in the program itself and even allows to run the tests while the user is using it.

In the remaining parts of this article, we first provide background V&V in general and them move on to review the concepts of Message Sequence Specification in the software development process. We continue with the presentation of aspect programming, a key part of our framework. Then, we describe our proposed framework. We later introduce an example of the use of such framework. We conclude with a brief discussion on limitations and advantages of our approach and the future work.

2 Background Review

2.1 Background on Verification & Validation

Software testing is involved in each stage of software life cycle, but how we test and what we test on each stage of software development is different, the nature and the goals of what is being tested are different. Based on [1], there are 8 types of testing in the life cycle: Unit testing is a code based testing which is performed by developers, this testing is mainly done to test each and individual units separately. The Unit testing can be done for small units of code, generally no larger than a class. Integration testing validates that two or more units work together properly, and inclines to focus on the interfaces specified in the low-level design. System testing reveals that the system works end-to-end in a production-like location to provide the business functions specified in the high-level design. Acceptance testing is conducted by business owners; the purpose of acceptance testing is to test whether the system does, in fact, meet their business requirements. Regression Testing is the testing of software after changes have been made to ensure that those changes did not introduce any new errors into the system. Alpha Testing is usually applied at the developer's site, with the presence of the developer. Beta Testing is done at the customer's site with no developer in site. Functional Testing is done for a finished application, its goal is to verify that the system provides all the required behavior.

In the context of V&V, black-box testing is often used for validation (i.e. are we building the right software?) and white-box testing is often used for

verification (i.e. are we building the software right?). In black-box testing, the test cases are based on the information from the specification. The software testers do not consider the internal source code of the test object. The focus of these tests is solely on the outputs generated in response to selected inputs and execution conditions. The software tester sees the software as a black box, where information is input to the box, and the box sends something back out. This can be done purely based on the requirement specification knowledge; the tester knows what to expect from the black box to send out and tests to make sure that the box sends out what it's supposed to send out.

Oppositely, white-box testing, also called structural testing, designs test cases based on the information derived from the source code. White-box testing is concern with the internal mechanism of a system, it mainly focuses on the control flow or data flow of the program. White-box and black-box testing are considered to complement each other. In order to test software correctly, it is essential to generate test cases from the specification and source code. This means that we must use white-box and black-box techniques on the software under development.

Each test technique, both white-box and black-box, must describe a test model and, at least, one coverage criteria. Test models describe how to generate test cases, it can be a graph, a table or a set of numbers. Coverage criteria are used to steer and stop the test generation process ([2]), they are usually boolean conditions. They have widely accepted means of assessing the quality of a test ([3]).

The same testing technique that we classified as white or black box, can be arranged as static or dynamic techniques. Static testing are those techniques where the code is not executed, it can be analyzed manually or by a set of tools. This type of testing checks the code, requirement documents, and design documents. Dynamic testing is done when the code is executed. Dynamic testing is performed when the code being executed is input with a value, the result or the output of the code is checked and compared with the expected output.

2.2 Background on Message Sequence Specification

In 1994, Kirani and Tsai ([4]) presented a technique called Message Sequence Specification that, in the context of an object-oriented program, describes the correct order in which the methods of a class should be invoked by its clients. The method sequence specification associated with an object specifies all sequences of messages that the object can receive while still providing correct behavior.

Their strategy used regular expressions to model the constraints over the correct order of the invocation of the methods i.e. the regular expression is the test model. Method names were used as the alphabet of a grammar which was then used to statically verify the program's implementation for improper method sequences. A runtime verification system identifies incorrect method invocations by checking for sequence consistency with respect to the sequencing constraints.

If a class C has a method M_1, this is noted as C_{M_1}. Sequence relationships between two methods were classified into three categories, sequential, optional,

and repeated. If the method M_1 of C should be invoked before the method M_2 of the same class, then this relationship is sequential and is represented as

$$C_{M_1} \bullet C_{M_2}$$

If one, and only one of the methods M_1 and M_2 can be invoked then this relationship is optional and is represented as

$$C_{M_1} | C_{M_2}$$

Finally, if the method M_1 can be invoked many times in a row then this is a repeated relationship and is represented as

$$(C_{M_1})^*$$

For example, if a class X has three methods *create*, *process* and *close*, a possible sequencing constraint based on Message Sequence Specification could look like

$$X_{create} \bullet (X_{process})^* \bullet X_{close}$$

If class X is part of a larger system S, then we could statically check the source code of S to see if all calls to X's methods follow the defined grammar. If a static analysis is not enough, we could implement a runtime verification system that tracks all calls to X's methods and checks dynamically the sequence of calls against the grammar.

This technique can also be used to test the robustness of a system. Continuing with the class X as an example, we can use the defined grammar to create method sequences that are not a derivation from the grammar, i.e. incorrect sequences methods. These new sequences can be used to test how the class handles a misuse. For example, how does the class X respond to the following sequence of calls:

$$X_{create} \bullet X_{close} \bullet X_{process}$$

Testing with the sequences generated by the grammar and with those that were not are two variations presented in [5], a work that extended the research done in [4].

2.3 Background in Aspect Oriented Programming

Aspect Oriented Programming [6] (AOP) is a programming paradigm that aims to increase modularity by allowing the separation of cross-cutting concerns. It does so by adding additional behavior to existing code without modifying the code itself, instead separately specifying which code is modified via a "pointcut" specification, such as "log all function calls when the function's name begins with *set*". This allows behaviors that are not central to the business logic (such as logging or testing) to be added to a program without cluttering the code, core to the functionality.

AOP entails breaking down program logic into distinct parts (so-called concerns). Nearly all programming paradigms support some level of grouping and encapsulation of concerns into separate, independent entities by providing abstractions (e.g., functions, procedures, modules, classes, methods) that can be used for implementing, abstracting and composing these concerns. Some concerns cut across multiple abstractions in a program, and defy these forms of implementation. These concerns are called cross-cutting concerns or horizontal

concerns. Logging exemplifies a crosscutting concern because a logging strategy necessarily affects every logged part of the system. Logging thereby crosscuts all logged classes and methods.

See for example Listing 1.1, a simple Java class for a bank account. If we need to log all the events in the account one way to do it is as the listing shows. The main disadvantages of this approach is that we are mixing the logic of the bank account class with the requirement of logging its events. By using AspectJ we can create an aspect, as in Listing 1.3 while the bank account class remains simpler 1.2. With this two classes, every time there is a call to *deposit* or *withdraw* the jre will execute the methods in the AspectJ aspect.

Listing 1.1. Classic example

```
public class Account {
protected int amount;

public Account()
{ this.amount = 0; }

public void deposit( int _amount
{ this.amount += _amount;
 Log.put("deposit for" + _amount);
 }

public void whitdraw( int _amount )
{ this.amount -= _amount;
 Log.put("whitdraw for" + _amount);
 } }
```

Listing 1.2. Clear code

```
public class Account {
protected int amount;

public Account()
{ this.amount = 0; }

public void deposit( int _amount
{ this.amount += _amount;
 }

public void whitdraw( int _amount )
{ this.amount -= _amount;
 } }
```

Listing 1.3. AspectJ code

```
public aspect AspectLogic {

before(int _amount):
call(void Account.deposit(int))
 && args(_amount) {
Log.put("deposit for " + _amount);
}

before(int _amount):
call(void Account.whitdraw(int))
 && args(_amount) {
Log.put("whitdraw for " + _amount);
 } }
```

By using AspectJ we can add checks to a source code with out the need to change the code itself. This is a very appealing feature in the context of software verification and validation. For more on AOP we recommend the book [6].

3 Previous Work on Testing with Message Sequence Specification

The technique presented in 1994 by Kirani and Tsai ([4]) was follow in 1999 by Daniels and Tsai in [5]. Also in 1999, Tsai et al. presented [8] where Message Sequence Specification was used to create template scenarios than later were used to create test cases. In 2003, Tsai used the technique [7] as a verification mechanism to the UDDI servers in the context of Web Services. In 2014, [9] introduced a Java-based tool for monitoring sequences of method calls, similar to our goal in this work they used annotations instead of AOP.

4 Our Proposal

Our goal in this work is to present a testing framework for object-oriented source code based on Message Sequence Specification by using AOP. AOP will allow us to create our test cases without the need to modify the source code, and those test cases will run automatically with each run of the program under test. The use of message sequence specification will allow the developer of each class to describe a grammar that will represent the correct behavior of such class. The framework will take these grammars, run the program and check that the methods are used according to the developer specification. We wanted to provide an easy to use framework, with an easy to read and understand representation of the methods correct usage. More particularly, we wanted a framework that the developer can use, without the need for a testing specialist.

The first thing the developer must do to use the framework is to create the regular expression associated with its class. This is the correct behavior or order in which the methods of the class should be called. In order to express this

grammar in a simple way, the developer is encouraged to use simple symbols that represent the methods. This means that it is not required to use the actual names of the methods in the grammar. But, in order to keep the grammar easy to read, at some point the developer must create a map between actual methods name and their short version.

Lets use Listing 1.4 as an example. Again we have a Bank Account class, but a bit more complex. In this case, the correct order to use the Account is: first the account must be created and then it must be verified. The first money movement in the account must be a deposit, after that we can deposit or withdraw money. Finally, the account must be closed.

Listing 1.4. Classic example

```
public class Account {
protected int amount;
protected boolean verify;

public Account()
{ this.amount = 0;
this.verify = false; }

public void verify()
{ this.verify = true; }

public void deposit( int _amount )
{ if (this.isVerify()) this.amount += _amount; }

public void whitdraw( int _amount )
{ if (this.isVerify()) this.amount -= _amount; }

public void close()
{ this.amount = 0;
this.verify = false; }

public boolean isVerify()
{ return this.verify; } }
```

Based on Message Sequence Specification, the grammar for the correct use of this class is as follows:

$$create \bullet verify \bullet deposit \bullet (deposit|withdraw)^* \bullet close$$

or as a more simpler expression, as:

$$c \bullet v \bullet d \bullet (d|w)^* \bullet x$$

As shown the names that we used in the grammar not necessarily have to be the actual methods name. This can be used to enhance the readability of the grammar, but in order to use this as it is, we need to map those name to the actual methods. For this goal, the framework offers a $Map < String, String >$

map; variable where the mapping is store. Before we can tell the framework about our grammar for the Bank Account class, we must first input the mapping as shown in the Listing 1.5. We then can introduce the grammar to the framework (Listing 1.6). With this two steps completed, it is now possible to execute the main program and let the framework check, at runtime, in which order are the methods of the Bank Account class being called.

Listing 1.5. Classic example

```
map.put("bank.Account.<init>", "c");
map.put("bank.Account.verify", "v");
map.put("bank.Account.deposit", "d");
map.put("bank.Account.withdraw", "w");
map.put("bank.Account.close", "x");
```

Listing 1.6. Classic example

```
Pattern regex = Pattern.compile("cvd(d|w)*x");
Matcher matcher = regex.matcher("");
```

Once the main program is running, our framework will keep a log of how the methods in the grammar are called. If a method call doesn't follow the grammar form, the framework will issue an alert in the console output. A trace of all the call will be output to the console as well.

The usefulness of this framework can be emphasized in the following section, where we use our development to find an error in a real case, an application developed in our laboratory.

5 Case Study. Rock.AR, a software solution for point counting

Point counting is the standard method to establish the modal proportion of minerals in coarse-grained igneous, metamorphic, and sedimentary rock samples. This requires observations to be made at regular positions on the sample, namely grid intersections. At each position, the domain expert decides to which mineral the respective grid point and its local neighborhood belongs. By counting the number of points found for each mineral, it is possible to calculate the percentages that these values represent of the total counted points. These percentages represent the approximate relative proportions of the minerals in a rock, which is a 2D section of a 3D sample. Rock.AR ([11]) is a visualization tool with a user-friendly interface that provides a semiautomatic point-counting method. It increases the efficiency of the point-counting task by reducing the user cognitive workload. This tool automates the creation of the grid used to define the point positions. This grid is overlaid on a predetermined image of a sample, allowing the count of minerals identified at the intersections of the grid lines. This method significantly reduces the time required to conduct point counting, it does not require an expensive ad hoc device to perform the job, and

it improves the consistency of counts. Among the methods available in the main class of this application, there are three of interest for us. Those are LoadSample(), MoveSelectedCell() and AddNewRockType(). In this program, the correct used of these methods is: First it must be at least one call to LoadSample(), then before a new rock type can be added, a point (also known as a cell) must be selected. We can create a grammar from this as such:

$$(LoadSample \bullet LoadSample^* \bullet (MoveSelectedCell \bullet MoveSelectedCell^* \bullet AddNewRockType^*)^*)^*$$

or in a more simpler way:

$$(l \bullet l^* \bullet (m \bullet m^* \bullet a^*)^*)^*$$

A particular observation of this grammar is that, as defined by the developer, before adding a new mineral type at least one cell must be selected, and before that a sample must be load. Using our framework, we created this grammar and we enter it into the framework. After this we ran the program several times and used Rock.AR for a while. In the third run of the program, the test framework detect an error and output the sequence of calls that didn't follow the grammar. The sequence was: :

$$l \bullet m \bullet m \bullet a \bullet l \bullet l \bullet a$$

The last three call on the sequence are the interesting ones. According to these, it was possible to call the method AddNewRockType() without having to call MoveSelectedCell() first. Using this information we discover that when the user load a second sample in the programm, there were two variables that were not initialized. Because of this, it was possible to call AddNewRockType() right after LoadSample() something that was not allow in the grammar. Before the use of our testing framework there was no evidence of the error found. The error does not generate any exception in execution, this is because since the variables were already initialized their old values were taken and the program continued its execution.

6 Conclusions & Future Work

In a world where technology is part of every person's daily life, software is a crucial element. The quality of software has become the most important factor and the developer need tools to assist them in their work in order to achieve such high quality. In this work, we presented a framework for white-box testing. Our framework combines Message Sequence Specification with Aspect Oriented Programming in order to achieve a tool to test the correct order in which methods in a class are being called. As we stated earlier, our goal was to create an easy to use framework for testing. A framework that the developer could use on its own code and a framework that will not interfere with the code under review. The intention of this framework is to find a particular type of error, which occurs when the order in which the methods of a class are called is relevant. As it was demonstrated in the case of study, the framework allows to detect errors that otherwise would be difficult to determine.

As for future work, a more expressive framework is under consideration. For the moment, we can only describe the order in which the methods are being called but not any information on how are they call or the state of the class instance. It would be useful to say that a method x must be called after method y if the value of an attribute is equal to 0.

Acknowledgment

This work was partially supported by the following research projects: PGI 24/N037 and PGI 24/ZN29 from the Secretaría General de Ciencia y Tecnología, Universidad Nacional del Sur, Argentina.

References

1. Jorgensen, Paul C. Software testing: a craftsman's approach. CRC Press, 2013.
2. Weileder, Stephan. Test models and coverage criteria for automatic model-based test generation with UML state machines. Diss. Humboldt University of Berlin, 2010.
3. Friske, Mario, Bernd-Holger Schlingloff, and Stephan Weileder. "Composition of Model-based Test Coverage Criteria." MBEES. 2008.
4. Kirani, Shekhar H., and W. T. Tsai. Specification and verification of object-oriented programs. Diss. University of Minnesota, 1994.
5. Daniels, F. J., and K-C. Tai. "Measuring the effectiveness of method test sequences derived from sequencing constraints." Technology of Object-Oriented Languages and Systems, 1999. TOOLS 30 Proceedings. IEEE, 1999.
6. Laddad, Ramnivas. AspectJ in action: practical aspect-oriented programming. Dreamtech Press, 2003.
7. Tsai, W. T., Paul, R., Cao, Z., Yu, L., Saimi, A. (2003, January). Verification of web services using an enhanced UDDI server. In Object-Oriented Real-Time Dependable Systems, 2003.(WORDS 2003). Proceedings of the Eighth International Workshop on (pp. 131-138). IEEE.
8. Tsai, W. T., Tu, Y., Shao, W., Ebner, E. (1999). Testing extensible design patterns in object-oriented frameworks through scenario templates. In Computer Software and Applications Conference, 1999. COMPSAC'99. Proceedings. The Twenty-Third Annual International (pp. 166-171). IEEE.
9. Nobakht, B., de Boer, F. S., Bonsangue, M. M., de Gouw, S., Jaghoori, M. M. (2014). Monitoring method call sequences using annotations. Science of Computer Programming, 94, 362-378.
10. Bai, X., Lu, H., Zhang, Y., Zhang, R., Hu, L., Ye, H. (2011, July). Interface-Based automated testing for open software architecture. In Computer Software and Applications Conference Workshops (COMPSACW), 2011 IEEE 35th Annual (pp. 149-154). IEEE.
11. Larrea, M. L., Castro, S. M., Bjerg, E. A. (2014). A software solution for point counting. Petrographic thin section analysis as a case study. Arabian Journal of Geosciences, 7(8), 2981-2989.

Método de inclusión de Hacking ético en el proceso de Testing de software

Ariel Giannone, Hernán Amatriain, Dario Rodriguez y Hernán Merlino.

Programa de Maestría en Sistemas de Información Universidad Tecnológica Nacional – Facultad Regional Buenos Aires. Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información
Grupo de Ingeniería de Explotación de Información y Grupo Investigación en Sistemas de Información
giannonearIel@gmail.com, hamatriain@gmail.com, dariorodriguez1977@gmail.com, hmerlino@gmail.com

Resumen. La infraestructura digital se hace cada vez más compleja e interconectada, la dificultad de lograr la seguridad en aplicaciones aumenta exponencialmente; y se plantea la necesidad de una actualización en las herramientas de testeo para la seguridad informática. Se propone un modelo de inclusión de hacking ético en el proceso de testeo de software.

Palabras clave. Seguridad Informática, Hacking ético, OWASP, Testeo de software, intrusión, pen testing.

1. Introducción

El crecimiento explosivo de Internet ha traído muchas cosas buenas tales como el comercio electrónico, facilidad en el acceso a grandes cantidades de almacenamiento de material de referencia, computación colaborativa, e-mail, nuevas vías para la publicidad, información distribuida, por nombrar unos pocos. A medida que la infraestructura digital se hace cada vez más compleja e interconectada, la dificultad de lograr la seguridad en aplicaciones aumenta exponencialmente. [OWASP, 2013] La información para la organización es un activo que debe ser protegido, la piratería es muy común en Internet y tiene afectado a la organización en términos de dinero, la pérdida de recursos y pérdida de imagen. [Sheoran & Singh , 2014]
Cuando el número de estas intrusiones informáticas destructivas se hicieron “famosas”, debido a la visibilidad del sistema o la magnitud del daño infligido, se convirtieron en noticias y los medios de comunicación recogieron estas historias. En lugar de usar el término preciso de "criminal informático", los medios de comunicación comenzaron a usar el término "Hacker" [Raymond, 1991] para referirse a las personas que irrumpen en los ordenadores para la diversión, la venganza, o su propia ganancia.
En su búsqueda de una manera de abordar el problema, las organizaciones informatizadas se dieron cuenta de que una de las mejores formas de evaluar la amenaza de intrusión a sus intereses sería tener profesionales independientes de

seguridad informática intentando entrar en sus sistemas. Este esquema es similar a tener auditores independientes entrando en una organización para verificar sus registros de contabilidad. En el caso de seguridad informática, estos "hackers éticos" emplean las mismas herramientas y técnicas que los intrusos, pero sin dañar el sistema de destino ni robar información.
Ahora bien, estas etapas deben realizarse en un marco de control, gestión y supervisión constante la cual otorgue tranquilidad y seguridad tanto al profesional que se "coloca" en los pies del criminal como a la organización en su totalidad. Es allí donde apunta este trabajo, poder incluir de manera segura y metódica la fase de revisión por hacking ético dentro del proceso de Testing de software.

2. Descripción del Problema

En general, las políticas de seguridad de la información o los controles por sí solos no garantizan la protección total de la información, ni de los sistemas de información, servicios o redes. Después de los controles que habitualmente se implementan, vulnerabilidades residuales probablemente permanezcan haciendo ineficaz la seguridad de la información y por lo tanto los incidentes son aun más probables. Una preparación insuficiente por una organización para hacer frente a este tipo de incidentes hará cualquier respuesta menos efectiva, y aumentar así el grado de impacto comercial potencial adverso. [ISO/IEC 27035:2011].Por otro lado para lograr éxito en las pruebas, este proceso debe ser planificado con antelación. Todos los aspectos técnicos, de gestión y estratégicos deben estar sumamente cuidados. La planificación es importante para todas las pruebas, ya sea desde un simple análisis de contraseña a una prueba de penetración completa en una aplicación web [Mayorga Jácome et al, 2015; Santos Castañeda, 2016; Onofa Calvopiña et al, 2016; López Vallejo, 2017].
Claro que el profesional de seguridad, al llevar a cabo un test de penetración como parte de su trabajo de hacking ético, necesita contar con ese tipo de lógica y tiene que aplicarla, más allá de utilizar las técnicas y herramientas open source [Comunidad Linux,2014; OISSG, 2012; GNU, 2014], comerciales o privadas [Tenable Network Security, 2014], dado que necesita imitar un ataque de la mejor manera y con el máximo nivel posible [Coronel Suarez, 2016; Hurtado Sandoval et al, 2016; López Alvarez, 2016; López Vallejo, 2017]. Para eso, tendrá que emplear todos los recursos que tenga a su alcance, utilizar al extremo sus conocimientos, poder de deducción y análisis mediante el razonamiento y así determinar qué es lo mejor que puede intentar, cómo, dónde y con qué. Por ejemplo, saber si un pequeño dato, por más chico o insignificante que parezca, le será útil y cómo proseguir gracias a él [Tori, 2008].
Tal como exponen diversos autores y organizaciones, se aconsejan herramientas, tareas y como planificar, pero no existe, hasta el momento, ninguna metodología o protocolo que establezca formalmente los pasos a seguir por un profesional dentro del proceso mismo de testing de software, que ayude a lograr un sistema con la menor cantidad de vulnerabilidades expuestas.

3. Solución Propuesta

Se propone desarrollar e incorporar un método de hacking ético para la evaluación de vulnerabilidades dentro del procedimiento mismo de Testeo de un sistema. Suministrando, de esta manera, a los encargados de testing en sectores de Seguridad Informática un grupo de actividades y herramientas que les brinde el soporte necesario para poder prevenir los problemas que en la actualidad son de creciente interés por las pérdidas económicas que conllevan.
Esta metodología se incluye en el proceso mismo de testeo de software, tal como detalla la figura 1, siendo aplicable a cualquier modelo de desarrollo de sistemas.

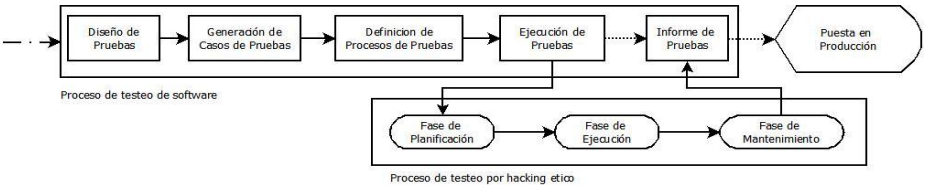


Figura 1. Inclusion del Proceso de Testeo por Hacking Ético

El proceso propuesto está compuesto por tres fases que se muestran en la Figura 2 y detallan en la Tabla 1.

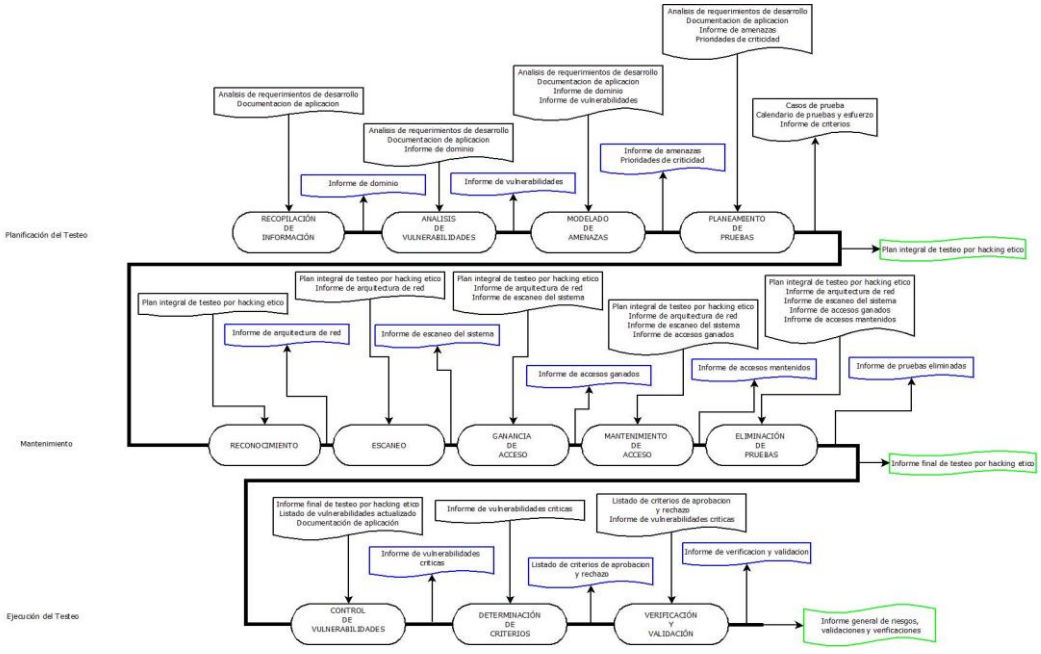


Figura 2. Proceso de Testeo por Hacking Ético

Proceso de testeo de hacking etico			
FASE I: PLANIFICACION DE TESTEO			
ETAPAS	**ACTIVIDADES**	**INSUMOS**	**PRODUCTOS**
RECOPILACION DE INFORMACION	1) Documentar tipo de aplicación 2) Establecer límites de componentes a someter en prueba 3) Delimitar alcance de la prueba 4) Identificar primeros riesgos asociados	• Análisis de requerimientos de desarrollo • Documentación de aplicación	• Informe de dominio
ANALISIS DE VULNERABILIDADES	5) Listar vulnerabilidades 6) Determinar probabilidad de vulnerabilidad 7) Realizar ranking de vulnerabilidades propicias	• Análisis de requerimientos de desarrollo • Documentación de aplicación • Informe de dominio	• Informe de vulnerabilidades
MODELADO DE AMENAZAS	8) Identificar información sensible en el sistema 9) Crear una descripción de la arquitectura 10) Descomponer la aplicación 11) Identificar las vulnerabilidades criticas 12) Documentar las vulnerabilidades criticas 13) Asignar prioridades a las vulnerabilidades criticas	• Análisis de requerimientos de desarrollo • Documentación de aplicación • Informe de dominio • Informe de vulnerabilidades	• Informe de amenazas • Prioridades de criticidad
PLANEAMIENTO DE PRUEBAS	14) Planificar pruebas 15) Armar casos de pruebas 16) Establecer criterios de aprobación y rechazo 17) Establecer tipo de acción (preventiva, correctiva, etc.) 18) Determinar grados de criticidad 19) Calendarizar hitos 20) Estimar esfuerzo 21) Elaborar plan	• Análisis de requerimientos de desarrollo • Documentación de aplicación • Informe de amenazas • Prioridades de criticidad	• Casos de prueba • Calendario de pruebas y esfuerzo (Gantt • Informe de criterios
PRODUCTO FINALDE LA FASE: PLAN INTEGRAL DE TESTEO POR HACKING ETICO			
FASE II: EJECUCION DEL TESTEO			
ETAPAS	**ACTIVIDADES**	**INSUMOS**	**PRODUCTOS**
RECONOCIMIENTO	22) Recopilar información inicial 23) Determinar el tamaño de la red 24) Identificar máquinas activas 25) Descubrir puertos abiertos y puntos de acceso 26) Identificar el sistema operativo 27) Mapear la red completa	• Plan integral de testeo por hacking ético	• Informe de arquitectura de red
ESCANEO	28) Detectar sistemas vivos en la red 29) Descubrir puertos activos 30) Descubrir el sistema operativo 31) Descubrir los servicios ejecutándose y presentes en el sistema 32) Descubrir direcciones IPs	• Plan integral de testeo por hacking ético • Informe de arquitectura de red	• Informe de escaneo del sistema
GANANCIA DE ACCESO	33) Recorrer direcciones 34) Investigar puertos 35) Explotar servicios y sistemas	• Plan integral de testeo por hacking ético • Informe de arquitectura de red • Informe de escaneo del sistema	• Informe de accesos ganados
MANTENIMIENTO DE ACCESO	36) Mantener el acceso 37) Asegurar el acceso exclusivo 38) Cargar, descargar y manipular datos, aplicaciones y configuraciones en el sistema 39) Usar el sistema comprometido para lanzar más ataques	• Plan integral de testeo por hacking ético • Informe de arquitectura de red • Informe de escaneo del sistema • Informe de accesos ganados	• Informe de accesos mantenidos
ELIMINACION DE PRUEBAS	40) Ocultar actos maliciosos. 41) Continuar el acceso al sistema de la víctima. 42) Sobrescribir el servidor, los sistemas y el registro de aplicaciones. 43) Elaborar Informe	• Plan integral de testeo por hacking ético • Informe de arquitectura de red • Informe de escaneo del sistema • Informe de accesos ganados • Informe de accesos mantenidos	• Informe de pruebas eliminadas
PRODUCTO FINALDE LA FASE: INFORME FINAL DE TESTEO POR HACKING ETICO			
FASE III: MANTENIMIENTO			
ETAPAS	**ACTIVIDADES**	**INSUMOS**	**PRODUCTOS**
CONTROL DE VULNERABILIDADES	44) Listar vulnerabilidades nuevas 45) Determinar grado de aplicabilidad	• Informe final de testeo por hacking ético • Listado de vulnerabilidades actualizado	• Informe de vulnerabilidades criticas

DETERMINACION DE CRITERIOS	46) Determinar grado de criticidad 47) Establecer criterios de aprobación y rechazo 48) Listar ranking de vulnerabilidades mas riesgosas	• Informe de vulnerabilidades criticas	• Listado de criterios de aprobación y rechazo
VERIFICACION Y VALIDACION	49) Ejecutar pruebas de verificación 50) Comparar con criterios establecidos 51) Determinar criticidad del riesgo 52) Establecer acciones a tomar 53) Validar resultados 54) Elaborar informe	• Listado de criterios de aprobación y rechazo • Informe de vulnerabilidades criticas	• Informe de verificación y validación
PRODUCTO FINALDE LA FASE: INFORME GENERAL DE RIESGOS, VALIDACIONES Y VERIFICACIONES			

Tabla 2. Proceso de Testeo por Hacking Ético

4. Descripción de Etapas y Actividades

Este proceso puede comenzar en paralelo a la fase de ejecución de prueba dentro del proceso mismo de testeo de software.
El modelo de proceso propuesto para el Testeo por Hacking Ético se encuentra dividido en tres fases principales: Fase de Planificación del testeo, Fase de Ejecución del testeo y Fase de Mantenimiento.

Fase de Planificación del Testeo

Tiene como objetivo interpretar la aplicación, identificar amenazas y vulnerabilidades, planificar y calendarizar las pruebas, entre otras acciones. Posee como entradas principales los Análisis de Requerimientos previos al desarrollo y la Documentación de la aplicación a someter al testeo, generando como producto final el Plan Integral de Testeo por Hacking Ético (PITHE). Esta fase se encuentra conformada por cuatro etapas que se explican a continuación.

Recopilación de Información: se encuentran las tareas correspondientes al entendimiento e interpretación de la aplicación que se someterá al testeo, por lo que se plantean las siguientes actividades: Documentación del tipo de aplicación, Establecimiento de límites de componentes a someter a pruebas, Delimitación del alcance de la prueba e Identificación de primeros riesgos asociados. Dando como producto final el Infrome de Dominio.

Análisis de vulnerabilidade: aparecen las actividades que ayudan a identificar las debilidades a las que la aplicación sería más propicia. Las tareas que la conforman son: Listado de vulnerabilidades, Determinación de probabilidad de vulnerabilidad y Realización de ranking de vulnerabilidades más propicias. Otrogando como producto final el Infrome de vulnerabilidades.

Modelado de Amenazas: donde se realiza una descomposición detallada y minuciosa de la aplicación, identificando y documentando la criticidad de las vulnerabilidades. Para llevar esto a cabo se plantean las siguientes actividades: Identificación de la información sensible dentro de la aplicación, Descripción de la arquitectura, Descomposición de la aplicación, Identificación de las vulnerabilidades críticas, Documentación de las vulnerabilidades y Asignación de prioridades de las vulnerabilidades críticas, donde los productos finales son el Informe de Amenazas y las Prioridades de Criticidad.

Planeamiento de Pruebas: la cual tiene como objetivo el ordenamiento, armado y planificación de las pruebas a realizar, esta es constituida por las actividades: Planificación de pruebas, Armado de casos de pruebas, Establecimiento de criterios de aprobación y rechazo y de tipo de acción, Determinación de grados de criticidad, Calendarización de hitos, Estimación de esfuerzos y por último, la Elaboración del plan integral de testeo por hacking ético. Dando como productos, los Casos de Prueba, Calendario de pruebas y esfuerzo y el Informe de criterios.
Habiendo cumplido con todas estas tareas y armado de productos, obtendremos como salida final el *Plan Integral de Testeo por Hacking Ético (PITHE).* Siendo precisamente el input para la fase siguiente.

Fase de Ejecución del Testeo

La cual tiene como objetivo principal la realización de las pruebas de forma ordenada, pautada y regida por las herramientas que se proponen para llevar adelante y arribar de la mejor forma al Informe Final del Testeo (IFT). Para lograr lo descripto anteriormente, esta fase es compuesta por cinco etapas complementarias.
Reconocimiento: se busca entender, descubrir y recopilar toda la información necesaria para llevar adelante las pruebas sobre la aplicación. Para lograr esto, se plantean estas actividades: Recopilación de información inicial, Determinación del tamaño de la red, Identificación de las máquinas activas, Descubrimiento de puertos abiertos y puntos de acceso, Rastreo del sistema operativo y Mapeo de la red completa. Obteniendo como producto el Informe de Arquitectura de Red.
Escaneo: donde se busca descubrir y determinar todas las características propias del sistema, las tareas que la conforman son: Detección de sistemas vivos en la red y Descubrimiento de puertos activos, sistema operativo, servicios en ejecución y presentes en el sistema y direcciones IPs.
Ganancia de acceso: donde se plantean las siguientes tareas: Escaneo de direcciones, Investigación de puertos y Explotación de servicios y sistemas. Dejando como producto el Informe de Escaneo de sistema. Generando como producto el Informe de Accesos Ganados.
Mantenimiento de acceso: la cual es integrada por las tareas: Mantenimiento de acceso, Aseguramiento de acceso exclusivo, Carga, descarga y manipulación de datos, aplicaciones y configuraciones en el sistema y Utilizar el sistema para lanzar más ataques. Produciendo el Informe de Accesos Mantenidos.
Eliminación de Pruebas: la cual tiene como objetivo el ocultamiento de huellas y la apertura para futuros accesos al sistema.Para esto tiene como tareas las siguientes: Ocultamiento de actos maliciosos, Continuación del acceso al sistema víctima, Sobre escritura del servidor, los sistemas y el registro de aplicaciones y el armado del informe de Pruebas Eliminadas.
Todas estas etapas, tareas, actividades y herramientas, nos llevan a obtener el *Informe Final de Testeo por Hacking Ético (IFTHE).* Este informe contiene la información de que tan crítico, vulnerable o protegido esta nuestro sistema próximo a salir a Producción.

Fase de Mantenimiento

La cual vela por que los sistemas que no son vulnerables en su puesta productiva, no sean posibles objetivos ante nuevas amenazas. Para lograr esto se proponen 3 etapas.
Control de Vulnerabilidades: busca constantemente actualizarse con las vulnerabilidades nuevas que surgen y mide que tan críticas son en nuestras

aplicaciones. Para esto tiene las siguientes actividades ordenadas: Listado de vulnerabilidades nuevas, Determinación de grado de aplicabilidad. Dando como producto el Informe de Vulnerabilidades criticas.
Determinación de Criterios: en ella las tareas Determinación del grado de criticidad, Establecimiento de criterios de aprobación y rechazo y Listado del ranking de vulnerabilidades más riesgosas, buscan determinar que tan crítico y con qué prioridad se deben corregir las vulnerabilidades, donde se genera un Listado de Criterios de Aprobación y Rechazo.
Verificación y Validación: donde las tareas a realizarse son: Ejecución de pruebas de verificación, Comparación con criterios establecidos, Determinación de la criticidad del riesgo, Establecimiento de las acciones a tomar, Validación de los resultados y la Elaboración del informe.
Esta etapa final tiene da como producto el Informe de Verificación y Validación. Todo este proceso, proporcionará como salida un *informe general de riesgos, validaciones y verificaciones. (IGRVV)*, el cual resumirá los riesgos, criticidades y consejos a seguir para el software puesto a prueba.

5. Métricas de vulnerabilidad

Durante la fase de Ejecución del Testeo se plantea el uso de métricas que nos indiquen el grado de vulnerabilidad del sistema al ser desarrollado. Las métricas propuestas son las siguientes:
Índice de posibilidad de acceso (IPA): es directamente proporcional a la cantidad de accesos ganados (CAG) e inversamente proporcional a la cantidad de puertos activos detectados (PAD). Es un índice que tiene un valor entre 0 y 1, siendo los valores cercanos a 0 los que nos dan una medida de un sistema más seguro y un valor cercano a la unidad un sistema más vulnerable.

$$IPA = \frac{CAG}{PAD} \qquad 0 <= IPA <= 1 \qquad \left| \qquad IPA\ \% = \frac{CAG}{PAD} * 100\ \% \qquad IPA << 1 \right.$$

Fórmula 3. Índice de posibilidad de acceso

Índice de acceso real (IAR): directamente proporcional a la cantidad de accesos ganados (CAG) e inversamente proporcional a los puertos activos reales (PAR) del sistema. Es un índice que tiene un valor entre 0 y 1, siendo los valores cercanos a 0 los que nos dan una medida de un sistema más seguro y un valor cercano a la unidad un sistema más vulnerable.

$$IAR = \frac{CAG}{PAR} \qquad 0 <= IAR <= 1 \qquad \left| \qquad IAR\ \% = \frac{CAG}{PAR} * 100\ \% \qquad IAR << 1 \right.$$

Fórmula 2. Índice de acceso real

Índice de detección de puertos (IDP): directamente proporcional a la cantidad de puertos activos detectados (PAD) e inversamente proporcional a los puertos activos reales (PAR) del sistema. Es un índice que tiene un valor entre 0 y 1, siendo los valores cercanos a 0 los que nos dan una medida de un sistema más seguro y un valor cercano a la unidad un sistema más vulnerable.

$$IDP = \frac{PAD}{PAR} \qquad 0 <= IDP <= 1 \qquad \left| \qquad IDP\ \% = \frac{PAD}{PAR} * 100\ \% \qquad IDP << 1 \right.$$

Fórmula 3. Índice de detección de puertos

6. Validación

Para validar la propuesta se construyó una maqueta la cual contiene una maquina virtual Linux donde se instaló la distribución Kali del sistema operativo y por otro lado dos maquinas virtuales mas con una distribución Ubuntu de Linux y Windows XP respectivamente.
Se realizaron dos aplicaciones web para someter a diferentes pruebas, las cuales son muy diferentes conceptualmente, ya que una es un blog de posteos y la otra un gestionador de archivos con login y ftp de archivos.
Tomando como base el modelo de proceso propuesto, se llevaron a cabo las tareas propuestas en cada una de las etapas de las tres fases, comenzando por la planificación, definiendo la estructura (creación de plantillas a seguir durante las subsiguientes fases) del proceso de testeo.
En primer lugar junto al análisis de requerimientos y la documentación de la aplicación, se obtuvo el Informe de dominio, una vez obtenido esto ayuda a realizar la segunda etapa, en la cual se realiza el informe de vulnerabilidades, donde se determina la probabilidad de vulnerabilidades y se ordenan generando un ranking de acuerdo a su propensión. Luego de tener las posibles debilidades listadas, se identificaron y documentaron las mas criticas armando el informe de amenazas y prioridades de criticidad.
Junto a toda esta información, se realiza la última etapa del planeamiento donde se armaron los casos de prueba, se calendariza y se identifican los criterios de aceptación y rechazo.
En las pruebas llevadas a cabo dentro de las vulnerabilidades detectadas hay muchas coincidentes, pero otras que son propias de cada escenario.
Para la fase de ejecución, se utilizaron múltiples herramientas para cada etapa, las cuales se brindan como ejemplo a continuación.
Para recopilar la arquitectura de la red y realizar el escaneo de puertos y servicios activos, se utilizaron herramientas instaladas en la distribución ya mencionada tales como Nmap, masscan SQLmap y wireshark . Los resultados son reflejados en los infromes de arquitectura y escaneo respectivamente.
Para las etapas de ganancia de acceso y mantenimiento, ambos escenarios fueron sometidos a las mismas herramientas, intersec, powersploit, inundator y Slowhttptest,

donde además de recorrer y explotar puertos, se intentó inundar la comunicación y denegar el servicio.
Por último se probaron herramientas forenses para eliminar las pruebas de intrusión.
Recopilando todos los informes se redacta el informe final de testeo por hacking ético, donde se detallan las vulnerabilidades detectadas y en qué medida se cumplen los criterios de rechazo y aprobación.
Como última fase, y ya pensada luego de la puesta en producción, se encuentra la fase de mantenimiento, donde se reutilizan algunas de las herramientas ya mencionadas y se busca mantener la aplicación libre de vulnerabilidades.

7. Conclusiones y futuras líneas de investigación

Durante las pruebas realizadas se pudo comprobar la utilidad de planificar las tareas de testeo de hacking ético, al tener una guía de actividades y consulta de herramientas para las tareas de testeo, recolección de información, análisis y toma de medidas correctivas.
Algunas de estas tareas pueden llegar a pasarse por alto de no contar con una planificación adecuada, junto a un conjunto de medidas a tener en cuenta y herramientas para utilizar.
Si bien se han realizado solamente dos casos de validación, la utilización de las métricas de vulnerabilidad ofrecen una visión general del estado inicial del sistema al ser desarrollado (en cuanto a vulnerabilidades se trate). El empleo de estas métricas y análisis de más proyectos de desarrollo de software siguiendo el presente modelo de proceso para incorporar el haking ético durante el testeo de software, harán posible el desarrollo de indicadores sobre estas métricas para, incluso, poder llegar a realizar una clasificación de sistemas de acuerdo a su grado de vulnerabilidad necesario y el que posee realmente. Así, un sistema de home banking tendrá un indicador de las métricas de vulnerabilidad mucho más ajustado que el de un blog hobista, por indicar un ejemplo.

8. Bibliografía

ANSI/IEEE, (2007). Draft IEEE Standard for software and system test documentation. ANSI/IEEE Std P829-2007.

Comunidad Linux. 2014 <http://www.linux.org/> Página Válida a 05/2018

Coronel Suarez, I.A. (2016). Aplicar hackeo ético para detección de vulnerabilidades mediante herramientas Open Source en las aplicaciones web de una institución de educación superior. Tesis Maestría de la Escuela Superior Politécnica del Litoral.

Evans, Bob (2001). The Sorry State of Software. InformationWeek 112.

GNU (2014) Operating System Sponsored by the Free Software Foundation. 2014/05/15 <http://www.gnu.org/> Página Válida a 05/2018

Hurtado Sandoval, M.E., Mendaño Mendaño, L.A. (2016). Implementación de técnicas de hacking ético para el descubrimiento y evaluación de vulnerabilidades de la red de una cartera de Estado. Tesis Grado. Escuela Politécnica Nacional de Ecuador.

ISO 27035:2011 (2010). Information technology – Security techniques – Information security incident management [Online]. http://www.iso.org/iso/catalogue_detail?csnumber=44379.

López Alvarez, D.M. (2015). Hacking ético para detección de vulnerabilidades de una empresa del sector de telecomunicaciones. Tesis Maestría de la Escuela Superior Politécnica del Litoral.

López Vallejo, M.R. (2017) Hacking ético. Vulnerabilidad de Sistemas Operativos en el acceso por contraseñas. Revista Publicando, 4 No 10. (1). 2017, 31-51. ISSN 1390-9304. <http://rmlconsultores.com/revista/index.php/crv/article/view/407/pdf_259> Página Válida a 05/2018

Mayorga Jácome, T., Quisaguano Belduma, F.J. (2015). Implementación de hacking ético para el análisis de vulnerabilidades, métodos de prevención y protección aplicados a la infraestructura de red de la empresa Construlec Cía. Ltda. en Quito Ecuador. Editorial: Quito: Universidad Israel.

OISSG (2012) Open Information Systems Securty Group. 2003 – 2012 <http://www.oissg.org/> Página Válida a 05/2018

Onofa Calvopiña, F.O., Pilatuña Chica, I. (2016). Análisis y evaluación de riesgos y vulnerabilidades del nuevo portal web de la Escuela Politécnica Nacional, utilizando metodologías de hackeo ético. Tesis Grado. Escuela Politécnica Nacional de Ecuador.

OWASP Top 10 - 2013 (2013). [Online]. http://www.owasp.org. Página Válida a 06/2018.

Palmer, Charles (2001). Ethical hacking, IBM Systems Journal, Vol. 40, N°3

Raymond E (1991). The New Hacker's Dictionary, MIT Press, Cambridge, MA

Santos Castañeda, D.M. (2016). Análisis y diagnóstico de vulnerabilidades informáticas en la red de datos de la empresa YOUPHONE Cía. Ltda. Utilizando Hacking Ético. Tesis Grado. Institucional de la Universidad de las Fuerzas Armadas ESPE.

Sheoran, Pankaj & Singh, Sukhwinder (2014). Applications of Ethical Hacking, International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463 Vol. 3 Issue 5, 05/2014 112-114

Sommerville, I., 2005. Ingeniería del Software. Séptima edición. Capítulo 1. Pearson Addison. ISBN: 84-7829-074-5.

Tenable Network Security (2014), provedora de la herramienta Nessus, 2014 <http://www.tenable.com/products/nessus?gclid=CK2xwJGavr4CFScHwwod9VMAZA> Página Válida a 05/2018

Tori C. (2008). Hacking Ético (1ra Ed). Buenos Aires: Mastroianni.

Software para Decodificación y Análisis de Archivos Binarios en Computadoras Navales

Mauricio J. Mancini[1,2], Christian L. Galasso[1,2] y Miguel A. Banchieri[1]

[1] Grupo SITIC - Dpto. de Ing. Electrónica, Facultad Regional Bahía Blanca, Universidad Tecnológica Nacional, Argentina.
[2] Servicio de Análisis Operativos, Armas y Guerra Electrónica, Base Naval Puerto Belgrano, Armada Argentina.
mancini_mauricio@hotmail.com, christian_galasso81@yahoo.com.ar, mbanch@frbb.utn.edu.ar

Abstracto. En el presente trabajo se detalla el diseño de un software de decodificación de archivos y análisis de grabaciones de eventos navales. Originalmente, las embarcaciones contaban con un programa operativo de características similares, que se ejecutaba sobre las computadoras de 24 bits de abordo. La única interfaz que se disponía era una terminal tipo teletipo donde el operador insertaba comandos y se imprimían los resultados sobre un papel, el cual se utilizaba posteriormente para analizar los resultados del evento naval realizado. Debido a sus múltiples defectos y problemas, se decidió realizar una reingeniería del mismo para poder sustituirlo con un software que se ejecute sobre una PC de uso comercial, brindándole al usuario mayores herramientas y facilidades a través de las tecnologías disponibles en la actualidad.

Keywords: Decodificación, reingeniería, computadoras navales.

1 Descripción del Sistema Actual de Análisis

Las unidades navales, recepcionadas a partir del año 1982, poseen una serie de computadoras conectadas en red, las cuales conforman un sistema en tiempo real duro. La gran mayoría de ellas poseen la capacidad de realizar una grabación de las maniobras y datos actuales, denominado DataLogging (DL). Esta tarea genera un archivo binario, con un formato propio del sistema, en el cual se almacenan gran cantidad de datos correspondientes a sensores, movimientos de actuadores, estados y cálculos propios del sistema naval, durante el tiempo que se esté ejecutando esta tarea de DL. Los datos registrados varían en función de la práctica y la consola donde se realice la grabación, totalizando 13 posibles escenarios diferentes. Al finalizar la grabación, se crea un archivo, denominado COPYCASS, donde se encuentran todos los datos grabados de la práctica. Este archivo posee un tamaño variable que es

directamente proporcional a la duración de la grabación, pudiendo llegar a 6 horas como máximo (aproximadamente 16 MB).

1.1 Procedimiento Original de Extracción de un DL

Una vez que se tiene el archivo que contiene los datos grabados durante la práctica, se debe extraer para que sean analizados posteriormente por el personal correspondiente. Para hacer esto, se debe cargar un programa sobre la computadora de abordo denominado PFRD (Programa de Formateo y Reducción de Datos). Una vez ejecutado, se deben insertar una serie de comandos, dependiendo del tipo de evento y de los datos que se deseen extraer, por la terminal tipo teletipo de abordo. Terminado este proceso, se comenzarán a escribir diversos grupos de información, referente a conjuntos de datos específicos de cada sector de la embarcación. Se escribirá una fila completa de datos (entre 6 y 15 variables diferentes) por cada intervalo de tiempo grabado, una vez finalizado, escribirá otra fila en la parte inferior correspondiente al siguiente intervalo, siguiendo este proceso hasta terminar el total de la grabación.

La duración del proceso de extracción varía en función de la duración total de la grabación realizada, aunque este difiere según la práctica realizada ya que cambian los intervalos de grabación y la cantidad de datos mostrados sobre el papel. Para una extracción completa de uno de los eventos más comúnmente grabados, el tiempo de extracción de 15 minutos de grabación se puede extender a más de 8 horas. La duración total se torna tan extensa ya que se deben imprimir, para 15 minutos de grabación, aproximadamente 100000 valores de variables diferentes, todo a través de un teletipo que funciona de manera electromecánica y tiene obvias limitaciones respecto a la velocidad máxima de impresión que puede lograr.

El gran tiempo necesario para realizar una extracción no es la principal deficiencia que posee el sistema actual. Hay que tener en cuenta que durante todo el proceso, se debe dar alimentación al barco de 115 V 400 Hz a través de generadores impulsados con motores de combustión. Esto conlleva a que no solo el proceso de extracción de una grabación es muy extenso sino que además costoso.

1.2 Análisis de los Resultados

Una vez que se obtiene el resultado del PFRD impreso en papel, personal especializado se encarga de analizar estos valores para juzgar y calificar el desempeño de los operadores en sus respectivos cargos. Estos deben buscar valores específicos a lo largo de todos los datos impresos, para luego verificar los datos de otros parámetros en ese mismo instante de tiempo. Esta tarea es ardua y laboriosa ya que el hecho de buscar valores específicos en datos impresos sobre un papel es muy ineficiente y vuelve al proceso de análisis muy extenso.

2 Desarrollo de Aplicación PFRD Engine

Debido a las abundantes problemáticas que posee el actual sistema de extracción de DL, se planteó sustituirlo por una versión más moderna, la cual pueda ser ejecutada sobre una PC de tipo comercial y brinde herramientas que faciliten el análisis.

Para lograr el desarrollo de un sustituto, se debían realizar diversas pruebas al programa original, ya que no se posee el código fuente de este y la documentación disponible es escasa. Como primer paso se debía conocer el formato de los archivos generados por el sistema para identificar el modo en que se almacena la información. Luego se tenía que poder asociar los datos leídos a las variables mostradas por el teletipo descubriendo en qué posición se encuentra cada dato y de qué manera se convierten para ser mostrados al operador. Finalmente se debía desarrollar un software que se ejecute sobre la PC el cual debe analizar el archivo, extraer los datos, convertir las variables mostradas al operador y entregarlas de una manera más sencilla para su posterior análisis.

Se decidió denominar a este software "PFRD Engine" ya que es un motor de cálculo para convertir la información de un archivo de DL a valores de parámetros interpretables por el operador encargado del análisis.

El programa se desarrolló íntegramente en lenguaje C [1] [2], ya que brinda todas las herramientas necesarias y es fácil de implementar. El entorno de desarrollo elegido fue, el software libre, CodeBlocks [5].

2.1 Estructura del Archivo Generado

El archivo de generado (COPYCASS), se compone de distintos bloques de diversos tipos y tamaños, pudiendo ser: Volumen Header, Mark, File Header, Preset Block, Data Block, Last Block, File Trailer y Device Trailer. La estructura de los archivos sigue un patrón predeterminado (Fig.(1)), necesario para que el programa operativo pueda analizar sus bloques e interpretar sus contenidos.

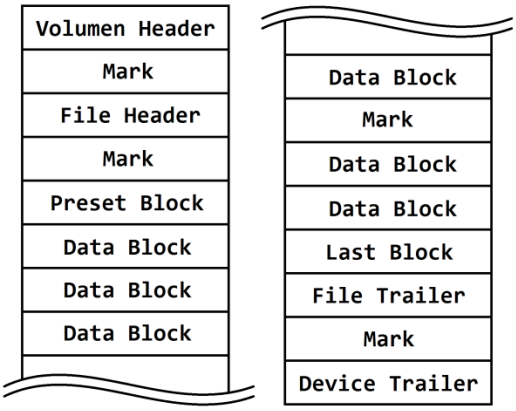


Fig. 1. En la imagen se puede ver la estructura del archivo COPYCASS segmentada en bloques de datos. Se identifica por el nombre a cada uno de ellos y se muestra el orden en el que estos se presentan normalmente en un archivo de DL. Se muestra el comienzo del archivo sobre la derecha y el final sobre la izquierda.

Como primer bloque de un archivo de DL, siempre se encuentra un Volumen Header. Este posee un tamaño de 255 bytes (85 palabras) de los cuales solo los 6 primeros contienen información. Comienza con los caracteres 'DL', seguido de cuatro números identificando al tipo de unidad naval. El resto del contenido no se toma en cuenta, aunque normalmente se completa con ceros. Seguido de este bloque se espera ver una marca (Mark). Dicho bloque solo posee dos bytes, los cuales están todos en 1 (0xFFFF). Se usa generalmente de separación, necesario cuando se leía en una cinta magnética por tiempos de aceleración y frenado. Como tercer bloque, se espera un File Header, el cual posee información más específica de la consola en la que se está realizando el DL. El tamaño de este bloque es de unos 81 bytes (27 palabras). Luego se vuelve a tener una marca sin información que sea de utilidad. En este punto suele aparecer un primer bloque denominado de Preset Block. Contiene información sobre el DL a realizar tal como, grupos, frecuencia de grabación de datos e información del sistema. Posee un tamaño variable, normalmente de 255 bytes (85 palabras), aunque en la primera palabra del bloque nos aclara su longitud. Llegado a este punto del archivo, es cuando aparece la información que se analizará posteriormente a través de unos bloques denominados Data Block. Estos contienen diversos grupos, dependiendo de cuál sea la práctica o ejercicio grabado. Dichos grupos son un conjunto de datos que almacenan información específica de cada subsistema en un intervalo de tiempo, tal como sensores, actuadores, rumbo, estados de la consola, otros. La longitud de los Data Block es usualmente de 255 bytes, aunque puede ser diferente en algunos casos, no obstante, en la primer palabra aclara cuál es su tamaño (Fig.(2)). Los diferentes grupos son almacenados en el Data Block hasta que complete su longitud predeterminada, pudiendo o no repetir un tipo de grupo dependiendo de la cantidad de

información que se grabe. De repetirse alguno en un mismo Data Block, significa que pertenece a un intervalo de tiempo diferente.

A partir de este punto se pueden recibir varios tipos de bloques en el archivo, usualmente son más Data Blocks cuya cantidad depende de la duración de la grabación. También se espera ver una marca cada 20 bloques de datos, solo a modo de separación. Por último, se lee un Last Block que nos muestra el final de la grabación del archivo. Se caracteriza por tener un tamaño de 255 bytes, siendo el primero un 1 y el resto de los 254 bytes 0's. Si se finalizó la grabación, el siguiente bloque es el File Trailer, de una longitud de 255 bytes de los cuales los primeros 3 contiene información mostrando los caracteres '/ / /'. Luego de este bloque se espera otra marca a modo de separación. Finalmente el último bloque del archivo es un Device Trailer, con una longitud de 81 bytes, cuyos primeros 9 poseen caracteres indicando que se trata del final del archivo.

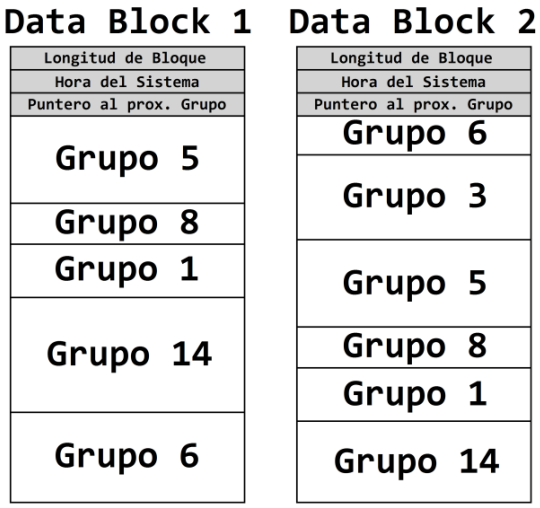


Fig. 2. En la imagen se puede ver un ejemplo de dos Data Blocks. En ellos se muestra cómo se organizan los datos en su interior separados por grupos. En el caso de la imagen, el *grupo 6* se termina de completar en el segundo bloque ya que completó el tamaño total del *Data Block 1*.

2.2 Decodificación del Archivo

El programa PFRD Engine, como primera tarea a ejecutar, realiza una lectura completa del archivo COPYCASS. Durante el recorrido del archivo se busca verificar que los bloques leídos correspondan al formato y orden de un DL. De encontrar algún bloque que no cumpla con el formato preestablecido, el programa detendrá su ejecución normal e informará al usuario el tipo de error y el número de bloque en el que se detectó. A medida que se leen los bloques, extrae su información útil y la

discrimina según corresponda. Como se mencionó previamente, los grupos poseen información específica de conjuntos de datos. Estos poseen un tamaño que varía según el grupo del que se trate y una estructura predeterminada. La primer palabra (24 bits) de cualquier grupo pertenece a un encabezado donde se dice el número de bloque y cuál es su longitud. A continuación, contiene datos de variables y parámetros, uno por cada palabra, aunque hay casos en que se utilizan variables booleanas y existen diversos estados dentro de una misma palabra.

Agrupa todos los datos de cada grupo por separado y se extraen datos de la grabación de los primeros bloques como el de Preset y encabezados. Finalizado este proceso, se posee cada grupo separado en orden temporal y listo para ser analizado.

2.3 Conversión de Datos

Teniendo los datos discriminados por grupos, se deben transformar a las variables mostradas por el PFRD. Para la extracción, el operador debe insertar unos nemónicos que simbolizan un conjunto de datos a extraer, como puede ser el conjunto de sensores del sistema (nemónico SEN), datos propios del barco (nemónico OSD) o datos de iniciales de la grabación (nemónico PRE), entre otros. En total existen 18 posibles conjuntos de datos para todas las grabaciones que se pueden realizar abordo. Los conjuntos obtienen la información que necesitan de grupos de datos específicos, uno o más, para luego transformarla y adaptarla.

La conversión de datos varía según el parámetro a mostrar, pudiendo ser una conversión lineal o a través de una operación algebraica. En las conversiones lineales, como se muestra en la Ec.(1), se divide al valor leído por una constante de conversión, que se relaciona con el tipo de unidad métrica del parámetro.

$$\text{resultado} = \text{valorLeido} / \text{ConstConversion} . \quad (\mathbf{1})$$

Las constantes de conversión se obtuvieron mediante diversas pruebas ya que no se encontró ningún manual donde se las mostrara. Inicialmente, se realizaban DL y se extraían por el método original para poder obtener las constantes de conversión. Hubo una gran cantidad de constantes que no se pudieron descifrar ya que muchos de los valores no variaban o se mantenían directamente en cero. Esto se debe a que los parámetros y variables de la embarcación rara vez tienen grandes variaciones para un periodo de tiempo corto, como puede ser el ejemplo de la temperatura ambiente o la presión atmosférica. Frente a esto, se decidió modificar manualmente las palabras de un archivo de DL y forzar a las variables a que tomen diferentes valores para observar la respuesta del programa original a cualquier tipo de entrada, incluso cuando esta tenga valores incoherentes.

$$\text{resultado} = \arctan(\text{valorLeido_1}, \text{valorLeido_2}) . \quad (\mathbf{2})$$

$$resultado = arctan(valorLeido_1, sqrt(valorLeido_2^2 + valorLeido_3^2)) . \quad (3)$$

$$resultado = sqrt(valorLeido_1^2 + valorLeido_2^2 + valorLeido_3^2) . \quad (4)$$

Otro tipo de conversión de datos se realiza a través de distintas operaciones algebraicas, como se muestra en las Ec.(2), (3) y (4). Estas son definidas en los manuales, aunque en la práctica se encontraron algunas diferencias y también el hecho de que las variables se deben escalar a sus respectivas unidades antes de realizar los cálculos pertinentes. En las ecuaciones se pone "valorLeido_X" haciendo referencia a una palabra especifica de un grupo de datos predeterminado.

Se encontraron ciertos valores que no correspondían con ninguna conversión de las definidas previamente. En este caso se tuvieron que colocar mayor cantidad de valores en los campos de los grupos para tener mayor cantidad de muestras. Hacer esto de manera manual, editando el archivo binario, era inviable ya que pueden existir 16.777.216 valores posibles (2^{24}). Es por eso que se desarrolló un software que generara archivos de DL con los valores que se deseen. De esta manera se pudo descubrir el algoritmo conversión de ciertos parámetros. El software desarrollado también fue de gran utilidad para testear el funcionamiento del PFRD Engine, pudiéndole ingresar una gran cantidad de datos con convinaciones de valores que difícilmente de puedan replicar en una grabación convencional.

2.4 Interfaz de Usuario

Una vez que se convierten todos los valores a unidades físicas interpretables por el operador, se le entregan a través de un documento de Excel (.csv). Se genera un archivo por cada conjunto de datos (SEN, OSD, PRE, etc) que el usuario desee extraer. Dentro del documento, se muestran en las columnas los distintos parámetros y variables predefinidas para cada conjunto y en las filas los tiempos en los que se midió cada uno de estos valores (Fig.(3)).

Grupo SEN							
IDENT:	222222	INT:	660 MS				
TIME	**VAR 1**	**VAR 2**	**VAR 3**	**VAR 4**	**VAR 5**	**VAR 6**	**VAR 7**
SEC	**SEC**	**Y**	**DEG**	**DEG**	**DEG**	**Y/S**	**Y/S**
0.00	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X
0.66	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X
1.32	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X
1.98	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X
2.64	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X
3.30	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X	XXX.X

Fig. 3. Se muestra conceptualmente como se entrega la información al operador. En las columnas, se pone *VAR X* haciendo referencia a los parámetros y variables propias de cada grupo. Debajo de ellas se muestra el tipo de unidad, como por ejemplo *SEC* (*segundos*), *Y*

(*yardas*), *DEG* (*grados*), *Y/S* (*yardas por segundo*), entre otros. En la parte superior se muestra el *nombre del grupo* (*SEN*), el *identificador* de la grabación (*IDENT*) y el *intervalo* de tiempo de la grabación (*INT*).

Al entregarse los resultados en un archivo compatible con Excel (o similar), se busca brindarle más herramientas al usuario, aprovechando las propias del software que son ideales para análisis de gran cantidad de datos numéricos. Entre las herramientas que se pueden utilizar están la generación de gráficas, búsqueda de valores específicos, filtrado de datos y utilización de operadores matemáticos y estadísticos sobre un conjunto de datos. De esta manera, el usuario podría detectar el valor máximo, o mínimo, o promedio, de miles de valores en una fracción de segundo, siendo que por el método original debería hacerlo manualmente leyendo cada uno de estos desde un papel.

3 Testeo del Programa

Durante el desarrollo del programa, se lo sometió a una serie de pruebas [3] [4] donde se comparaban los resultados obtenidos con los que se mostraban en el PFRD original. Primero se crearon diversos archivos de DL donde se forzaban los valores. Se introdujeron gran cantidad de ellos para simular todas las posibilidades y observar como respondían ambos programas. Finalmente se confeccionaron planillas comparativas entre ambos PFRDs demostrando el correcto funcionamiento del nuevo software.

El segundo procedimiento de pruebas fue el de recolectar diversas pruebas de prácticas realizadas con anterioridad, para ser extraídas nuevamente pero esta vez con el PFRD Engine. Se volvieron a confeccionar las planillas mostrando los resultados de ambos programas, pero esta vez sobre archivos de prácticas reales abordo.

Superadas las dos pruebas de testeo se determinó el correcto funcionamiento del nuevo software y de su respuesta frente a cualquier tipo de archivo, ya sea con valores reales de una práctica, como de valores forzados que difícilmente se puedan encontrar en una grabación real o simulación.

4 Futuros Desarrollos

Actualmente se encuentra en desarrollo un nuevo software de análisis de archivos de DL. Este utiliza al PFRD Engine como motor de cálculo para obtener todos los valores ya convertidos, pero brinda al operador más herramientas que faciliten el análisis de las prácticas navales.

Se plantearon, en conjunto con el personal especializado, objetivos y funciones que se deben implementar en el nuevo desarrollo de software. Entre ellos se puede nombrar la búsqueda de valores específicos, navegación temporal del archivo,

capacidad de visualizar parámetros específicos de distintos conjuntos de datos o grupos facilitar las comparativas y la generación de archivos de tipo PDF protegidos para mostrar los resultados completos de manera segura y no a través de archivos fácilmente corruptibles.

Otro proyecto a futuro, es el de incorporar un software desarrollado previamente. Este programa estudia los eventos navales a través de videos tomados a bordo de la embarcación. Actualmente, posee una serie de errores o bugs que dificultan su utilización. Se planteó desarrollar un sustituto de este programa e incorporarlo al de análisis de archivos de DL que utiliza el PFRD Engine como motor de cálculo. De esta manera se podrían agrupar todas las herramientas de análisis de prácticas navales dentro de un mismo software.

Referencias

1. Darnell, P.A., Margolis, P.E.: Software Engineering in C. Springer, New York (1988)
2. Pagnoni, A.: Project Engineering, Computer-Oriented Planning and Operational Decision Making. Springer, Berlin Heidelberg (1990)
3. Stewart, G.: Testing Safety-Related Software, A Practical Handbook. Spring, London (1999)
4. Heitmeyer C.: A Model-Based Approach to Testing Software for Critical Behavior and Properties. In: Petrenko A., Simão A., Maldonado J.C. (eds) Testing Software and Systems. ICTSS 2010. Lecture Notes in Computer Science, vol 6435. Springer, Berlin, Heidelberg (2010)
5. Open source software CodeBlocks, http://www.codeblocks.org

Propuesta de un Modelo de Proceso de Implantación de Sistemas Informáticos (MoProIMP)

Marisa Panizzi [1,2], Rodolfo Bertone[3], Alejandro Hossian[4]

[1] Programa de Doctorado en Ciencias Informáticas. Facultad de Informática. Universidad Nacional de La Plata.
[2] Departamento de Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional. Facultad Regional Buenos Aires, Medrano 951 (C1179AAQ), C.A.B.A, Argentina
[3] Instituto de Investigaciones en Informática LIDI. Facultad de Informática. UNLP – CIC
[4] Grupo de Investigación en Aplicaciones de Sistemas Inteligentes en Ingeniería. Universidad Tecnológica Nacional – Facultad Regional Neuquén, Av. Pedro Rotter S/N Barrio Uno, (8318), Neuquén, Argentina
marisapanizzi@outlook.com;pbertone@lidi.info.unlp.edu.ar;alejandrohossian@yahoo.com.ar

Resumen. A partir del análisis de metodologías de desarrollo de software, se ha observado que varias de ellas no contemplan el proceso de implantación del sistema o, si lo hacen, no lo desarrollan o gestionan de manera integral.
Como respuesta a las vacancias identificadas, se propone definir MoProIMP, un modelo que permite sistematizar el proceso de implantación de software. Este modelo puede ser utilizado en conjunto con la metodología de desarrollo de la organización productora de software.
MoProIMP identifica fases, actividades, tareas (entradas y salidas) y dependencias existentes en una implantación de software, gestionando sistémicamente cada una de ellas. Por último, se presenta un caso de estudio que sirve como validación inicial al modelo propuesto.

Palabras Clave: procesos de software, implantación de sistemas informáticos, modelo de proceso, desarrollos a medida.

1 Introducción

Se puede definir implantación de sistemas informáticos como el conjunto de actividades y tareas necesarias que permiten la transferencia del producto software finalizado a su ambiente de utilización por parte de la comunidad usuaria. [1]
La ingeniería de software ha desarrollado un conjunto de metodologías y estándares para la construcción de sistemas informáticos. Algunos de las mismos pertenecen a los denominados “tradicionales o robustas” [2] y otros pertenecen al grupo de los denominados “ágiles” [2]. A pesar de la diferencia de alcance existente entre una metodología [4], un estándar [5][6] o un método [7], en este artículo, los autores han decidido darles un tratamiento equivalente, en base en como cada una de ellos soportan el proceso de implantación.

En los trabajos [1] y [3] se han presentado una revisión sistemática de un conjunto de metodologías o estándares de desarrollo de software con el propósito de estudiar el proceso de implantación de sistemas informáticos.

A partir del análisis de las metodologías y estándares revisados, se identifican una serie de deficiencias y/o carencias para la ejecución exitosa del proceso de implantación de sistemas informáticos. A partir de las ausencias identificadas, se propone un **Mo**delo de **Pro**ceso de **IMP**lantación de sistemas informáticos, que se denomina MoProIMP. Este proceso determina fases, actividades, tareas, técnicas de representación y procedimientos de ejecución de cada tarea, sistematizando la implantación de software.

La siguiente sección presenta una propuesta de solución, en la cual se describe y justifica la arquitectura del modelo de proceso propuesto. Posteriormente se presenta el primer caso de estudio y luego las conclusiones obtenidas y las propuestas de mejoras sobre el proceso.

2 Modelo de Proceso de IMPlantación (MoProIMP)

La arquitectura de MoProIMP presenta su modelo de ciclo de vida adoptando 5 grupos de procesos. Esto surge a partir del análisis de PMI BOOK [8]. Estos grupos son: *Iniciación, Planificación, Ejecución, Seguimiento y Control y Cierre*. El motivo de esta elección, se debe a que PMI BOOK es un estándar reconocido a nivel mundial y de aplicación en la industria del software a nivel nacional. A cada uno de estos procesos, el modelo propuesto lo denominará *fase*.

Para llevar a cabo la aplicación del modelo en la industria del software en la Argentina y que se realice de manera escalonada, se han adoptado tres de los niveles de capacidad, como los propuestos en el estándar CMMI-DEV [9]. Los niveles de capacidad considerados son: *nivel 1= Realizado, nivel 2= Gestionado y nivel 3= Definido*. No se ha considerado el nivel *0= Incompleto* ya que significa la no realización o la realización parcial de ese proceso en la organización. Los tres niveles de capacidad considerados se han analizado a un nivel de granularidad de las tareas propuestas en el modelo desarrollado. La elección de considerar los niveles de capacidad en lugar de los niveles de madurez, se debe a que en el contexto de la industria Argentina, son pocas las empresas que han alcanzado niveles de madurez 4 y 5, por lo tanto esto dificultaría llevar a cabo la validación del modelo propuesto.

El modelo adopta un grupo de tareas propuestas en la metodología Métrica versión 3[10] por considerarse una de las metodologías más completas, realizando algunas adaptaciones en referencia a la industria nacional.

En este trabajo, se presentan las fases, las actividades y sus objetivos, las tareas (identificando las vinculaciones) y los elementos de dependencia entre las mismas (los elementos de entradas y las salidas).

En la figura 1, se ilustran las fases, las actividades y las interdependencias entre las actividades que componen al modelo propuesto. Para la representación del modelo se ha utilizado la técnica SADT (Structured Analysis and Design Technique) [11] con algunas adaptaciones.

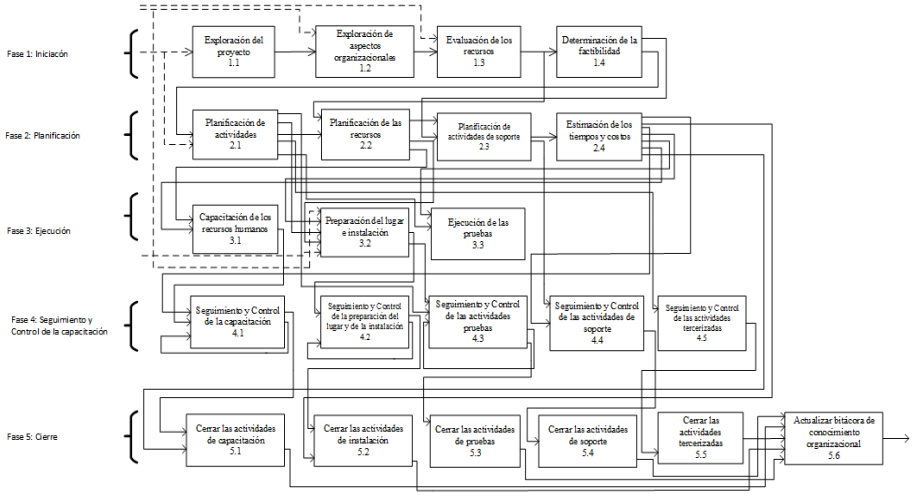


Figura 1. Proceso de Implantación (**MoProIMP**).

La **Fase 1: Iniciación**, se compone de cuatro actividades: *Exploración del proyecto*, cuyo objetivo consiste en la realización de un estudio de las características del proyecto y de la documentación de los requisitos establecidos para el mismo; *Exploración de los aspectos organizacionales*, cuyo objetivo consiste en la identificación de los mecanismos de documentación, comunicación para la implantación en el contexto organizacional y la gestión de configuración; *Evaluación de los recursos,* cuyo objetivo consiste en la revisión de los recursos necesarios para llevar a cabo la implantación y *la Determinación de la factibilidad* cuyo objetivo consiste en identificar los riesgos y definir la factibilidad de la implantación. En la Tabla 1, se presenta la **Fase 1: Iniciación,** con las tareas de cada actividad, el nivel de capacidad determinado para la tarea y las entradas y las salidas de cada una de las tareas.

La **Fase 2: Planificación**, se compone de cuatro actividades: *Planificación de las actividades,* cuyo objetivo consiste en establecer un conjunto de actividades que conforman el proyecto de implantación considerando los entregables, las pruebas y las actividades tercerizadas; *Planificación de los recursos* cuyo objetivo consiste en la determinación de los recursos tanto humanos como tecnológicos; *Planificación de actividades de soporte,* cuyo objetivo consiste en la identificación de las características a medir en el transcurso de la implantación, la determinación de los elementos de gestión de configuración y establecer la gestión de riesgos ; *Estimación de tiempos y costos,* cuyo objetivo consiste en la definición de los tiempos y costos para el desarrollo de las actividades de implantación. En la Tabla 2, se presenta la **Fase 2: Planificación,** con las tareas de cada actividad, el nivel de capacidad determinado para la tarea y las entradas y las salidas de cada una de las tareas.

Tabla 1. Fase 1: Iniciación, actividades, tareas, nivel de capacidad, entradas y salidas

Fase	Actividades	Tareas	NC	Entradas	Salidas
1. Iniciación	1.1.	1.1.1.Revisar el plan del proyecto	2	Plan del proyecto	**Reporte de situación de la implantación**
		1.1.2. Revisar documento de requisitos	2	Documento de Requisitos	
		1.1.3. Revisar la arquitectura del producto software	2	Documento de arquitectura del software	
	1.2.	1.2.1. Revisar el protocolo de documentación de la organización	3	Protocolo de documentación Reporte de situación de la implantación	Reporte de situación de la implantación
		1.2.2. Revisar el protocolo de comunicación de la organización	3	Protocolo de comunicación Reporte de situación de la implantación	Reporte de situación de la implantación
		1.2.3. Revisar el protocolo de la gestión de la configuración	2	Protocolo de gestión de la configuración Reporte de situación de la implantación	Reporte de situación de la implantación
	1.3.	1.3.1 Determinar los recursos humanos existentes en la organización	2	Reporte de situación de la implantación Listado de usuarios Listado de recursos humanos de sistemas (técnicos)	Reporte de recursos humanos internos Reporte de situación de la implantación
		1.3.2.Determinar los recursos tecnológicos existentes en la organización	2	Reporte de situación de la implantación Documento de infraestructura	Reporte de recursos tecnológicos Reporte de situación de la implantación
	1.4.	1.4.1 Identificar los riesgos	2	Reporte de situación de la implantación	Reporte de analisis de riesgos Reporte de situación de la implantación
		1.4.2.Realizar el estudio de factibilidad	2	Reporte de situación de la implantación	Reporte de factibilidad Reporte de situación de la implantación

Tabla 2. Fase 2: Planificación, actividades, tareas, nivel de capacidad, entradas y salidas.

Fase	Actividades	Tareas	NC	Entradas	Salidas
2. Planificación	2.1.	2.1.1.Definir las actividades asociadas a la implantación	2	Plan del proyecto Reporte de situación de la implantación	Plan de implantación Documento de carga de datos Guía de instalación
		2.1.2. Definir los hitos entregables	2	Plan del proyecto Plan de implantación Reporte de situación de la implantación	Plan de implantación (con hitos entregables)
		2.1.3. Definir las actividades de verificación y validación	2	Plan del proyecto Reporte de situación de la implantación Plan de implantación	Plan de implantación (con pruebas de Documento de pruebas
		2.1.4. Planificar actividades tercerizadas	2	Plan del proyecto Reporte de situación de la implantación Plan de implantación	Plan de implantación: actividades tercerizadas Acuerdo con proveedores
	2.2.	2.2.1.Planificar los recursos humanos	2	Reporte de situación de la implantación Reporte de recursos humanos internos Plan de implantación	Reporte de recursos humanos requeridos Plan de implantación: asignación de recursos
		2.2.2. Planificar los recursos tecnológicos	2	Reporte de recursos tecnológicos Plan de implantación	Reporte de recursos tecnológicos requeridos Plan de implantación
		2.2.3. Planificar la capacitación de los recursos humanos	2	Reporte de recursos humanos requeridos Plan de implantación Reporte de situación de la implantación	Plan de implantacion: capacitación equipo técnico Plan de implantacion: capacitación de usuarios finales Plan de implantación
		2.2.4. Planificar la adecuación de recursos tecnológicos del ambiente de instalación	2	Reporte de situación de la implantación Diagrama de arquitectura del software Plan de implantación	Plan de implantación: instalación Diagrama de infraestructura Diagrama de despliegue
	2.3.	2.3.1.Determinar las métricas a emplear	2	Plan del proyecto Plan de implantación Reporte de situación de la implantación	Reporte de métricas
		2.3.2. Determinar los elementos de gestión de configuración	2	Plan del proyecto Plan de implantación Reporte de situación de la implantación	Reporte de elementos de gestión de configuración
		2.3.3. Planificar la gestión de riesgos	3	Plan de implantación Reporte del análisis de riesgos	Plan de implantación: gestión de riesgos Plan de implantación: plan de mitigación de riesgos
	2.4.	2.4.1.Estimar los tiempos de las actividades de la implantación	2	Plan de implantación	Plan de implantación: calendario
		2.4.2. Estimar los costos de las actividades de la implantación	2	Plan de implantación: calendario	Plan de implantación Costo estimado

La **Fase 3: Ejecución**, se compone de tres actividades: *Capacitación de los recursos humanos,* cuyo propósito consiste en el diseño y la realización de la capacitación para la comunidad usuaria y el equipo técnico; *Preparación del lugar e instalación,* cuyo propósito consiste en la preparación de la infraestructura, la instalación del producto software y la carga de datos; *Ejecución de las pruebas,* cuyo propósito consiste en la realización de las pruebas del ambiente de instalación, de la instalación y las de aceptación. En la Tabla 3, se presenta la **Fase 3: Ejecución** con las tareas de cada actividad, el nivel de capacidad determinado para la tarea y las entradas y las salidas de cada una de las tareas.

Tabla 3. Fase 3: Ejecución, actividades, tareas nivel de capacidad, entradas y salidas.

Fase	Actividades	Tareas	NC	Entradas	Salidas
3. Ejecución	3.1.	3.1.1 Diseñar la capacitación de los usuarios finales	1	Plan de implantación: capacitación de usuarios finales	Materiales para la capacitación. Reporte de la capacitación de los usuarios finales
		3.1.2. Diseñar la capacitación de los técnicos	2	Plan de implantación: capacitación del equipo técnico	Materiales para la capacitación. Reporte de la capacitación de los técnicos
		3.1.3. Capacitar a los usuarios finales	1	Materiales para la capacitación Reporte de la capacitación de los usuarios finales	Reporte de asistencia de los usuarios finales
		3.1.4. Capacitar a los técnicos	2	Materiales para la capacitación Reporte de la capacitación de los técnicos	Reporte de asistencia de los técnicos
	3.2.	3.2.1Preparar el ambiente de instalación	1	Plan de implantación: instalacion Diagrama de infraestructura	Reporte del ambiente de instalación
		3.2.2.. Instalar el software	1	Documento de arquitectura del software Plan de implantación: instalacion Diagrama de despliegue Guía de instalación	Reporte de la instalación del producto software
		3.2.3.Realizar la carga de datos	1	Set de datos Documento de carga de datos Plan de implantación: instalacion	Reporte de la carga de datos
	3.3.	3.3.1Realizar las pruebas de implantación	2	Plan de implantación: pruebas de instalación Documento de pruebas	Reporte de Resultados de pruebas de instalación
		3.3.2. Realizar las pruebas de aceptación	2	Plan de implantación: pruebas de aceptación Documento de pruebas	Reporte de Resultados de pruebas de aceptación

La **Fase 4: Seguimiento y Control**, se compone de cinco actividades: *Seguimiento y Control de la capacitación,* cuyo objetivo consiste en la realización del seguimiento y control de las actividades de capacitación como así también de realizar los ajustes necesarios; *Seguimiento y Control de la preparación del lugar y de la instalación,* cuyo objetivo consiste el monitoreo y ajuste de las actividades de preparación del lugar y de la instalación; *Seguimiento y Control de las actividades de pruebas,* cuyo objetivo consiste en la realización del seguimiento y control de las actividades de pruebas y de los ajustes necesarios; *Seguimiento y Control de las actividades de soporte,* cuyo objetivo consiste la monitoreo de la gestión de configuración, el cumplimiento de las métricas y del cumplimiento de la mitigación de los riesgos si se presentaron; *Seguimiento y Control de las actividades tercerizadas,* cuyo objetivo consiste en monitorear las actividades que realizan los proveedores como así también el cumplimiento de los acuerdos establecidos. En la Tabla 4, se presenta la **Fase 4: Seguimiento y Control** con las tareas de cada actividad, el nivel de capacidad determinado para la tarea, las entradas y las salidas de cada una de las tareas.

La **Fase 5: Cierre**, se compone de seis actividades: *Cerrar las actividades de capacitación,* cuyo propósito consiste en la registración del cierre formal de las actividades de capacitación; *Cerrar las actividades de instalación* cuyo propósito consiste en la registración del cierre de las actividades de instalación; *Cerrar las actividades de pruebas,* cuyo propósito consiste en la registración del cierre formal de las actividades de las pruebas de instalación y pruebas de aceptación*; Cerrar las actividades de soporte,* cuyo propósito consiste en registrar la finalización de las actividades de soporte; *Cerrar las actividades tercerizadas,* cuyo propósito consiste en registrar los resultados de las actividades realizadas por los proveedores y el grado de cumplimento de los acuerdos establecidos; *Actualizar bitácora de conocimiento organizacional,* cuyo propósito consiste en la registración de los resultados del proyecto para capitalizar el conocimiento de la organización. En la Tabla 5, se presenta la **Fase 5: Seguimiento y Control** con las tareas de cada actividad, el nivel de capacidad determinado para la tarea, las entradas y las salidas de cada una de las tareas.

Tabla 4. Fase 4: Seguimiento y control, actividades, tareas nivel de capacidad, E/S.

Fase	Actividades	Tareas	NC	Entradas	Salidas
4. Seguimiento y Control	4.1.	4.1.1. Realizar el seguimiento de la capacitación de los usuarios finales	2	Plan de implantación: capacitación de usuarios finales	Plan de implantación: capacitación de usuarios finales
				Reporte de asistencia de los usuarios finales	Reporte de acciones correctivas (capacitación)
		4.1.2. Realizar el seguimiento de la capacitación del equipo técnico	2	Plan de implantacion: capacitación del equipo técnico	Plan de implantacion: capacitación del equipo técnico
				Reporte de asistencia del equipo técnico	Reporte de acciones correctivas (capacitación)
	4.2.	4.2.1. Monitorear la preparación para la instalación	2	Plan de implantación: instalación	Reporte del ambiente de instalación
				Reporte del ambiente de instalación	Reporte de acciones correctivas
		4.2.2. Realizar el seguimiento de la instalación del software	2	Plan de implantación: instalación	Plan de implantación: instalación
				Reporte de la instalación del producto software	Reporte de la instalación del producto software
					Reporte de acciones correctivas
		4.2.4.Realizar el seguimiento de la carga de datos	2	Plan de Implantación: instalación	Plan de Implantación: instalación
				Reporte de la carga de datos	Reporte de la carga de datos
					Reporte de acciones correctivas
	4.3.	4.3.1. Realizar el seguimiento de las pruebas de instalación del software	2	Documento de pruebas	Plan de implantación: pruebas de instalación
				Plan de implantación: pruebas de instalación	Reporte de acciones correctivas
				Reporte de Resultados de pruebas de Instalación	Reporte de pruebas
		4.3.2. Realizar el seguimiento de las pruebas de aceptación	2	Documento de pruebas	Plan de implantación: pruebas de aceptación
				Plan de implantación: pruebas de aceptación	Reporte de acciones correctivas
				Reporte de Resultados de pruebas de Aceptación	Reporte de pruebas
	4.4.	4.3.1. Realizar el seguimiento de la gestión de configuración	2	Plan de implantación	Reporte de resultados de la gestión de configuración
				Reporte de elementos de gestión de configuración	
		4.3.2. Revisar las métricas	2	Plan de implantación	Reporte de resultados de las métricas
				Reporte de métricas	
		4.3.3. Revisar los riesgos	3	Plan de implantación: gestión de riesgos	Reporte de resultados de la revisión de riesgos
				Plan de implantación: plan de mitigación de riesgos	
	4.5.	4.5.1. Realizar el monitoreo de las actividades tercerizadas	2	Plan de implantación: actividades tercerizadas	Reporte de resultados de las actividades tercerizadas
					Plan de implantación: actividades tercerizadas
		4.5.2. Realizar el monitoreo del acuerdo con los proveedores	2	Plan de implantación: actividades tercerizadas	Reporte de resultados del acuerdo con los proveedores
				Acuerdo con proveedores	

3 Caso de estudio

Se presenta un caso de estudio para realizar una validación inicial del modelo propuesto. Se consideraron las tareas correspondientes a un nivel de capacidad 2 = gestionado de cada una de las fases de MoProIMP, debido a que la organización se encuentra en un nivel de capacidad equivalente. El caso de estudio corresponde a la implantación de un sistema de gestión de agencias de publicidad para Latinoamérica. La empresa en la que se desarrolló la experimentación del modelo, es una empresa multinacional que se encuentra en la República Argentina y ofrece servicios de consultoría. El sistema de gestión es un ERP (en inglés, Enterprise Resource Planning) desarrollado a medida para la gestión de empresas de publicidad; cuenta con los siguientes módulos: clientes, proveedores, contabilidad, tesorería, administración y parámetros (módulo donde se crean entidades maestras y parametrización del sistema), rendiciones de gastos y seguridad. Los escenarios posibles para el proceso de implantación pueden ser: 1) Empresa nueva en un país nuevo (implica la apertura de una sucursal del grupo de agencias en un país nuevo y esto conlleva a la configuración de todos los aspectos impositivos del país), 2) Empresa nueva en un país existente (se cuenta con experiencia de los aspectos impositivos del país) o 3) Empresa nueva en instalación existente (implica la replicación de una instalación, se utiliza el mismo servidor de una sucursal existente

sin compartir la base de datos). Estos escenarios diferentes indican distintos niveles de complejidad en el proceso de implantación. El primer escenario es el más complejo y es sobre el cual se aplicó el modelo propuesto.

Como resultado de la evaluación de la aplicación del modelo construido, se generó un informe técnico pormenorizado. A continuación, se resumen las principales fortalezas y debilidades que se han presentado durante la ejecución de la implantación del sistema de gestión de agencias de publicidad.

Tabla 5. Fase 5: Cierre, actividades, actividades, tareas nivel de capacidad, entradas y salidas.

Fase	Actividades	Tareas	NC	Entradas	Salidas
5. Cierre	5.1.	5.1.1Registrar finalización de la capacitación de los usuarios finales	2	Reporte de acciones correctivas (capacitación) Plan de implantación: capacitación de usuarios finales	Reporte de finalización de la capacitación de los usuarios finales
		5.1.2. Registrar finalización de la capacitación del equipo técnico	2	Reporte de acciones correctivas (capacitación) Plan de implantación: capacitación del equipo técnico	Reporte de finalización de la capacitación del equipo técnico
	5.2.	5.2.1. Registrar la finalización de la preparación del ambiente de instalación	2	Plan de Implantación: instalación Reporte del ambiente de instalación Reporte de acciones correctivas	Reporte del ambiente de instalación
		5.2.2. Registrar la finalización de la instalación	2	Plan de Implantación: instalación Reporte de la instalación del producto Reporte de acciones correctivas	Reporte de la instalación del producto software
		5.2.3. Registrar la finalización de la carga de datos	2	Plan de Implantación: instalación Reporte de la carga de datos Reporte de acciones correctivas	Reporte de la carga de datos
	5.3.	5.3.1. Registrar finalización de las actividades de pruebas de instalación	2	Plan de implantación: pruebas de instalación// Reporte de pruebas	Reporte de finalización de las actividades de pruebas de instalación
		5.3.2. Registrar finalización de las actividades de pruebas de aceptación	2	Plan de implantación: pruebas de aceptación//Reporte de pruebas	Reporte de finalización de las actividades de pruebas de aceptación
	5.4.	5.4.1. Registrar finalización de las actividades de gestión de configuración	2	Reporte de resultados de gestión de configuración Plan de implantación	Reporte de la finalización de las actividades de gestión de configuración
		5.4.2. Registrar finalización de las actividades de medición	2	Plan de implantación Reporte de resultados de las métricas	Reporte de la finalización de las actividades de medición
		5.4.3. Registrar finalización de la gestión de riesgos	3	Reporte de los resultados de la revisión de riesgos	Reporte de la gestión de riesgos
	5.5.	5.5.1. Registrar finalización de las actividades tercerizadas	2	Reporte de los resultados de las actividades tercerizadas Plan de implantación: actividades	Reporte de finalización de las actividades tercerizadas
		5.5.2. Registrar finalización de acuerdos con proveedores	2	Reporte de los resultados del acuerdo con los proveedores	Reporte de finalización de acuerdos con proveedores
	5.6.	5.6.1. Actualizar los resultados de las actividades de capacitación	3	Reporte de la finalización de la capacitación de los usuarios finales Reporte de la finalización de la capacitación del equipo técnico	Bitácora de lecciones aprendidas: Registro de los resultados de
		5.6.2. Actualizar los resultados de las actividades de instalación	3	Reporte del ambiente de instalación Reporte de la instalación del producto Reporte de la carga de datos	Bitácora de lecciones aprendidas: Registro de los resultados de la
		5.6.3. Actualizar los resultados de las pruebas de verificación	3	Reporte de la finalización de las actividades de pruebas de instalación	Bitácora de lecciones aprendidas: Registro de resultados de las pruebas de
		5.6.4. Actualizar los resultados de las pruebas de validación	3	Reporte de la finalización de las actividades de pruebas de aceptación	Bitácora de lecciones aprendidas: Registro de resultados de las pruebas de
		5.6.5. Actualizar los resultados de las actividades de gestión de configuración	3	Reporte de la finalización de las actividades de gestión de configuración	Bitácora de lecciones aprendidas: Registro de resultados de la gestión de configuración
		5.6.6. Actualizar los resultados de las mediciones	3	Reporte de la finalización de las actividades de medición	Bitácora de lecciones aprendidas: Registro de resultados de las métricas
		5.6.7. Actualizar los resultados de los riesgos	3	Reporte de la gestión de riesgos	Bitácora de lecciones aprendidas: Registro de resultados de los riesgos
		5.6.8. Actualizar los resultados de las actividades tercerizadas	3	Reporte de la finalización de las actividades tercerizadas	Bitácora de lecciones aprendidas: Registro de resultados de las actividades
		5.6.9. Actualizar los resultados de los acuerdos con los	3	Reporte de la finalización de acuerdos con proveedores	Bitácora de lecciones aprendidas: Registro de resultados de las actividades

En la **Fase 1: Iniciación,** se destacan los siguientes puntos favorables; la confección *del reporte de situación de la implantación* que sintetiza todos los puntos

claves para el proceso: alcances del proyecto, estrategia de implantación, uso de una wiki como espacio formal para compartir documentación de manera interna y con el cliente. Respecto a los recursos humanos, se revisó el listado de los *usuarios clave* que son usuarios del cliente que participaron por parte de la empresa en definiciones de diseño y en la aprobación de los productos terminados entregados. La comunidad usuaria estaba conformada por treinta usuarios. Se definieron los responsables del proceso de implantación, los representantes del cliente y del equipo de la consultora, quienes están en condiciones de comprometer a sus respectivas organizaciones en la aprobación del proyecto, y en la negociación de las fechas y alcances del mismo. Se le entregó al cliente, *el reporte de recursos tecnológicos* con las especificaciones necesarias para la implantación de la solución, las cuales han sido definidas por la consultora. El cliente se comprometió a adquirir e instalar el entorno de hardware y software de base necesario para ejecutar la aplicación, en la medida que el equipamiento que poseían no permitiese la instalación del producto software. Se logró determinar la factibilidad del proceso de implantación; se definieron los siguientes riesgos posibles: 1) fallas en el ambiente de instalación, 2) fallas en la instalación del producto software, 3) fallas en la carga inicial de datos y 4) fallas en el ambiente de capacitación y sus estrategias para mitigarlos. Dado que el escenario de implantación, era una empresa nueva en un país nuevo, se definió llevar a cabo una implantación modular, ya que no existía ningún sistema anterior. La aplicación del modelo ha resultado satisfactoria, dado que todas las tareas se han desarrollado sin inconvenientes. Las únicas tareas que no se han desarrollado son las correspondientes al *nivel 3=definido,* dado que la consultora cuenta con procesos equivalentes a un *nivel 2=gestionado;* esto se puede considerar como un aspecto de mejora no solo para el proceso en estudio, sino también en el resto de los procesos de la organización.

En la **Fase 2: Planificación**, se confeccionó el *plan del proceso de implantación*, en el cual se consideraron las tareas, sus duraciones, los hitos de entregas y las pruebas a realizar (instalación y aceptación). Se definieron, además, las responsabilidades para el cliente y la consultora. Del análisis realizado para este proyecto, se decidió la realización de las actividades de manera interna, sin contratación de proveedores. Se logró conformar el equipo, dos analistas-capacitadores quienes participarán en el proyecto desde su inicio, un técnico-implantador, un implantador experto y el líder de proyecto que ocupa el rol de coordinador de la implantación. En esta instancia se definió un usuario clave de soporte quién será responsable del mantenimiento una vez que se haya considerado finalizada la implantación. Se generó la especificación de la infraestructura que deberá preparar el cliente para el ambiente de instalación. Se planificó la capacitación para la comunidad usuaria y la capacitación para el responsable técnico del cliente, encargado del futuro del mantenimiento. La capacitación a la comunidad usuaria, se desarrolló en las instalaciones del cliente con cursos de dos semanas de duración y de media jornada. Se definieron como métricas el costo y el tiempo. Se definió la wiki como espacio para el almacenamiento de los versionados de los artefactos: manuales de usuario, documentación de la preparación del ambiente de instalación, de la instalación, de la carga de datos; y también se definirán los usuarios de la Wiki con sus correspondientes niveles de seguridad. Se calendarizaron las actividades del plan. Se calculó el costo en horas de los recursos humanos requeridos, la traducción de horas a un valor monetario es realizada por otra Área de la consultora.

Como resultado de la fase 2, se encontraron dos debilidades. En primer lugar, la trasferencia de la responsabilidad de la preparación del ambiente de instalación al cliente, sin la supervisión de la consultora. Esto es un punto crítico para el momento en que debe realizarse la instalación del producto software. En segundo lugar, las métricas escasas que la consultora emplea en sus procesos.

En la **Fase 3: Ejecución,** se desarrollaron los materiales para el curso de capacitación, los cuales se almacenaron en la wiki de la consultora. Se identificaron los dos grupos de usuarios a capacitar. Se elaboró un instructivo de carga de datos y parametrización del sistema para ser utilizado en el curso, se definió un procedimiento para la realización del mantenimiento del sistema de gestión para el cliente. Se realizó la capacitación en las instalaciones del cliente. Los capacitadores ayudaron a verificar el procedimiento de mantenimiento. Se realizó la instalación del sistema. Durante la misma, se presentaron algunos inconvenientes, lo que generó una consulta remota con el implantador experto de la consultora. Durante la capacitación, se realizó la carga inicial de datos por tratarse de un sistema nuevo. Se realizó la prueba de aceptación con el usuario.

El mayor inconveniente de la fase 3, ha sido que la preparación del ambiente de instalación haya sido realizada por el cliente y que no fue adecuadamente supervisada por la consultora. El resto de las actividades/tareas propuestas en MoProIMP se desarrollaron de manera satisfactoria.

En la **Fase 4: Seguimiento y Control,** el coordinador de la implantación realizó el seguimiento de las tareas, registrando sistemáticamente el monitoreo de las actividades. Las actividades de capacitación resultaron de acuerdo al plan. Se registró adecuadamente el desvío ocurrido y las acciones correctivas propuestas para la tarea preparación del lugar de instalación. Las tareas de instalación, carga de datos al sistema ERP y pruebas se han desarrollado de acuerdo a la planificación prevista. Se observó que la wiki ha sido gestionada correctamente como repositorio de versionados, dado que los recursos humanos de la consultora y del cliente se encuentran familiarizados a esta modalidad de trabajo. El único inconveniente que se ha presentado en esta fase, ha sido un mínimo retraso en el tiempo necesario. Este retraso (algunas horas) pudo ser subsanado sin afectar las tareas procedentes. El indicador tiempo se cumplió satisfactoriamente, no obstante el indicador horas/recurso humano se desvió, dado que el implantador experto brindó asistencia adicional en la solución de la preparación del ambiente de instalación.

En la **Fase 5: Cierre,** se realizó el cierre formal de las actividades de capacitación, de preparación del lugar, de instalación del sistema ERP, de carga de datos, de pruebas, de la gestión del versionado de artefactos y de los indicadores utilizados en el proceso. Una debilidad encontrada, es la ausencia de un espacio institucional para compartir las prácticas realizadas en este proyecto de implantación del sistema ERP.

4 Conclusiones y futuros trabajos

En este trabajo, se ha logrado poner en práctica el modelo propuesto en una implantación de un sistema de gestión de agencias de publicidad para Latinoamérica,

utilizado por una empresa multinacional que ofrece servicios de consultoría como otra unidad de negocio en el ámbito de la República Argentina.

Se ha logrado aplicar el modelo en el nivel de capacidad 2 (gestionado) definido en esta propuesta y esto ha permitido confirmar algunas debilidades que siguen persistiendo en el proceso de implantación en la industria nacional; entre las que podemos mencionar métricas específicas, la gestión de riesgos, espacios institucionales para compartir conocimiento, roles con las competencias específicas para la realización del proceso.

MoProIMP ha permitido llevar adelante la gestión del proceso de implantación dado que su arquitectura de Fases, actividades, tareas y las dependencias entre las mismas, contribuye a que el modelo sea autogestionado.

Como trabajo futuro se identifica: (a) la necesidad de construcción de los artefactos de entrada y salida para cada una de las tareas del modelo empleando los lineamientos propuestos por algún estándar y (b) continuar con la experimentación del modelo en otras empresas de la industria nacional.

Referencias

1. Panizzi, M., Bertone R., Hossian A. Proceso de Implantación de Sistemas Informáticos – Identificación de vacancias en Metodologías Usuales. Libro de Actas de la V Conferencia Iberoamericana de Computación Aplicada CIACA 2017. Pag.207 -215. ISBN 978-989-8533-70-8. Vilamoura, Algarve, Portugal. (2017).
2. Carvajal Riola, J. Metodologías Ágiles: Herramientas y modelo de desarrollo para aplicaciones JAVA EE como metodología empresarial. Tesis Final de Máster en Tecnologías de la Información - UPC. Barcelona. (2008). https://upcommons.upc.edu/bitstream/handle/2099.1/5608/500%2015.pdf?sequence=1. Página vigente al 05/05/2018.
3. Panizzi, M., Hossian, A., García-Martínez, R. Implantación de Sistemas: Estudio Comparativo e Identificación de Vacancias en Metodologías Usuales. Libro de Actas del XXII Congreso Argentino de Ciencias de la Computación (CACIC 2017). Pág. 546-555. ISBN 978-987-733-072-4. Universidad Nacional de San Luis. (2016).
4. Iglesias Fernández, C. Definición de una Metodología para el desarrollo de sistemas Multiagente. Tesis Doctoral. Universidad Politécnica de Madrid. Madrid. (1998) http://www.upv.es/sma/teoria/agentes/tesiscif.pdf. Página vigente al 08/05/2018.
5. RAE, Diccionario de la Lengua Española, Real Academia Española. http://dle.rae.es/ Página vigente al 27/06/2017.
6. IEEE 610. IEEE Standard Glossary of Software Engineering Terminology. IEEE Std 610.12-1990., 1990.
7. Maya E., Métodos y técnicas de investigación, México. (2014)
8. A Guide to the Project Management Body of Knowledge. (PMBOK® Guide) – Fifth Edition. 2013 Project Management Institute. ISBN 978-1-935589-67-9.
9. CMMI® para Desarrollo, Versión 1.3 Mejora de los procesos para el desarrollo de mejores productos y servicios. TECHNICAL REPORT. Editorial Universitaria Ramón Areces. (2010).
10. PAe, Métrica versión.3. Portal de Administración Electrónica. Gobierno de España. (2001).
11. Marca David A. and McGowan, Clement L. "SADT: Structured Analysis and Design Techniques". McGraw-Hill, New York, NY. (1988).

Enfoque Holístico de Calidad: Aplicación de una Estrategia de Selección de Alternativas

Celina Vicenz, Nerea Waiman, María Fernanda Papa,
Pablo Becker, and Luis Olsina

GIDIS_Web, Facultad de Ingeniería, Universidad Nacional de La Pampa
General Pico, La Pampa 6360, Argentina
{celinavicenz90,nereawaiman}@gmail.com
{pmfer,beckerp,olsinal}@ing.unlpam.edu.ar

Resumen Hoy las empresas de software poseen oportunidades y desafíos importantes, ya que si bien pueden expandirse rápidamente al mercado mundial, la fuerte competencia complica su permanencia. Por lo que deberían diferenciarse de sus competidores ofreciendo productos innovadores, de alta calidad y a costos competitivos. En esta dirección, este trabajo muestra cómo aplicar una estrategia de selección de alternativas teniendo en cuenta las vistas de calidad y costo. La estrategia forma parte del Enfoque Holístico de Calidad Multipropósito y Multinivel, el cual permite la alineación de objetivos a distintos niveles organizacionales y la utilización de una familia de estrategias de medición, evaluación y/o cambio que brindan información útil para tomar decisiones.

Keywords: Estrategia, Calidad-Costo, Selección de Alternativas

1. Introducción

Las organizaciones de software se desarrollan en entornos globales y competitivos los cuales exigen un gran esfuerzo para mantenerse en el mercado. Una forma de lograrlo es ofreciendo productos innovadores de alta calidad a costo competitivo, ya que el mercado requiere productos o servicios con altos índices de desempeño costo-calidad. Consecuentemente, la organización debería optimizar sus recursos y procesos no sólo en el aspecto de calidad sino también de costos intentando tomar decisiones con un mínimo grado de incertidumbre.

La toma de decisiones se define como "la selección de un curso de acción entre varias alternativas" y suele considerarse como parte integral de la gestión de una organización [1]. En este sentido, las decisiones surgen en todos los niveles organizacionales como operativo, táctico y estratégico. El objetivo del análisis de decisiones es mejorar la calidad de las mismas, caracterizando apropiadamente los componentes del proceso de decisión. Según [2], es necesario contar con un marco conceptual apropiado, generar alternativas creativas y factibles, contar con información relevante y confiable, tener claridad de criterios de decisión, utilizar modelos adecuados y tener un compromiso con la acción.

El proceso de decisión posee tres grandes actividades: 1) Establecer premisas; 2) Identificar y evaluar alternativas teniendo en claro las metas de la organización; y 3) Elegir la alternativa que contribuya en mayor medida al logro de dichas metas. La calidad de dicho proceso está ligada a la calidad de la información con la que se cuenta y la calidad del análisis de los datos disponibles.

Si bien las decisiones a niveles jerárquicos pueden ser críticas y desicivas para la organización, las decisiones de los niveles inferiores no deben ser minimizadas y deberían estar alineadas a las tomadas por los niveles superiores. Más allá de la importancia de la toma de decisiones en la organización muchas veces la selección de una alternativa no se evalúa utilizando métodos, herramientas y procedimientos apropiados [3]. Este suele ser el caso de las empresas de software incipientes donde el personal debe desempeñar muchas funciones simultáneamente. Aunque incipientes, estas organizaciones suelen tener en cuenta la evaluación de calidad de los productos o servicios ofrecidos dado que es un factor de diferenciación en el mercado. Esto las lleva a familiarizarse con enfoques de evaluación que ayuden a conocer de forma objetiva la calidad de sus productos/servicios, e incluso, compararlos con los provistos por sus competidores con la finalidad de generar ciclos de mejora continua.

En este sentido el Enfoque Holístico de Calidad Multinivel y Multipropósito (de ahora en más enfoque HCMyM) [4] es de utilidad, ya que permite establecer metas de necesidad de información de medición y evaluación (ME) vinculadas con metas de negocio a distintos niveles organizacionales. Adicionalmente, por medio de la utilización de estrategias integradas específicas brinda datos e información útil para la toma de decisiones. Dichas estrategias fueron diseñadas para dar soporte a las categorías de propósitos de evaluación [5,6]: Caracterizar y entender; Controlar y cambiar; y Adoptar alternativa. Siendo la última categoría la que nos interesa en este trabajo, específicamente teniendo en cuenta la selección de alternativas donde se involucran "vistas de calidad y costo".

La contribución de este trabajo es la presentación de un caso de estudio de la Estrategia GOCAMES *(Goal-Oriented Context-Aware Measurement, Evaluation and Selection)* perteneciente a la familia de estrategias del enfoque HCMyM, con el objetivo de ejemplificar su instanciación en un caso actual, sencillo y fácil de entender como es la compra de celulares (smartphones) corporativos. Este ejemplo se puede extrapolar a cuestiones más relevantes en la vida organizacional como pueden ser la selección de COTS, de personal capacitado, entre otros. El resto del artículo se estructura como sigue: La Sección 2 brinda los fundamentos de la estrategia GOCAMES utilizada para seleccionar una alternativa en el caso de estudio detallado en la Sección 3. La Sección 4 comenta los trabajos relacionados y la Sección 5 presenta las conclusiones y trabajos futuros.

2. Estrategia GOCAMES

GOCAMES es una estrategia de medición, evaluación y selección que forma parte de la familia de estrategias del enfoque HCMyM. Este enfoque está basado en tres principios: 1) la definición de metas de negocios y de necesidades de

información a diferentes niveles organizacionales; 2) la definición de diferentes propósitos de evaluación para las metas; y 3) la concepción de una familia de estrategias de medición, evaluación y cambio que ayuda al logro del propósito de la meta. GOCAMES da soporte a metas con el propósito de 'seleccionar alternativa' incluidas en la categoría de propósito 'adoptar alternativa' [5]. Es una estrategia integrada, orientada a metas, sensible al contexto y centrada en la necesidad de información organizacional. Se considera integrada porque define simultáneamente tres capacidades, a saber:

1. Un *proceso* definido formalmente en SPEM desde distintas perspectivas que indica, entre otras cuestiones, qué hacer y en qué orden. Garantizando de este modo que los resultados sean reproducibles y comparables (ver Fig. 1).
2. Un *marco metodológico* que indica cómo implementar cada actividad del proceso. Por ejemplo, a partir de la especificación de plantillas de métricas e indicadores propone métodos de medición y evaluación, entre otros.
3. Un *marco conceptual* apoyado en las ontologías de "medición y evaluación" y de "vistas de calidad". Mientras que la primera promueve la uniformidad y consistencia en los términos utilizados en las capacidades anteriores, permitiendo no sólo la especificación de los metadatos necesarios para que los resultados sean consistentes y no ambiguos, sino también, evitando imprecisiones al comunicarlos o analizarlos. La segunda define conceptos de vista de calidad y de costo [7], las cuales representan una asociación entre una categoría de entidad de alto nivel –por ejemplo, recurso, proceso, producto, etc.– y un foco de calidad y/o costo –calidad de recurso, calidad del proceso, calidad interna o calidad externa. Estos conceptos son necesarios, junto al propósito de la meta, para seleccionar de la familia de estrategias la estrategia más adecuada, teniendo en cuenta la cantidad y tipos de vistas.

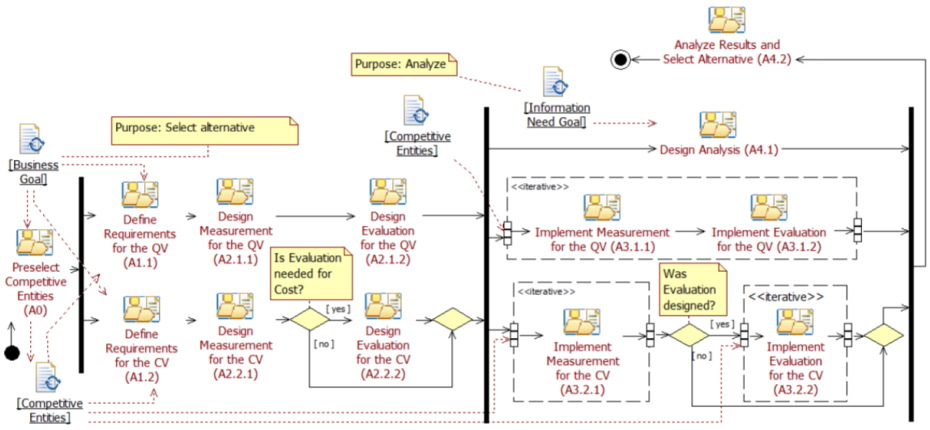


Figura 1. Proceso genérico de GOCAMES definido desde las perspectivas funcional y de comportamiento especificado en SPEM.

3. Aplicación de la Estrategia GOCAMES

El escenario se desarrolla dentro de una empresa de software que desea comprar celulares de última generación para sus empleados. Estos recursos, por un lado, permitirán la comunicación entre los integrantes de los distintos sectores de la empresa y, por el otro, se utilizarán como una herramienta de testeo en las aplicaciones web a desarrollar. Por lo que se parte de la siguiente meta de negocio: *Comparar un conjunto de características de calidad y costo de smartphones con el fin de seleccionar aquel que posea el índice de calidad-costo más adecuado.*

De ella se desprende que su propósito de evaluación es "adoptar una alternativa" teniendo en cuenta dos vistas (calidad y costo) de la entidad smartphone perteneciente a la supercategoría producto.

El proceso de la estrategia (recordar Fig. 1) comienza con la actividad *Preseleccionar entidades competitivas (A0)*, que requiere como entrada la "Meta de negocio" y produce como salida la selección de "Entidades competitivas". A partir del criterio de selección, que indica incluir las últimas versiones de celulares lanzados hasta la primera mitad de año 2017 de las marcas predominantes del mercado, se seleccionaron los modelos SAMSUNG S8, HUAWEI P9 y IPHONE 7. Mientras que Garbarino, Frávega, Musimundo y Megatone fueron elegidos considerando la disponibilidad de al menos uno de los productos seleccionados en sus sitios web.

Para una mejor organización del trabajo, las actividades se dividieron en las fases de diseño e implementación. Como se puede apreciar en la Fig. 1 éstas deben ser desdobladas y realizadas para las dos vistas intervinientes.

3.1. Diseño

Definir los requisitos no funcionales (A1) produce un documento que incluye la necesidad de información y el modelo de características y atributos a evaluar. El propósito de la necesidad de información es la selección del smartphone y el proveedor más adecuado respecto a su índice de desempeño calidad-costo.

La 1^{ra} columna de la Tabla 1 muestra un extracto del modelo de evaluación. La *Calidad* se define como el grado en el cual el producto satisface los requisitos de calidad implícitos y explícitos [8]. Las dimensiones consideradas son Eficiencia, Adecuación funcional y Compatibilidad [8] que se miden a partir de atributos que un usuario busca a la hora de comprar un celular [9]. El *Costo* en este estudio se representa como el gasto económico de comprar y trasladar el producto.

El atributo a ilustrar es *1.3.1.1.1 Interoperabilidad de comunicación con bandas de red (ICBR)* definido como el grado de interoperabilidad del celular con las diferentes bandas de red existentes, perteneciente a la subcaracterística Interoperabilidad con tecnologías previas (1.3.1.1).

Diseñar la medición y la evaluación para la vista de calidad (A2.1) es la siguiente actividad. En la fórmula (1) se muestra la métrica indirecta Tasa de interoperabilidad de comunicación con bandas de red (TICBR) utilizada para

cuantificar el atributo 1.3.1.1.1 del ejemplo.

$$TICBR = \left(\frac{CFD2G + CFD3G + CFD4G}{CMFD2G + CMFD3G + CMFD4G} \right) \quad (1)$$

donde $CFD2G$, $CFD3G$, $CFD4G$ representan la cantidad de frecuencias disponibles en el celular y $CMFD2G$, $CMFD3G$, $CMFD4G$ son la cantidad máxima de frecuencias en 2G, 3G y 4G, respectivamente.

Tabla 1. Fragmento del Modelo de Evaluación donde los atributos se resaltan en cursiva. El árbol de requisitos no funcionales completo se encuentra disponible en http://sfing.ing.unlpam.edu.ar/presentacion.pdf.

Características/Subcaracterísticas/Atributos	**P.**	**Op.**			
1.**Calidad**	0,5	C–	63,50	74,11	63,38
1.1.Eficiencia	0,3	CA	38,26	76,79	76,24
1.1.1.Capacidad	1,0	CA	38,26	76,79	76,24
1.1.1.1.Capacidad de memoria	0,2	C-	13,94	76,62	68,01
1.1.1.1.1.*Memoria RAM*	0,4		25,00	50,00	50,00
1.1.1.1.2.*Memoria interna*	0,4		75,00	100,00	75,00
1.1.1.1.3.*Memoria externa*	0,2		00,00	100,00	100,00
1.1.1.2.Capacidad de procesamiento	0,2	A	87,50	100,00	100,00
1.1.1.2.1.*Frecuencia del procesador*	0,5		100,00	100,00	100,00
1.1.1.2.2.*Núcleos disponibles*	0,5		75,00	100,00	100,00
1.1.1.3.Capacidad de la batería (...)	0,2	A	47,50	72,50	72,50
1.1.1.4.Capacidad de la cámara (...)	0,2	D-	61,42	62,20	71,86
1.1.1.5.Capacidad de la pantalla (...)	0,2	A	50,00	75,00	75,00
1.2.Adecuación funcional	0,33	A	59,40	73,40	47,60
1.2.1.Funcionalmente apropiado (...)	0,4	A	60,00	80,00	80,00
1.2.2.Funcionalmente completo (...)	0,6	A	59,00	69,00	26,00
1.3.Compatibilidad	0,33	A	100,00	72,09	67,44
1.3.1.Interoperabilidad (Interop.)	1,0	A	100,00	72,09	67,44
1.3.1.1.Interop. con tecnologías previas	1,0	A	100,00	72,09	67,44
1.3.1.1.1.*Interop. de comunicación con bandas de red*	0,4		100,00	72,09	67,44
1.3.1.1.2.*Interop. de comunicación con bluetooth*	0,6		80,00	100,00	80,00
2.**Costo**			**30298**	**21848**	**14999**
2.1.*Costo de compra*			29999	21499	14999
2.2.*Costo de envío*			299,00	349,00	0,00

La actividad *Diseñar la medición para la vista de costo (A2.2.1)* define la métrica indirecta Valor Económico Total de la Compra (VETC), resultado de sumar el precio de compra y de envío del smartphone.

En el *Diseño de la evaluación para la vista de calidad (A2.1.2)* se especifican los indicadores (elementales y derivados) junto con la escala y los criterios de decisión. La escala se define con un valor real, absoluto, cuya unidad es porcentaje. Los criterios de decisión se fijan con tres categorías, según lo que se espera de la calidad del dispositivo, a saber: valores entre [0 - 50) son considerados insatisfactorios; entre [50 - 75) marginales; y entre [75 - 100] satisfactorios.

La fórmula $NS_ICBR = TICBR * 100$ es parte de la definición del indicador elemental que transforma el valor medido del atributo ICBR a un valor de indicador elemental (NS_ICBR), interpretable con el criterio de decisión.

La definición de los indicadores derivados permiten conocer el nivel de satisfacción alcanzado por las características y subcaracterísticas presentes en el

modelo de evaluación (1^{ra} columna de la Tabla 1) a partir del modelo de agregación lógica de preferencias (LSP) [10], cuya función se especifica en (2).

$$ID_{(r)} = (P_1 * VI_1^r + P_2 * VI_2^r + \cdots + P_m * VI_m^r)^{1/r} \quad (2)$$

donde ID representa el valor de indicador derivado a calcular, VI_i son los valores de los indicadores del nivel inmediato inferior; $0 <= VI_i <= 100$, P_i son los pesos que cumplen con $(P_1 + P_2 + \cdots + P_m) = 1$; $P_i > 0$; $i = 1 \cdots m$ y r es un coeficiente conjuntivo/disyuntivo para el modelo LSP.

Además, para cada elemento del modelo de calidad se determinó un peso P (importancia relativa respecto a los elementos de su nivel) y para cada característica se adicionó un operador r (ver 2^{da} y 3^{ra} columnas de la Tabla 1).

Al *Diseñar la evaluación para la vista de costo (A2.2.2)* se definen la escala y los criterios de decisión para interpretar el costo agregado. Mientras que, la escala se especifica con un valor real y la unidad pesos, los criterios de decisión contienen cuatro categorías que determinan lo que el usuario esta dispuesto a pagar por un dispositivo, a saber: valores entre [1000 - 5000) son considerados baratos; entre [5000 - 15000) económicos; entre [15000 - 25000) razonables; y mayor a 25000 caros.

La actividad *Diseñar el análisis* (A4.1) permite anticipar la mejor forma de comunicar los resultados y las operaciones a realizar según escalas y unidades.

3.2. Implementación

Los datos obtenidos en *Implementar la medición de la vista de calidad (A3.1.1)* se recolectaron de las páginas oficiales de los smartphones y del sitio de comparación de especificaciones técnicas www.gsmarena.com. A modo de ejemplo, la Tabla 2 presenta los valores medidos necesarios para calcular la fórmula (1) y los resultados de su aplicación (negrita). Para la medición de los costos *(A3.2.1)* se sumó el costo de compra y envío (Tabla 3) publicados en los sitios de venta elegidos. La recolección se efectúo en la última quincena de diciembre de 2017.

Tabla 2. Valores medidos y calculados (negrita) para el atributo *1.3.1.1.1 Interoperabilidad de comunicación con bandas de red*

	CFD2G	CFD3G	CFD4G	CMFD2G	CMFD3G	CMFD4G	**TICBR**
SAMSUNG	4	5	22	10	8	25	**7,209**
	4	7	18	10	8	25	**6,744**
	10	8	25	10	8	25	**1,00**

Los valores calculados TICBR (Tabla 2) no respetan la escala establecida en los criterios de decisión, requiriendo una transformación para su interpretación. La actividad *Implementar la evaluación para la vista de calidad (A3.1.2)* es la encargada de dicha transformación, no siendo necesaria para la vista de costo.

Por ejemplo, a partir de la ecuación $TICBR*100$ se transforma el valor medido del atributo *1.3.1.1.1* para el smartphone Samsung S8 de 7,209 a 72,09 %. De este modo, se aplican los modelos elementales a todos los atributos medidos pertenecientes al modelo de evaluación para la vista de calidad.

Tabla 3. Valores del costo de los smartphones en los distintos proveedores, entre paréntesis el precio de envío.

	GARBARINO	FRÁVEGA	MUSIMUNDO	MEGATONE
SAMSUNG	$21499+($299)	$21499+($349)	-	-
HUAWEI	$14999+($299)	$14999+($349)	$14999+($0)	$15499+($249)
	$29999+($299)	-	-	-

Luego, a partir de los valores calculados de los indicadores elementales y aplicando la fórmula (2) se obtiene el nivel de satisfacción en el desempeño de las subcaracterísticas y características (ver últimas columnas de la Tabla 1).

En la actividad *Analizar los resultados y seleccionar alternativa (A4.2)* se informaron e interpretaron los resultados obtenidos siguiendo lo diseñado en la actividad *(A4.1)*. La Fig. 2 expone que el smartphone Samsung S8 alcanza un valor de 74,11 % de calidad, siendo este la mejor alternativa a seleccionar respecto a la calidad. Por el contrario, el Huawei P9 presenta el valor más bajo (63,38 %) como consecuencia de valores insatisfactorios y marginales en las características Adecuación funcional y Compatibilidad, respectivamente. Con respecto al costo, los resultados sugieren seleccionar el Huawei P9 porque si bien no existen diferencias de calidad con el Iphone 7, posee un mejor índice calidad-costo. Se puede concluir que la mejor alternativa es el Samsung S8 dado que posee valores más altos respecto a la calidad y su valor económico es razonable. El lugar de adquisición es Garbarino donde la diferencia, respecto de Frávega, radica en el costo de envío.

4. Trabajos Relacionados

En la Sección 1 se ha destacado el valor de contar con un Enfoque de Calidad que permita alinear objetivos a diferentes niveles organizacionales, ayude en la toma de decisiones y reúna los tres principios mencionados en la Sección 2.

Respecto al primer principio, existe bibliografía [11,12,13] que indica la importancia de la vinculación de los objetivos de medición con sus metas a diferentes niveles organizacionales ya que ayuda a la toma de decisiones. Esto es porque las metas de medición y evaluación pueden proporcionar información significativa para saber en qué medida se ha logrado un objetivo. En [14,15,16,17] se presentan distintas clasificaciones de propósitos de evaluación, el segundo principio requerido. Sin embargo, entre estos trabajos no existe un consenso en su categorización, ni son usados para seleccionar una estrategia que provea información

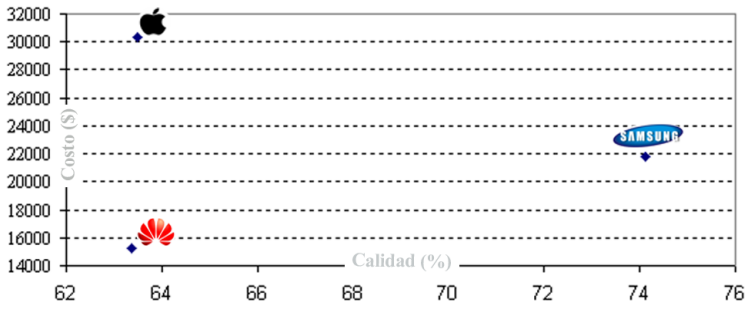


Figura 2. Índice calidad-costo para cada uno de los smartphone preseleccionados y el proveedor Garbarino.

para la toma de decisiones. Por último, en [18] se realiza una revisión sistemática de literatura donde surge que existen pocas estrategias de ME documentadas que integren simultáneamente las especificaciones de proceso y método, con la capacidad de la base conceptual. Los enfoques considerados relevantes son:

GQM *(Goal Question Metric)* [19] sostiene que la medición es un mecanismo para obtener información señalando la importancia de comprender qué datos recolectar partiendo de qué se quiere saber (meta). Su limitación es que no explicita el alineamiento entre los objetivos de medición y las metas de negocio [20,21]. GQM$^+$Strategies [11], mejora GQM, utilizado estrategias para alinear objetivos de negocio a diferentes niveles. Sin embargo este enfoque carece de especificaciones de proceso formales para las diferentes estrategias y no menciona propósitos de evaluación ni vistas de calidad/costo [7].

Goal-Driven Measurement describe un proceso general para la definición de objetivos de medición destinados a ayudar a cumplir los objetivos de la organización. Dicho proceso comienza con el establecimiento de objetivos de negocio a un nivel organizacional, de los cuales surgen preguntas que permiten identificar información cuantitativa de interés para los integrantes de la organización [22]. Utilizando el modelo de GQM identifica las entidades y atributos a medir. Este enfoque, si bien utiliza la terminología y plantillas de GQM, no explicita la definición de conceptos tales como objetivo de negocios, propósito de evaluación, etc., ni tampoco especifica estrategias de ME asociadas a propósitos de evaluación y vistas de calidad/costo.

GQ(I)M *(Goal-Question-Indicator-Measurement)* [12] entrelaza los aspectos de Goal-Driven Measurement [13] con el paradigma de Balanced Scorecard [23] descomponiendo objetivos estratégicos en sub-objetivos para cada uno de los cuadrantes (Financiero, Enfoque en el cliente, Procesos internos, y Aprendizaje y crecimiento). Sin embargo, no se apoya en el uso de estrategias integradas para cumplir con los propósitos de las metas organizacionales a partir de objetivos de necesidad de información de ME.

5. Conclusiones y Trabajos Futuros

Este trabajo documenta la instanciación de la estrategia GOCAMES con el propósito de seleccionar la mejor alternativa de compra de smartphones, teniendo en cuenta el índice de calidad-costo, según las necesidades específicas de una empresa de software (Sección 3). Si bien la ejemplificación desarrollada es un caso sencillo, fácil de entender y actual donde sus resultados permiten una toma de decisiones con mayor grado de información que si se hace ad hoc, se puede extrapolar a tomas de decisiones de mayor envergadura a cualquier nivel organizacional, como pueden ser la selección de COTS, de personal capacitado o de un nuevo local comercial, entre otros, donde la información requerida para la toma de decisiones debe ser más abarcativa y rigurosa. Como puede ser la inclusión de características de usabilidad del smartphone, para el ejemplo presentado.

Como GOCAMES (Sección 2) pertenece a la familia de estrategias integradas del Enfoque Holístico de Evaluación de Calidad Multipropósito y Multinivel, provee gran parte de los componentes que [2] considera necesarios al momento de tomar decisiones. El marco conceptual y el proceso formalmente definido propuesto por el enfoque y sus estrategias, hace que la organización sepa cómo llevar a cabo cada actividad garantizando información relevante y confiable. También, favorece al planteo de alternativas factibles, criterios de decisión claros y modelos de evaluación apropiados porque la alineación entre los objetivos a distintos niveles hace que todo esfuerzo operativo o táctico, esté dirigido hacia el logro de las metas estratégicas de la organización. Por lo tanto es un enfoque robusto, al alcance de las organizaciones que permite atacar el problema detectado por [3] de que la selección de una alternativa no siempre se evalúa utilizando métodos, herramientas y procedimientos apropiados y repetibles.

Como se discutió en la Sección 4 existen enfoques similares al utilizado pero ninguno de ellos posee los tres principios simultáneamente, a saber: alineación de las metas y objetivos a diferentes niveles organizacionales, categorías de propósitos para seleccionar la estrategia más adecuada a la necesidad de información y una familia de estrategias que ayuden al logro del propósito de la meta.

Como trabajo futuro se va especificar el patrón de diseño que agrupa a estrategias que consideran vistas de calidad/costo, como ya se ha realizando con el resto de los própositos [7].

Referencias

1. Harrison, E.F.: The Managerial Decision-Making Process. 5^{ta}edición, Boston (1999)
2. Matheson, D., Matheson, J.E.: The smart organization. Harvard Business School Press, Boston, MA (1998)
3. Castillo Hernández, M.J.: Toma de decisiones en las empresas: entre el arte y la técnica : metodologías, modelos y herramientas. Ediciones UniAndes, Colombia (2008)
4. Rivera, B., Becker, P., Papa, M.F., Olsina, L.: A Holistic Quality Evaluation, Selection and Improvement Approach driven by Multilevel Goals and Strategies. In. CLEI Electronic, vol. 19, no. 3, pp. 1–28 (2016)

5. Olsina, L., Becker, P.: Family of Strategies for different Evaluation Purposes. In: Proc. of the XX Conf. Iberoamericana en Soft. Engineering (CIbSE), CABA, Arg., Published by Curran Associates, pp. 221-234, ISBN 978-99967-839-2-014 (2017)
6. Tebes, G., Peppino, D., Becker, P., Papa, M.F., Rivera, M.B., Olsina, L.: Family of Evaluation Strategies: A Practical Case for Comparing and Adopting Strengths. Journal of Computer Science & Technology, vol. 18, no. 1, pp. 48–60, DOI: 10.24215/16666038.18.e06 (2018)
7. Rivera, M.B., Becker, P., Olsina, L.: Quality Views and Strategy Patterns for Evaluating and Improving Quality: Usability and User Experience Case Studies. Journal of Web Engineering, Rinton Press, USA, 15:(5&6), pp. 433-464 (2016)
8. ISO/IEC 25010, System and Software engineering–System and software Quality Requirements and Evaluation (SQuaRE)–System and software quality models (2011)
9. Rosero Barzola, C. Montalvo Ruilova, H.: La disonancia cognitiva como factor motivador en las decisiones del consumidor: caso smartphones. RETOS, Revista de Ciencias de la Administración y Economía (2015)
10. Dujmovic, J.J.: A Method for Evaluation and Selection of Complex Hardware and Software Systems. In: Proc. 22^{nd} International Conference for the Resource Management and Performance Evaluation of Enterprise, pp. 368-378 (1996)
11. Basili, V., Lindvall, M., Regardie, M., Seaman, C., Heidrich, J., Jurgen, M., Rombach, D., Trendowicz, A.: Linking Software Development and Business Strategy through Measurement, IEEE Computer, 43:(4), pp. 57-65 (2010)
12. Goethert, W., Fisher, M.: Deriving Enterprise-Based Measures Using the Balanced Scorecard and Goal-Driven Measurement Techniques, Software Engineering Measurement and Analysis Initiative, CMU/SEI-2003-TN-024 (2003)
13. Park, R., Goethert, W., Flora,c W.: Goal-Driven Software Measurement. A Guidebook, TR. CMU/SEI-96-HB-002, USA (1996)
14. Briand, L., Differding, Ch., Rombach, D.: Practical Guidelines for Measurement-based Process Improvement, Software Process Improvement and Practice Jnal., 2:(4), pp. 253-280 (1996)
15. Fenton, N., Pfleeger, S.,: Software Metrics: a Rigorous and Practical Approach, 2nd Ed., PWS Publishing Co. (1996)
16. PMBOK: A Guide to the Project Management Body of Knowledge, (2013)
17. Preece, J., Rombach, D.: A Taxonomy for Combining Software Engineering and Human-Computer Interaction Measurement Approaches: Towards A Common Framework, Int'l Journal of Human-Computer Studies, 41:(4), pp. 553-583 (1994)
18. Papa, M.F.: Toward the Improvement of a Measurement and Evaluation Strategy from a Comparative Study, In LNCS 7703, Springer: Current Trends in Web Engineering, ICWE Int'l Workshops, pp. 189-203 (2012)
19. Basili, V., Caldiera, G., Rombach, D.: Goal, Question, Metric Paradigm, Enc. of Software Eng., J.J. Marciniak, Ed., John Wiley & Sons, Vol. 1, pp. 528-532 (1994)
20. Buglione, L., Abran, A.: Balanced Scorecards and GQM: What are the Differences. In: Proc. of the 3^{rd} European Software Measurement Conference (FESMA-AEMES 2000), páginas 18–20 (2000)
21. Trendowicz, V.B.A., Heidrich, M.K.J., Münch, C.S.J., Rombach, D.: Aligning Organizations Through Measurement-The GQM+Strategies Approach. Springer, ISBN: 978-3-319-05046-1, ISSN: 2193-8199, doi:10.1007/978-3-319-05047-8 (2014)
22. Rombach, D., Ulery, B.: Improving Software Maintenance through Measurement, IEEE, Proc. ISSN 0018-9219, 77:(4), pp. 581-595 (1989)
23. Kaplan R., Norton D.: The Balanced Scorecard: Translating Strategy into Action, Harvard Business Press (1996)

Green BPM: ciclo de vida de procesos de negocio incorporando aspectos ambientales.

Anahí Soledad Rodríguez[1], Luis Damián Candia[1], Patricia Bazan[1], Viviana Miriam Ambrosi[1, 2], Nestor Castro[1], Francisco Javier Díaz[1], Irina Benitez[1]

[1] LINTI – Facultad de Informática – UNLP
[2] Comisión de Investigaciones Científicas CIC-BA

arodriguez@linti.unlp.edu.ar, dcandia@linti.unlp.edu.ar, pbaz@info.unlp.edu.ar, vambrosi@info.unlp.edu.ar, ncastro@isis.unlp.edu.ar, jdiaz@info.unlp.edu.ar , ibenitez@linti.unlp.edu.ar

Resumen La adopción del paradigma de BPM es cada vez más frecuente en las organizaciones ante los beneficios que esto conlleva en términos de mejoras en la productividad. Por otra parte, la responsabilidad con el medio ambiente es también un compromiso creciente de las organizaciones y que además impacta en los aspectos productivos y tecnológicos. La incorporación de conceptos de tecnología dentro de los procesos de negocio contribuye a la mejora continua de los mismos. A su vez, el concepto de Green IT considera los aspectos tecnológicos en lo que respecta a la protección del medio ambiente durante el ciclo de vida de las TI (Tecnologías de Información). Este artículo integra los conceptos de Green IT a BPM mediante la incorporación de indicadores "verdes" dentro del ciclo de vida de los procesos de negocio y la incorporación de "Alertas Verdes" que permiten definir un ciclo de vida Green BPM y su aplicación a un caso de estudio.

Palabras Clave: Green BPM. Ciclo de vida BPM. Green IT. Desarrollo Sostenible.

1 Introducción

La evidencia creciente de cambios climáticos abruptos ha provocado que cada vez más los gobiernos, instituciones e individuos tomen consciencia y reconozcan la necesidad de un desarrollo sostenible de la vida y de la organización [1].

La Gestión de Procesos de Negocio (*Business Process Management* – BPM) prepara a las organizaciones para nuevos retos y cambios en el entorno, además de estar adaptada a las nuevas tendencias y tecnologías.

La utilización de BPM en las organizaciones contribuye al ciclo de mejora continua y a su eficiencia, como resultado de contar con procesos de negocio identificados, modelados, desplegados y monitorizados [2] [3] [4]. Este ciclo de mejora continua se sustenta fuertemente en la definición de los indicadores que se utilizan en la etapa de monitoreo. El desafío que encuentran los procesos de negocio de las organizaciones en torno al cuidado del medio ambiente es que sus indicadores

no solamente optimicen la productividad sino también la sostenibilidad de las actividades de los procesos de la organización.

El rol de las organizaciones en términos de sostenibilidad es fundamental para el entorno, y además ha llevado a las mismas a alcanzar un mayor potencial socioeconómico. Por lo cual las organizaciones y los gobiernos están cada vez más interesados en tener más responsabilidades ante el cuidado del medio ambiente, además de cumplir con los compromisos internacionales, post COP 21en París [5], y por las propias regulaciones regionales y nacionales.

Las TI sostenibles deben entenderse como una tendencia de las nuevas tecnologías, no solo vinculada a los componentes electrónicos, sino también a su uso eficiente, pero minimizando el impacto ambiental, maximizando su viabilidad económica y propiciando una mejora del contexto social [6]. El concepto de Green IT considera la protección del medio ambiente en todo el ciclo de vida de TI.

En este sentido, los procesos de negocio, ya establecidos en las organizaciones, no están ajenos a esta realidad, por lo que es necesario trabajar teniendo en cuenta esta problemática inminente y los requerimientos asumidos por los gobiernos y empresas a nivel internacional.

Surge así el concepto de proceso de negocio "verde" que requiere que se definan indicadores de mejoras vinculada a aspectos ecológicos y además demanda mejorar las notaciones de modelado y obtener la trazabilidad necesaria para monitorizar los procesos según los indicadores definidos.

En este sentido, las organizaciones pueden estar interesadas en medir cuan sostenibles son sus procesos, por ejemplo midiendo los recursos que consumen [4]. Ante estos nuevos retos se debe analizar como incorporar temas de cuidado del medio ambiente en el ciclo de vida de los procesos de negocio lo cual conduce a definir el concepto de Green BPM [7].

En este artículo se propone la definición de las etapas de un ciclo de vida Green BPM con la incorporación de Alertas Verdes, y la aplicación de esta metodología en un caso de estudio de los procesos del Programa E-Basura [8].

El trabajo se organiza de la siguiente manera: en la Sección 2 se presentan los antecedentes en torno a los indicadores de rendimiento y Green BPM, incluyendo conceptos vinculados con procesos "verdes". En la Sección 3 se presentan la definición de las etapas del ciclo de vida BPM en el contexto de Green IT, la incorporación de indicadores ambientales en el ciclo de vida de los procesos de negocio y la incorporación de Alertas Verdes que contribuyen a la mejora continua del proceso. En la Sección 4 se analiza el nuevo mecanismo en un caso de estudio para el ciclo de vida Green BPM. En la Sección 5 se presentan las conclusiones y trabajos futuros.

2 Antecedentes en Indicadores de Rendimiento y Green BPM

Las organizaciones comúnmente optimizan sus procesos de negocio dentro de las dimensiones clásicas como costo, calidad, tiempo, flexibilidad pero en la actualidad, con la creciente preocupación con el medio ambiente, podemos analizar una nueva dimensión: la sostenibilidad [7]. Con la incorporación de esta nueva dimensión se deben tener en cuenta nuevos indicadores como por ejemplo, consumo de energía, consumo de agua, desperdicios ocasionados por la producción del producto en la

ejecución del proceso de negocio y las actividades del mismo [9] que pueden generar un impacto negativo en el medio ambiente. El análisis y medición de estos nuevos indicadores permitirían aplicar cambios al proceso de negocio para hacerlo más eficiente y efectivo con el medio ambiente.

Es muy importante para las organizaciones tener en cuenta los KPI (*Key Performance Indicators* o Indicadores Clave de Rendimiento). Los KPI son métricas cuantificables que reflejan el desempeño de una empresa en el contexto del logro de sus metas y objetivos más amplios. La ITU (*International Telecommunication Union*) identificó que indicadores claves podían ser útiles para una organización. Los KPI ayudan a las empresas a implementar estrategias vinculando varios niveles de una organización (unidades de negocios, departamentos e individuos) con objetivos y puntos de referencia claramente definidos [10, 11]. Hay cinco grupos principales de KPI ambientales que normalmente se clasifican por instituciones del sector TIC y no TIC: la energía, las emisiones de gases de efecto invernadero, residuos, agua, y otros.

Los indicadores claves ambientales (*Environmmental* KPI) son de tipo directo e indirecto [26] y también encontramos de tipo estático y dinámico. El dinámico es el que se genera del uso real de un dispositivo, por ejemplo el consumo de energía de una impresora; mientras que los estáticos son independientes del uso, por ejemplo servidores, en el cual su uso es constante [27].

El concepto de Green BPM se encuentra en las primeras fases de las investigaciones, los autores Opitz, Krüp y Kolbe en su trabajo de investigación [12] proponen la siguiente definición: *"Green BPM es la suma de todas las actividades de gestión soportadas por los sistemas de información que ayudan a monitorizar y reducir el impacto ambiental de los procesos de negocios en sus etapas de diseño, mejora, implementación u operación, así como guiar hacia un cambio cultural dentro del ciclo de vida del proceso".*

La intención detrás de Green BPM es la incorporación de objetivos ecológicos en la gestión de procesos comerciales [13]. Para incorporar la sostenibilidad a BPM se debe ejecutar cada fase del ciclo de vida teniendo en cuenta cuestiones ecológicas. También es muy importante el uso de conceptos, métodos, técnicas y herramientas para considerar los diferentes grados de impacto en el medio ambiente, la economía y la sociedad [14].

La bibliografía es muy reciente en cuanto a cómo medir y monitorizar un Green BPM [24]. En [27] se analizan los aspectos de investigación basada en la revisión de Green BPM que sin dudas marcan un camino hacia la definición de KPIs en torno a su impacto ambiental.

El ciclo de vida de BPM se compone de fases cíclicas que no implican necesariamente un orden temporal, pero sí una dependencia lógica, según lo define M.Weske en [25]. Si bien hay variantes en cuanto a las diferentes etapas o fases que componen este ciclo de vida, podemos decir que las mismas se identifican como: 1- diseño, que es estudio de la situación desde el punto de vista técnico como organizacional, 2 - modelado, que identifica, revisa, valida y representa los procesos de negocios en un modelo, 3- ejecución, que implica el despliegue del proceso modelado en una entorno ejecutable, 4- monitoreo, registra los rastros de ejecución de las distintas instancias del proceso alimentando los KPI y 5- optimización, revisa los rastros de ejecución para aplicar los cambios y cerrar el ciclo de mejora continua.

En la tabla 1 se muestra la vinculación entre Green IT y el ciclo de vida BPM de Weske. Dentro del ciclo de vida verde se tienen en cuanta las etapas de ejecución y evaluación. A su vez, se define la incorporación de un componente para el procesamiento de Alertas Verdes

.

Tabla 1 – Vinculación entre el Ciclo de Vida Verde y el tradicional.

<table>
<tr><th>Fase de BPM</th><th colspan="2">Ciclo de Vida Verde</th></tr>
<tr><td>Diseño</td><td rowspan="4" colspan="2">Análisis y Diseño Verde</td></tr>
<tr><td>Diseño</td></tr>
<tr><td>Diseño</td></tr>
<tr><td>Modelado</td></tr>
<tr><td>Monitoreo</td><td colspan="2">Monitoreo Verde</td></tr>
<tr><td rowspan="2">Optimización</td><td>Optimización y Evaluación Verde</td><td rowspan="2">Alerta Verde</td></tr>
<tr><td>Ejecución Verde</td></tr>
</table>

3 Definición de las etapas del ciclo de vida BPM en el contexto de Green IT

Según los antecedentes presentados en torno a Green BPM se desprende que existen definiciones precisas para este concepto en cuanto a la incorporación de indicadores ecológicos y a las mejoras en torno al impacto climático, que requieren de un marco tecnológico que las implemente, además de una metodología o ciclo de vida verde.

Actualmente, las organizaciones han alcanzado cierto grado de madurez en cuanto a la adopción del paradigma BPM, incluso utilizando herramientas específicas para su despliegue (BPMS). En este sentido, la ejecución de procesos de negocio es un desafío ya resuelto y hay una gran variedad de alternativas tal como se analiza en [19].

Dado que el ciclo de vida de BPM es una mecanismo cíclico donde todas las fases se vinculan y se alimentan unas a otras, resulta conveniente definir las diferentes etapas a la luz de los aspectos ambientales y determinar los mecanismos precisos que ayudan a considerarlos.

Diseño y Modelado Verde

Esta etapa incluye los aspectos de actitud, estrategia y gobernabilidad verdes a través del análisis de *Key Ecological Indicators* (KEIs). Se pueden diseñar distintas alternativas de procesos, por ejemplo en una organización destinada a la fabricación de productos la utilización de materias primas alternativas no contaminantes. También se podría permitir que el cliente elija alguna alternativa con diferentes impactos ambientales [18]. Esto llevará a tener nuevos diseños en los modelos de procesos,

donde se incorporen notaciones. Por ejemplo para las emisiones dióxido de carbono o para el consumo de energía de las actividades o del proceso completo, tal como se presenta en [1] donde se propone una notación que agrega simbología gráfica de aspectos ecológicos. Para facilitar la lectura de ciertos indicadores ecológicos se podrían incorporar en el diseño la utilización de sensores de IoT (*Internet of Things*). Por lo que resulta indispensable tener una notación para ser incorporados en el modelado del mismo. Meyer y Sperner son dos autores que analizaron la manera de simbolizar dispositivos IoT y toda la tecnología que rodea a los mismos en el modelado de BPM [20, 21, 22].

Ejecución Verde

En esta etapa entra en juego el uso de un BPMS para dar soporte a la implementación y despliegue del proceso [15]. Por lo cual hay que evaluar que BPMS tiene soporte para la incorporación de los KEI tanto antes como durante la ejecución del proceso [18]. La incorporación de estos KEI antes de la ejecución puede hacerse en la etapa de modelado pero en el caso de la incorporación en ejecución es necesario contar con mecanismos para incluirlo sin interrupción de la ejecución del proceso [4]. El mecanismo para incorporar indicadores en tiempo de ejecución requiere un BPMS con motores de ejecución que permitan [23]:

- *Notificaciones*: el uso de notificaciones en el BPMS, las cuales son avisos que se disparan ante determinados eventos dentro de la ejecución de un proceso, estos eventos pueden ser: la creación de un nuevo ítem de trabajo, cambios de prioridad, mensajes internos, nuevas etiquetas. Ante cualquiera de dichos eventos se notifica al participante del proceso involucrado.
- *Comentarios etiquetados*: las actividades pueden ser comentadas o no, en los comentarios se pueden incluir etiquetas a otros usuarios, el nombre de una actividad o bifurcador o el nombre de una variable de proceso. Estos registros luego serán utilizados en la monitorización de la ejecución del proceso.
- *Definición de tareas ad-hoc sin modelar*: Para crear nuevas tareas "no previstas" en el modelado del proceso, que se comportarán como sub-tareas del proceso.
- *Notificaciones a los responsables de los procesos*: Para permitir la comunicación entre los participantes del proceso y el responsable del mismo.

Durante la ejecución, la fuente de información para la lectura de los indicadores puede estar provista por sistemas de información, sensores o entrada manual, por lo tanto BPMS también debe admitir distintos orígenes de datos.

Monitoreo Verde

En esta etapa se evalúan las instancias del proceso. El monitoreo al igual que la ejecución, se realiza a través de un BPMS [15]. En el monitoreo verde los rastros clásicos del proceso, se amplían con la información ecológica provista en fases anteriores, por lo que los métodos y técnicas deben considerar este tipo de registros [18]. Las técnicas de monitoreo se verán afectadas dependiendo del momento en que se tenga en cuenta el indicador, ello puede ser antes o durante la ejecución, por lo cual el monitoreo puede ser:

- Estático: evalúa los registros obtenidos de la ejecución de las instancias del proceso, que impacten en el medio ambiente y produce los *dashboards* que se utilizarán para la etapa de evaluación.
- Dinámico: considera rastros enriquecidos por información obtenida de manera dinámica durante la ejecución del proceso, produciendo indicadores no previstos (dinámicos) en la etapa de modelado, lo cual genera nuevos *dashboards* para la etapa de evaluación.

Alertas Verdes: Las alertas son señales producidas en la etapa de Ejecución y que afectan al Monitoreo, como indica la Figura 1. Estas señales se disparan al superarse determinados umbrales definidos al promulgar el proceso y que impactan en el medio ambiente. Las alertas verdes pueden generar distintas acciones, como por ejemplo:

- Informar a un responsable sobre una situación anómala.
- Detener la ejecución del proceso ante una situación que exceda rangos establecidos para el cuidado del medio ambiente.
- Lanzar/disparar la ejecución de un nuevo proceso encargado de la recuperación de los errores encontrados.
- Definir distintos tipos de alertas verdes con distintos grados de impacto negativo para el medio ambiente. Por ejemplo ciertos impactos pueden requerir detener un proceso por completo, y en otros casos disparar otra acción. Por lo tanto, los umbrales deberán ser definidos previamente para posibilitar la generación de alertas en el caso de exceder estos valores. Durante el monitoreo dinámico serán involucrados dichos indicadores dinámicos obtenidos durante la ejecución del proceso.

Optimización y Evaluación Verde:

La optimización verde del proceso también se puede llevar a cabo con la utilización de patrones definidos en [9]. Por ejemplo se pueden detectar recursos con consumos excesivos de energía para lo cual se puede aplicar el "Patrón 3: Cambio de Recursos" el cual propone seleccionar recursos en función de su impacto ambiental. También se proponen dos patrones los cuales proponen el cambio de una tarea manual a automática o viceversa, esto se analiza en los patrones: "Patrón 6: Proceso automatizado" o "Patrón 7: Rendimiento en el proceso humano/manual".

La evaluación del proceso tiene en cuenta los aspectos tradicionales del proceso más los que involucran el cuidado del medio ambiente, y de esta forma poder contribuir a la mejora continua del proceso. Las técnicas tradicionales de evaluación deberán ser cambiadas por técnicas que tengan soporte para indicadores verdes. Las "partes interesadas" de los procesos podrán identificar los KEI y así obtener optimizaciones que produzcan procesos más sostenibles con respecto al impacto ambiental [18].

La Figura 1 muestra las etapas del ciclo de vida tradicional de BPM definido por Weske (en color azul) y la definición de un modelo de ciclo de vida de un Green BPM (en color verde) basados en [18] considerando la gestión de Alertas Verdes.

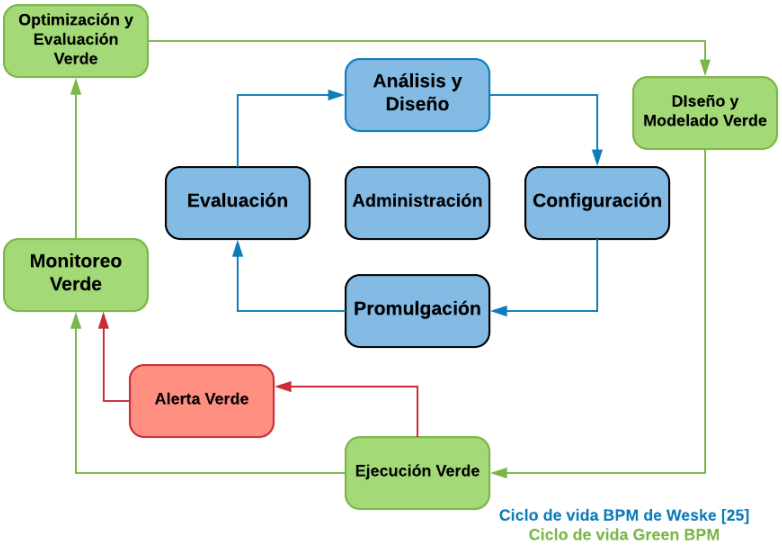


Figura 1 – Ciclo de Vida de Procesos de Negocio con consideraciones "verdes"

4 Caso de Estudio

Se toma como caso de estudio, para aplicar las etapas definidas en la metodología propuesta para completar un ciclo de vida de procesos de negocio que considere aspectos medioambientales, los procesos del Programa E-Basura.

El Programa E-Basura [8] surgió en el año 2009 como iniciativa del Laboratorio de Investigación en Nuevas Tecnologías Informáticas (LINTI) [16] de la Facultad de Informática de la Universidad Nacional de La Plata (UNLP) [17]. Comenzó como un Proyecto de Extensión Universitaria para reforzar el compromiso de la Universidad Pública con la Sociedad. En 2017 fue declarado Programa por resolución de la propia UNLP, y en 2018 se convertirá en una Planta Piloto Experimental en temas de residuos electrónicos en base a un convenio de colaboración internacional con la ITU. Entre los objetivos del programa, se busca generar conciencia y educar tanto en la UNLP como en la sociedad sobre los graves problemas que pueden generar al ambiente y la salud la incorrecta disposición final de los RAEE; propiciar la reutilización tecnológica a través del reacondicionamiento de equipos provenientes del sector de las TIC y su posterior donación para posibilitar el acceso a la tecnología en los sectores más desfavorecidos de la sociedad; reciclar y valorizar materiales; formar en oficios y contribuir a una disposición final segura de los materiales irrecuperables.

Para llevar a cabo las tareas el Programa adopta el paradigma BPM (*Business Process Management*) para documentar, optimizar y dar flexibilidad a sus tareas diarias a través de la gestión de sus procesos.

Del caso de estudio mencionado fue seleccionado el proceso piloto denominado *Molienda de Plásticos*. El objetivo del mismo es realizar el seguimiento de las tareas vinculadas a preparación y procesamiento de lotes de plásticos provenientes de los residuos electrónicos, de forma tal de poder ser reinsertados en el mercado como insumo productivo.

Diseño y Modelado Verde

En esta etapa se agrega la lectura de indicadores ecológicos, la cual se realizará a través de una red de sensores. En la Figura 2 se presenta la etapa de "Diseño y Modelado Verde" del proceso piloto de *Molienda de plástico* llevado a cabo en la Planta Piloto Experimental de Residuos Electrónicos. En este proceso se medirá el nivel de **contaminación sonora y de aire** producido por la propia molienda de plásticos para el reciclaje de materiales de descarte industrial de plásticos de RAEE a través de una red de sensores. Para el modelado se utiliza la notación propuesta en el artículo [28] para la incorporación de las lecturas de indicadores ecológicos, como se observa en la Figura 2, para la cual se utilizan:

- Una entidad física: un elemento del mundo real que se conecta a la red de sensores y actuadores, que es modelada como una notación de texto denominada `MedioAmbiente`.
- Una entidad virtual: la representación virtual de la entidad física, que es modelada en el diagrama como un repositorio de datos llamado `MedioAmbiente`
- Un Dispositivo de IoT: el componente que interactúa en forma directa con los sensores y actuadores, que es modelado con un lane llamado `Dispositivo de IOT - Raspberry`

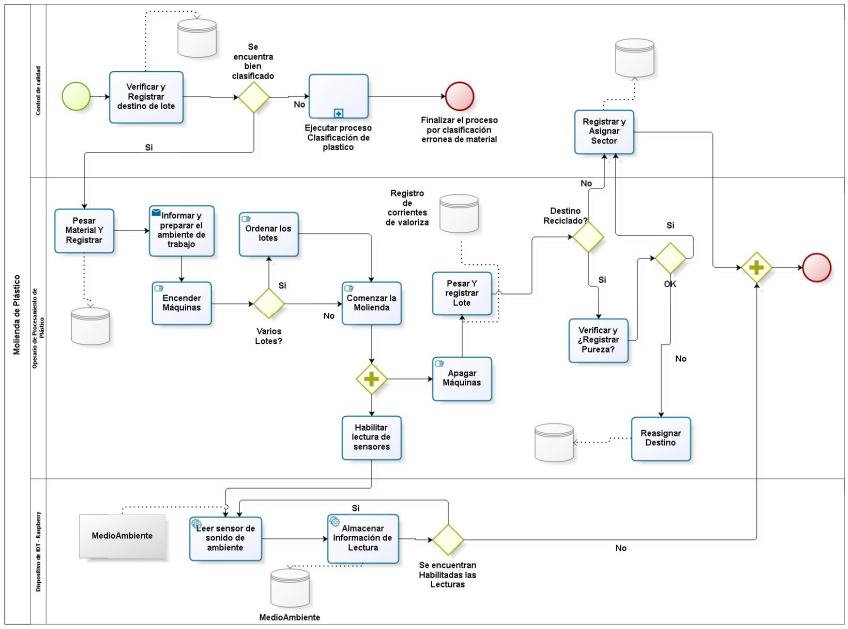


Figura 2 – Proceso Molienda de Plástico

Ejecución Verde

La etapa de ejecución dentro del ciclo de vida de los procesos de negocio no solo controla la ejecución sino que su resultado aporta a la trazabilidad de los procesos. En la ejecución del proceso se tendrá en cuanta la lectura de los indicadores de la **contaminación sonora** y **del aire** antes propuesto. Para el caso en que durante la ejecución surja la necesidad de incluir un indicador dinámico, este indicador será incorporado en tiempo de ejecución utilizando la técnica propuesta en [4] y que alcanza dos aspectos: 1- incorporación de nuevos indicadores para las actividades en un BPMS específico y la definición de cómo medirlos y 2- definición de un mecanismo para incorporar, en ejecución, estos nuevos indicadores mediante actividades automáticas.

Monitoreo Verde

Como se mencionó en la Sección *Definición de las etapas del ciclo de vida BPM en el contexto de Green IT*, la etapa del "Monitoreo Verde" contemplará las trazas generadas por los indicadores verdes de contaminación sonora y del aire antes mencionado.

Ante la necesidad de la lectura de un indicador en ejecución (dinámico), sostenido por el concepto de **Alertas Verdes**, habilitaría a que los procesos en ejecución reaccionen ante señales, como por ejemplo, la superación de un umbral definido como seguro para el medio ambiente, y se puedan tomar acciones como notificar al responsable del proceso mediante el sistema de mensajería propuesto en [20], detener el proceso, ejecutar un proceso alternativo para recuperar fallos o resolver los resultados negativos provocados por la alerta.

Optimización y Evaluación Verde

Cierre del ciclo de mejora continua con características de cuidado del medio ambiente. Esta etapa considera los indicadores verdes y aplica el patrón: "*Pattern 1: Green Compensation*" definidos en [8], en el cual se puede realizar un análisis para identificar posibles estrategias de optimización. El desafío es encontrar una solución adecuada para compensar el impacto negativo en el medio ambiente sin cambiar el proceso.

5 Conclusiones y Trabajos Futuros

En el presente artículo se propone la incorporación de Alertas Verdes dentro de las etapas de un ciclo de vida Green BPM. Durante este ciclo de vida se tienen en cuenta no solo los indicadores definidos durante la etapa de "Diseño y Análisis" sino también nuevos indicadores incorporados durante la ejecución del proceso. Estos nuevos indicadores provocan nuevas trazas de ejecución en tiempo real y permiten realizar un monitoreo dinámico basados en la sostenibilidad.

La incorporación de Alertas Verdes juega un papel muy importante dentro del ciclo de mejora continua, permitiendo una evaluación y mejora que aumentaría la eficiencia, eficacia y efectividad del proceso verde.

Se logra aplicar el ciclo de vida verde en un caso de estudio dentro del Programa E-Basura logrando una mayor visibilidad de las actividades de sus procesos con la

implementación del enfoque BPM, y además adherir a los objetivos del Programa en cuanto al cuidado del medio ambiente.

Los trabajos futuros de investigación consistirán en la propuesta de técnicas y herramientas para la generación de alertas verdes y monitoreo y así poder mejorar los procesos de negocio reduciendo el impacto ambiental. Con esto se pueden tener trazas de ejecución enriquecidas con nuevas estadísticas sostenibles.

Referencias

1. RECKER, Jan C. Green, Greener, BPM?. BPTrends, 2011, vol. 5, no 7, p. 1-8.
2. LISTA, Ernesto Amaru Galvis; ZABALA, Mayda Patricia González. Herramientas para la gestión de procesos de negocio y su relación con el ciclo de vida de los procesos de negocio: Una revisión de literatura. Ciencia e Ingeniería Neogranadina, 2014, vol. 24, no 2, p. 3.
3. Miguel Alcalá, Conferencia IBM, Gestión de procesos de negocio en la nueva era tecnológica, Diciembre 2014.
4. RODRÍGUEZ, Anahí S., Bazán, P., Ambrosi, V. M., & Díaz, F. J et al. Modelado y ejecución de procesos de negocio con conceptos de Green IT. En XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016). 2016.
5. 21a Conferencia de las Partes (COP) de la Convención Marco de Naciones Unidas sobre Cambio Climático (CMNUCC), http://www.un.org/sustainabledevelopment/es/Cop21/
6. GINER, Javier Muñoz; RINCÓN, Yuresky Rojas. Nuevas tendencias en tecnologías verdes-Green IT para la Gestión en Organizaciones. En II Congreso Iberoamericano SOCOTE-Soporte al Conocimiento con la Tecnología-y VII Congreso SOCOTE Universidad Politécnica de Valencia. 2010.
7. SEIDEL, Stefan; RECKER, Jan; VOM BROCKE, Jan. Green business process management. En Green Business Process Management. Springer, Berlin, Heidelberg, 2012. p. 3-13.
8. Programa E-Basura, https://e-basura.linti.unlp.edu.ar/
9. SEIDEL, Stefan; RECKER, Jan; VOM BROCKE, Jan. Green business process management. En Green Business Process Management. Springer, Berlin, Heidelberg, 2012. p. 3-13.
10. ITU General specifications and KPIs, https://www.itu.int/dms_pub/itu-t/oth/4B/04/T4B0400000B0009PDFE.pdf
11. OPITZ, Nicky; KRÜP, Henning; KOLBE, Lutz M. Environmentally Sustainable Business Process Management-Developing a Green BPM Readiness Model. En PACIS. 2014. p. 12.
12. OPITZ, Nicky; KRUP, Henning; KOLBE, Lutz M. Green Business Process Management--A Definition and Research Framework. En System Sciences (HICSS), 2014 47th Hawaii International Conference on. IEEE, 2014. p. 3808-3817..
13. REITER, Markus; FETTKE, Peter; LOOS, Peter. Towards green business process management: concept and implementation of an artifact to reduce the energy consumption of business processes. En System Sciences (HICSS), 2014 47th Hawaii International Conference on. IEEE, 2014. p. 885-894.
14. MAGDALENO, Andréa Magalhães; DUBOC, Leticia; BETZ, Stefanie. How to Incorporate Sustainability into Business Process Management Lifecycle?. En International Conference on Business Process Management. Springer, Cham, 2016. p. 440-443.

15. BAZÁN, Patricia. Un modelo de integrabilidad con SOA y BPM. 2009. Tesis Doctoral. Facultad de Informática.
16. Laboratorio de Investigación en Nuevas Tecnologías Informáticas, https://www.linti.unlp.edu.ar/
17. Universidad Nacional de La Plata, https://unlp.edu.ar/
18. NOWAK, Alexander; LEYMANN, Frank; SCHUMM, David. The differences and commonalities between green and conventional business process management. En Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on. IEEE, 2011. p. 569-576.
19. DELGADO, Andrea, et al. A systematic approach for evaluating BPM systems: case studies on open source and proprietary tools. En IFIP International Conference on Open Source Systems. Springer, Cham, 2015. p. 81-90.
20. MEYER, Sonja, et al. Towards modeling real-world aware business processes. En Proceedings of the Second International Workshop on Web of Things. ACM, 2011. p. 8.
21. MEYER, Sonja; RUPPEN, Andreas; MAGERKURTH, Carsten. Internet of things-aware process modeling: integrating IoT devices as business process resources. En International conference on advanced information systems engineering. Springer, Berlin, Heidelberg, 2013. p. 84-98.
22. DE MEYER, Thibault. Integrating the internet of things into business process management.
23. BAZÁN, Patricia, et al. Mejora de la monitorización y ejecución de procesos de negocio con integración y socialización. En Computing Conference (CLEI), 2015 Latin American. IEEE, 2015. p. 1-12.
24. RÖGLINGER, Maximilian; PÖPPELBUß, Jens; BECKER, Jörg. Maturity models in business process management. Business Process Management Journal, 2012, vol. 18, no 2, p. 328-346.
25. Weske, M. (2012). Business process management architectures. In Business Process Management (pp. 333-371). Springer, Berlin, Heidelberg.
26. JONES, L. G. Environmental Key Performance Indicator, Reporting Guidelines for UK Business. Queen's Printer and Controller. London, 2006.
27. OPITZ, Nicky; KRÜP, Henning; KOLBE, Lutz M. Environmentally Sustainable Business Process Management-Developing a Green BPM Readiness Model. En PACIS. 2014. p. 12.
28. CANDIA, Damián, et al. Integración de IoT en el modelado de procesos de negocio. En XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste). 2018.

Problemas que afectan a la Calidad de Software en Entrega Continua y Pruebas Continuas

Maximiliano A. Mascheroni[1,2], Emanuel Irrazábal[1],

[1] Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste. Corrientes, Argentina.
[2] Facultad de Informática. Universidad Nacional de La Plata.
La Plata, Buenos Aires, Argentina.

{mascheroni, eirrazabal}@exa.unne.edu.ar

Abstract. Muchas organizaciones que desarrollan software de manera ágil buscan adoptar el enfoque de entrega continua, el cual permite la liberación del software a producción en cualquier momento. Sin embargo, uno de los principales desafíos es mantener la calidad del mismo gestionando adecuadamente el tiempo que se invierte en las pruebas. En este sentido, se han reportado numerosos problemas relacionados con la calidad del software al adoptar este enfoque. Asimismo, algunos autores y empresas han propuesto soluciones para hacer frente a algunos de ellos, pero surge el interrogante de si las pruebas continuas son el elemento clave para estas prácticas. En este trabajo, se presenta un resumen de los principales problemas que afectan a la calidad de software en entrega continua y la relación que existe entre los mismos. Al final, se propone como resultado del análisis de estudios previos una definición al concepto de “pruebas continuas".

Keywords: entrega continua, calidad de software, pruebas continuas.

1 Introducción

El desarrollo ágil de software ha incorporado una gran variedad de buenas prácticas y procesos para producir software de manera más veloz y eficiente [1]. Una de las más conocidas es la Integración Continua (CI). CI es una práctica del desarrollo de software, en donde los desarrolladores integran el código fuente con frecuencia y cada integración es verificada por un sistema que construye el código y lo prueba automáticamente [2]. Esta práctica en sus inicios estaba soportada solamente por pruebas unitarias. Pero con la aparición de nuevas herramientas como Selenium WebDriver[1], niveles más altos de pruebas podían ser automatizados, y así introducidos en el proceso de CI. Sin embargo, para ejecutar pruebas de manera automática sobre una versión del código en ejecución, es necesario el despliegue del

[1] Selenium Webdriver es una herramienta que permite la automatización de la interacción con páginas web. https://www.seleniumhq.org/projects/webdriver/

mismo a un ambiente de desarrollo. Esto dio lugar a la aparición de nuevas herramientas para automatizar el despliegue de código, y así incorporar este proceso, junto con la ejecución de pruebas automatizadas de distintos niveles en CI.

Posteriormente, se añade la idea de realizar despliegues automatizados en producción. Por un lado, algunas compañías buscaban automatizar el proceso completo, desde la integración del código fuente, hasta el despliegue a producción. Por otro lado, otras empresas, deciden automatizar todo el flujo, pero el despliegue a producción se realizaría de manera automática al presionar un botón cuando el cliente lo indique. El primer enfoque se denomina Despliegue Continuo (DC) y el segundo Entrega Continua (CD, del inglés Continuous Delivery). En todas estas prácticas (CI, DC y CD), las pruebas automatizadas cumplen un rol fundamental, ya que deben brindar a los desarrolladores confiabilidad en cada versión del producto que se construye. Para ello, los autores de CD proponen un patrón llamado *conducto de despliegue*, más conocido en inglés como Deployment Pipeline (DP) [3]. Con DP, el flujo se divide en diferentes etapas, donde a medida que se avanza de una etapa a la otra, se aumenta la confiabilidad, usualmente al costo de más tiempo de ejecución [4].

De esta manera, encontrar un equilibrio entre un tiempo de ejecución pequeño y un lote de pruebas que aseguren la calidad de cada entregable es una tarea difícil de llevar a cabo. En la literatura, pueden encontrarse revisiones y casos de estudio que reportan problemas relacionados con CD, y la mayor parte de ellos afectan a la calidad.

Con el fin de contribuir a la apertura de nuevas líneas de investigación en el área de ingeniería de software, en este trabajo se realiza un resumen de los problemas más frecuentes reportados en CD, que afectan a la calidad del software. Asimismo, se realiza un análisis de definiciones de "pruebas continuas", proponiendo una más general.

Además de esta sección introductoria, en la sección 2 se resume brevemente en qué consiste CD y sus elementos principales. En la sección 3 se describen los problemas que afectan a la calidad. La sección 4 representa el apartado de pruebas continuas. Finalmente, en la sección 5 se perfilan conclusiones y trabajos futuros.

2 Entrega Continua

Según Humble y Farley [3], CD fue creado para resolver el siguiente interrogante: “Si alguien tiene una buena idea, ¿cómo se la hace llegar a los usuarios lo más rápido posible?”. CD se define formalmente como un enfoque en el cual los equipos mantienen la producción de software en ciclos cortos de tiempo, asegurando que el producto pueda ser lanzado de manera fiable en cualquier momento [3, 5]. El objetivo es lanzar a producción una versión del software libre de defectos “con solo apretar un botón” [3]. Los principales beneficios de CD son [3, 4]: riesgo más bajo al realizar el despliegue; progreso real; y feedback inmediato por parte del usuario final.

La clave principal de CD es permitir a los clientes o la gente involucrada en el negocio, requerir que el desarrollo actual de una versión del software pueda ser lanzado a producción en el mismo momento en que se lo solicita. Sin embargo, para

lograrlo se deben cumplir ciertos requisitos. Los autores recomiendan las siguientes prácticas para la implementación de CD [3, 4]:

- Integración continua.
- Gestión de la configuración, no solo del código fuente, sino de cualquier artefacto relacionado al desarrollo del software, como por ejemplo esquemas de bases de datos, archivos de configuración, etc.
- Automatización de pruebas de software, en la mayor cantidad de niveles posibles (unitarias, de integración, funcionales, etc.)
- Generación de una cultura centrada en la comunicación, colaboración e integración entre desarrolladores de software y profesionales de operaciones en las tecnologías de la información (DevOps).

Además, para que el producto se encuentre libre de defectos, es necesario que los lotes de pruebas automatizados brinden confiabilidad. Como se mencionó anteriormente, DP es un patrón que permite dividir el flujo en etapas. Aún no existe un modelo estándar para la aplicación de las etapas de DP [5], y cada empresa lo implementa según sus necesidades. Sin embargo, el DP presentado por Humble y Farley [3] tiene las siguientes etapas (ver Fig. 1):

1. Introducción de cambios: incluye la compilación del código, ejecución de pruebas unitarias, análisis estático del código fuente y creación de artefactos.
2. Ejecución de pruebas de aceptación automatizadas.
3. Ejecución de pruebas no funcionales automatizadas: se deben incluir pruebas de capacidad, carga y sobrecarga.
4. Ejecución de pruebas manuales.

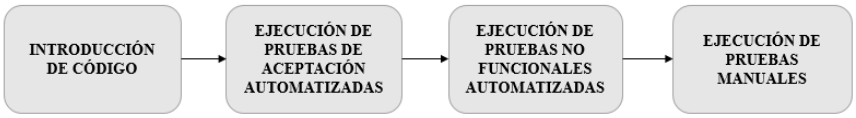


Fig. 1. Deployment Pipeline (conducto de despliegue).

En la actualidad, existen herramientas en el mercado que facilitan la implementación de las etapas de CD. Por otro lado, algunas empresas optan por integrar herramientas de CI con herramientas de DC. Finalmente, existen módulos y plugins de las propias herramientas de CI que permiten el despliegue de software. Algunos ejemplos de estas herramientas pueden verse en la Tabla 1, donde se especifica su aplicación (CI, DC, CD, plugin o módulo), el tipo de licencia (libre, comercial o ambos) y finalmente su sitio web. Sin embargo, si bien las mismas facilitan la implementación de las etapas de DP, todavía no existe un guía, modelo o lista de factores claves para el diseño, desarrollo y ejecución de pruebas en CD. En la literatura, pueden encontrarse revisiones sistemáticas que estudian los problemas de CD, y mencionan que aún se encuentran inconvenientes relacionados a las pruebas en este tipo de enfoque. Además, empresas y casos de estudios, reportan que mantener la calidad de software resulta un desafío al implementar CD.

Tabla 1. Herramientas más populares para implementar DP.

Herramienta	Aplicación	Licencia	URL
GoCD	Entrega Continua	Libre	https://www.gocd.org/
Jenkins	Integración Continua	Libre	https://jenkins.io/
Travis CI	Integración Continua	Ambos	https://travis-ci.org/
Bamboo	Integración Continua	Comercial	https://atlassian.com/software/bamboo
Strider CD	Despliegue Continuo	Libre	https://github.com/Strider-CD
Team City	Integración Continua	Ambos	https://jetbrains.com/teamcity/
Chef	Despliegue Continuo	Ambos	https://www.chef.io/chef
Ansible Tower	Despliegue Continuo	Comercial	https://www.ansible.com/products/tower
Buddy	Entrega Continua	Ambos	https://buddy.works/
Puppet	Despliegue Continuo	Ambos	https://puppet.com/
Wercker	Entrega Continua	Ambos	http://www.wercker.com
Jenkins Pipelines	Plugin	Libre	https://jenkins.io/solutions/pipeline/
Bamboo Deployments	Módulo	Comercial	https://atlassian.com/software/bamboo
Puppet Pipelines	Módulo	Comercial	https://puppet.com/products/puppet-pipelines

3 Problemas que afectan a la calidad de software

Actualmente, CD es un concepto que está siendo adoptado por empresas que desarrollan software de manera ágil. Como se mencionó anteriormente, esta disciplina soporta prácticas ágiles y reduce los tiempos de lanzamiento de nuevas versiones del producto al mercado, de semanas a solo horas. Sin embargo, de acuerdo a Prusak [6], "la industria aún no ha cerrado el círculo cuando se trata de implementar un proceso completo de entrega continua". Si bien la literatura contiene instrucciones y buenas prácticas de cómo adoptar CD, su implementación ha sido un desafío en la práctica [7]. Es por ello que para algunos autores [6, 8, 9], las pruebas continuas (Continuous Testing) son el elemento faltante en el proceso de desarrollo continuo, que incluye las prácticas mencionadas.

Algunas organizaciones aún no han podido adoptar CD completamente [10], y otras han encontrado muchos obstáculos [5, 10–15]. La mayor parte de los problemas reportados están relacionados principalmente con pruebas e integración [7].

A continuación, se presenta una lista de problemas relacionados con la calidad de software, que han sido reportados en revisiones de la literatura, casos de estudio y artículos empíricos, cuyo objetivo era el estudio y la implementación de CI, DC o CD.

- **Pruebas que consumen mucho tiempo de ejecución** [15–17]: existen muchos tipos de pruebas que pueden implementarse en CD. Sin embargo, ejecutar un gran lote de pruebas es una tarea que lleva mucho tiempo. Además, los cambios son introducidos con más frecuencia, y por eso, es necesario la ejecución de regresiones lo más rápido posible.
- **Pruebas no deterministas (Flaky Tests)** [15, 18–21]: Una prueba no determinista es aquella que podría generar un resultado positivo o negativo por la misma versión del software. Son pruebas que producen los llamados "falsos positivos" y por su inestabilidad son más conocidas como "flaky tests". Una de las características más importantes de CD es la confiabilidad, y pruebas que fallan aleatoriamente no son confiables.

- **Automatización de pruebas de interfaz gráfica de usuario** [22–25]: Las pruebas de interfaz de usuario eran ejecutadas de manera manual, pero con la aparición de herramientas como Selenium WebDriver, fue posible comenzar a automatizarlas. Sin embargo, la interfaz de usuario es la parte de la aplicación que cambia con más frecuencia y realizar pruebas sobre la misma, puede ocasionar "flaky tests".
- **Resultados de ejecución de pruebas ambiguos** [19, 26]: En entornos de desarrollo continuo, los desarrolladores deben ser notificados de cualquier defecto introducido, detectado a través de las pruebas. Cuando los resultados de la ejecución no son claros, es decir, no detallan el error, su causa, el lugar donde ocurrió, capturas de pantalla, etc., se tienen resultados ambiguos.
- **Automatización de pruebas web con contenido dinámico** [8, 25, 27]: Las aplicaciones web más modernas, utilizan tecnologías como Ajax, React o Angular, que generan contenido dinámico. Incorporar pruebas automatizadas de este tipo de tecnologías en un DP es complejo, ya que una determinada página web pudo haber alcanzado el estado esperado final, pero el contenido dinámico aún se encuentra cargando.
- **Pruebas en Big Data** [28]: Big data es el proceso de utilizar grandes conjuntos de datos que no se pueden procesar utilizando técnicas tradicionales. Realizar verificaciones sobre este conjunto de datos es un nuevo reto que involucra varias técnicas y herramientas que aún están madurando, y por lo tanto es difícil incorporarlas en CD.
- **Pruebas de datos** [29]: Los datos son muy importantes para diferentes tipos de sistemas y los errores en estos son costosos. Si bien las pruebas de software han recibido gran atención en diferentes niveles, las pruebas de datos han sido poco tenidas en cuenta. Es por ello, que no existen soluciones totales, para la implementación de pruebas de datos en un entorno de CD.
- **Pruebas en dispositivos móviles** [30]: Las pruebas móviles han traído consigo muchos desafíos. Entre ellos se encuentran: el proceso de pruebas en sí, los niveles y tipos de pruebas a considerar, los diferentes tipos de dispositivos y los costos de las pruebas automáticas. Todos estos factores dificultan la implementación de pruebas en dispositivos móviles en CD.
- **Pruebas automatizadas de requerimientos no funcionales** [16, 20]: Si bien las pruebas unitarias, de integración y funcionales, han sido estudiadas ampliamente en la literatura e implementadas por muchas organizaciones en la práctica, las pruebas no funcionales han sido poco consideradas.
- **Pruebas de aplicaciones compuestas por servicios en la nube** [31]: Actualmente, existe una gran cantidad de servicios en la nube, y el tamaño de los datos que deben manejar es muy grande. Probar flujos compuestos por estos servicios es muy complejo. Además, es necesario proporcionar garantías de calidad de servicio (QoS, Quality of Service) [32].
- **TaaS en entrega continua** [33]: Las pruebas como un servicio (TaaS), es un modelo en cual las pruebas son ejecutadas por un proveedor de servicios en lugar de un equipo de pruebas perteneciente a la misma organización que desarrolla el software. Realizar TaaS en CD, resulta una tarea muy compleja, puesto que los componentes y procesos del proveedor de servicios deben adaptarse a las características del DP de la organización contratante.

- **Pruebas de servicios web** [34–36]: Si bien existen muchas herramientas para probar este tipo de arquitecturas, es difícil integrarlas en CD.

Los problemas mencionados representan dificultades para la implementación del proceso de pruebas en CD correctamente. Además, los mismos se encuentran relacionados entre sí, de tal manera, que la ocurrencia de uno produce directamente otro, o lo puede producir. Estas relaciones se muestran en la Fig. 2.

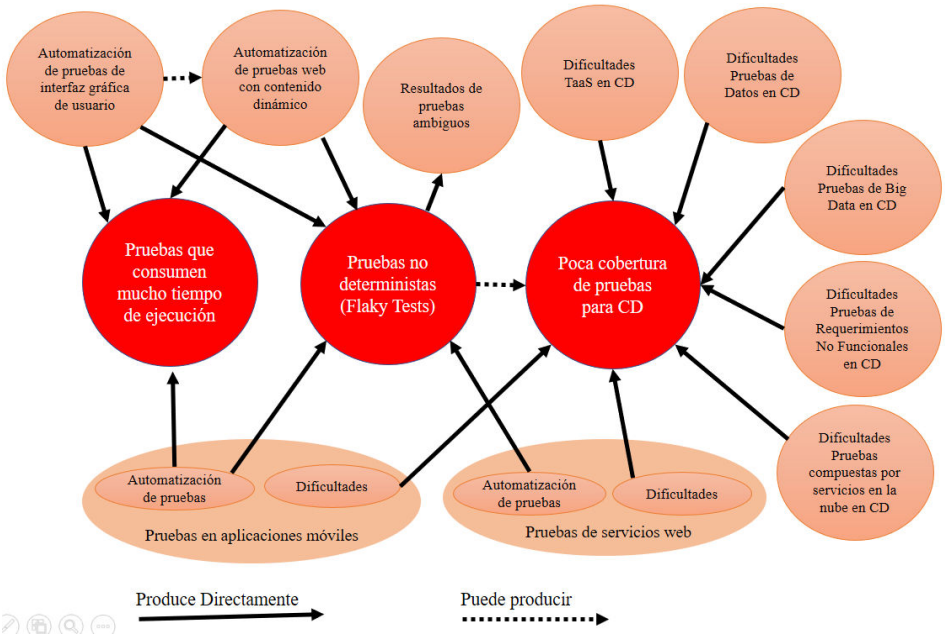


Fig. 2. Relación entre los problemas de pruebas en CD.

En la Fig. 2 puede apreciarse que los problemas más importantes son las ***pruebas que consumen mucho tiempo de ejecución*** y las ***pruebas automatizadas no deterministas***. Además, en base a los problemas relacionados con dificultades en la implementación de ciertos tipos de pruebas, se ha agregado un problema más: ***poca cobertura de pruebas en CD***. Estos factores han sido impedimentos para algunas organizaciones en la adopción de CD.

4 Pruebas Continuas

Las pruebas continuas o Continuous Testing (CT), es un término introducido por Edward Smith en el año 2000 [37], como un componente adicional de TDD en la ejecución de pruebas unitarias. Consistía en un proceso de ejecución de pruebas durante todo el día (continuamente), mientras se desarrollaba el código. Sin embargo, este concepto ha ido evolucionando con el correr del tiempo.

En al año 2003, Saff y Ernst [38], introdujeron el concepto de CT como una manera de reducir el tiempo que se pierde al ejecutar las pruebas unitarias. Para ello, entre los años 2004 y 2005 [39], proponen un plugin para el entorno de desarrollo integrado eclipse, el cual permite la ejecución de pruebas unitarias y de integración, mientras el desarrollador escribe código en su computadora. En el año 2010, este concepto se extiende a otros tipos de pruebas [40]: pruebas de especificación, pruebas de diseño, pruebas sobre el código fuente, pruebas funcionales, pruebas no funcionales, pruebas de instalación, pruebas de soporte y pruebas de mantenimiento.

Posteriormente, en el año 2011, en [31] se presenta CT para cloud computing. Los autores afirmaron que CT puede ser usado para realizar pruebas en aplicaciones SaaS. Además, como las aplicaciones están compuestas de servicios, CT puede ser llevado a cabo antes y después de la composición de estos servicios, incluso durante su ejecución como un sistema de monitoreo (Continuous Monitoring).

En el 2013, Google define CT como la "ejecución de cualquier tipo de prueba tan pronto como sea posible, para detectar cualquier tipo de defecto relacionado con un cambio realizado por el desarrollador" [41].

Entre los años 2015 y 2016, surgieron nuevos enfoques relacionados a CT. Erder y Pureur [42], consideran a CT como la utilización de técnicas de automatización para mejorar la velocidad de las pruebas mediante la integración de las etapas de desarrollo y pruebas. Para Moe et al. [43], CT consiste en probar el código fuente inmediatamente después de ser integrado en el repositorio, realizando regresiones. Virmani [44], expresa que CT consiste en automatizar cada uno de los casos de pruebas, ejecutándolos en cada integración sin intervención del usuario. Rossi et al. [45], presentan CT para desarrollo móvil, y se lo define como "el proceso de ejecutar todas las pruebas continuamente en un laboratorio de dispositivos móviles". Finalmente, Duvall et al. [46], afirman que "en un ambiente de CI, las pruebas deben ser ejecutadas continuamente".

El concepto de CT ha ido evolucionando con el correr de los años. En el comienzo, solo contemplaba la ejecución de pruebas unitarias de manera continua, especialmente en el entorno de trabajo del desarrollador. En la actualidad, no solo se aplica a las pruebas unitarias, sino a todo tipo de prueba que pueda ser automatizada (Ver Fig. 3).

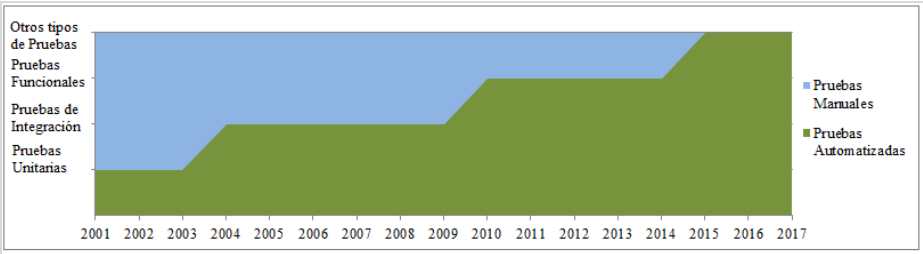


Fig. 3. Evolución de las pruebas automatizadas durante los años.

Esto podría indicar que la inclusión de las diferentes etapas de pruebas durante los años en las definiciones de CT, está relacionada con la aparición de herramientas y técnicas de automatización. Sin embargo, todas estas definiciones se basaron en el enfoque de Smith [37] y Saff y Ernst [38, 39], pero aún no existe una definición

formal para CT que englobe a todas las propuestas presentadas hasta la actualidad. De esta manera, tomando el análisis de las definiciones anteriores, podemos definir a CT como **el proceso de automatizar y ejecutar cualquier tipo de caso de prueba tan pronto como sea posible, para brindar a los desarrolladores retroalimentación rápida y así detectar defectos críticos antes de salir a producción**.

A partir de esta definición, y tomando soluciones existentes, es posible la construcción de un conjunto de procedimientos para atacar los problemas de la implementación del proceso de pruebas en entornos de CI, DC y CD.

5 Conclusiones y trabajo futuro

En este trabajo se presentó un relevamiento de una parte del estado del arte de la calidad del software en entornos ágiles, especialmente en aquellos que utilizan el enfoque de entrega continua. El mismo abarcó problemas comunes que afectan a la calidad, las relaciones que existen entre ellos, y las pruebas continuas.

Dentro de los problemas más comunes, se encuentran la ejecución de lotes de pruebas (regresiones) de larga duración, los resultados de pruebas no deterministas (flaky tests), y la baja cobertura de pruebas.

Por otro lado, diferentes autores afirman que las pruebas continuas son la clave para resolver los problemas de calidad en entrega continua. Así, se realizó un análisis de los diferentes enfoques que contemplan a las pruebas como continuas, y partiendo del mismo, se propuso una definición más concreta de pruebas continuas.

Como trabajo futuro se propone un análisis de las soluciones a estos problemas, para la elaboración de un modelo basado en las pruebas continuas. El mismo estará formado por un conjunto de procedimientos que una organización pueda seguir, para implementar el proceso de pruebas en entornos ágiles que buscan alcanzar exitosamente los enfoques de entrega continua.

Agradecimientos

Este trabajo se enmarca en dos proyectos de investigación: “Metodologías y herramientas emergentes para contribuir con la calidad del software” (PI 17F018 SCyT UNNE), y "Análisis e Implementación de tecnologías emergentes en sistemas computacionales de aplicación regional" (PI 17F017 SCyT UNNE).

Referencias

1. Cohn, M.: Succeeding with agile: Software development using scrum. Addison-Wesley Educational Publishers Inc, Upper Saddle River, NJ (2009).
2. Fowler, M.: Continuous Integration, https://martinfowler.com/articles/continuousIntegration.html.
3. Humble, J., Farley, D.: Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation. Pearson Professional, Upper Saddle River, NJ (2010).

4. Fowler, M.: Continuous Delivery, https://martinfowler.com/bliki/ContinuousDelivery.html.
5. Chen, L.: Continuous Delivery: Huge Benefits, but Challenges Too. IEEE Software. 32, 50–54 (2015).
6. Prusak, O.: Continuous testing: The missing link in the continuous delivery process, https://www.blazemeter.com/blog/continuous-testing-missing-link-continuous-delivery-process.
7. Laukkanen, E., Itkonen, J., Lassenius, C.: Problems, causes and solutions when adopting continuous delivery — A systematic literature review. Information and Software Technology. 82, 55–79 (2017).
8. Mascheroni, M.A., Irrazábal, E.: Continuous Testing and solutions for testing problems in Continuous Delivery: A Systematic Literature Review, (in press).
9. BlazeMeter: Continuous Testing in practice. Completing the Continuous Delivery Process, https://info.blazemeter.com/shift-left-testing, (2015).
10. Debbiche, A., Dienér, M., Svensson, R.B.: Challenges when adopting continuous integration: A case study. In: Product-Focused Software Process Improvement. pp. 17–32. Springer, Cham (2014).
11. Fitzgerald, B., Stol, K.-J.: Continuous software engineering and beyond: Trends and challenges. In: Proceedings of the 1st International Workshop on Rapid Continuous Software Engineering. pp. 1–9. ACM, New York, NY, USA (2014).
12. Claps, G.G., Berntsson Svensson, R., Aurum, A.: On the journey to continuous deployment: Technical and social challenges along the way. Information and Software Technology. 57, 21–31 (2015).
13. Fitzgerald, B., Stol, K.-J.: Continuous software engineering: A roadmap and agenda. Journal of Systems and Software. 123, 176–189 (2017).
14. Mäntylä, M.V., Adams, B., Khomh, F., Engström, E., Petersen, K.: On rapid releases and software testing: a case study and a semi-systematic literature review. Empir Software Eng. 20, 1384–1425 (2015).
15. Neely, S., Stolt, S.: Continuous Delivery? Easy! Just change everything (well, maybe it is not that easy). In: 2013 Agile Conference. pp. 121–128 (2013).
16. Chen, L.: Continuous Delivery: Overcoming adoption challenges. Journal of Systems and Software. 128, 72–86 (2017).
17. Brooks, G.: Team pace keeping build times down. Agile Conference. pp. 294–297 (2008).
18. Laukkanen, E., Lehtinen, T.O.A., Itkonen, J., Paasivaara, M., Lassenius, C.: Bottom-up adoption of continuous delivery in a stage-gate managed software organization. In: Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. pp. 45:1–45:10. ACM, New York, NY, USA (2016).
19. Cannizzo, F., Clutton, R., Ramesh, R.: Pushing the boundaries of testing and continuous integration. In: Agile 2008 Conference. pp. 501–505 (2008).
20. Leppänen, M., Mäkinen, S., Pagels, M., Eloranta, V.P., Itkonen, J., Mäntylä, M.V., Männistö, T.: The highways and country roads to continuous deployment. IEEE Software. 32, 64–72 (2015).
21. Debbiche, A., Dienér, M.: Assessing challenges of continuous integration in the context of software requirements breakdown: A case study, (2014).
22. Alégroth, E., Feldt, R., Ryrholm, L.: Visual GUI testing in practice: challenges, problems and limitations. Empir Software Eng. 20, 694–744 (2015).
23. Borjesson, E., Feldt, R.: Automated system testing using visual gui testing tools: A comparative study in industry. In: Verification and Validation 2012 IEEE Fifth International Conference on Software Testing. pp. 350–359 (2012).
24. Pradhan, L.: User interface test automation and its challenges in an industrial scenario, (2012).
25. Suwala, P.: Challenges with modern web testing, (2015).

26. Ståhl, D., Bosch, J.: Automated software integration flows in industry: A multiple-case study. In: Companion Proceedings of the 36th International Conference on Software Engineering. pp. 54–63. ACM, New York, NY, USA (2014).
27. Huizinga, D., Kolawa, A.: Automated defect prevention: Best practices in software management. John Wiley & Sons, Hoboken, N.J (2007).
28. Garg, N., Singla, S., Jangra, S.: Challenges and techniques for testing of big data. Procedia Computer Science. 85, 940–948 (2016).
29. Muşlu, K., Brun, Y., Meliou, A.: Data debugging with continuous testing. In: Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering. pp. 631–634. ACM, New York, NY, USA (2013).
30. Muccini, H., Francesco, A.D., Esposito, P.: Software testing of mobile applications: Challenges and future research directions. In: 2012 7th International Workshop on Automation of Software Test (AST). pp. 29–35 (2012).
31. Tsai, W.T., Zhong, P., Balasooriya, J., Chen, Y., Bai, X., Elston, J.: An approach for service composition and testing for cloud computing. In: 2011 Tenth International Symposium on Autonomous Decentralized Systems. pp. 631–636 (2011).
32. Barquet, L.A., Tchernykh, A., Yahyapour, R.: Performance Evaluation of Infrastructure as Service Clouds with SLA Constraints. Computación y Sistemas. 17, 401–411 (2013).
33. Tilley, S., Floss, B.: Hard problems in software testing: Solutions using testing as a service. Morgan & Claypool Publishers (2014).
34. Mascheroni, M.A., Irrazábal, E.: A Design Pattern Approach for RESTful tests: A case study. Obras colectivas en ciencias de la computación. Cali, Colombia (2018).
35. Savchenko, D.I., Radchenko, G.I., Taipale, O.: Microservices validation: Mjolnirr platform case study. 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). pp. 235–240 (2015).
36. Mascheroni, M.A., Irrazábal, E.: Framework para la creación y ejecución de pruebas automatizadas sobre servicios REST. Presented at the XXII Congreso Argentino de Ciencias de la Computación (CACIC). (2016).
37. Smith, E.: Continuous testing. Presented at the 17th International Conference on Testing Computer Software (2000).
38. Saff, D., Ernst, M.D.: Reducing wasted development time via continuous testing. 14th International Symposium on Software Reliability Engineering, pp. 281–292 (2003).
39. Saff, D., Ernst, M.D.: Continuous Testing in Eclipse. Electronic Notes in Theoretical Computer Science. 107, 103–117 (2004).
40. Burgin, M., Debnath, N.: Intelligent testing systems. In: 2010 World Automation Congress. pp. 1–6 (2010).
41. Penix, J.: Large-scale test automation in the cloud (Invited Industrial Talk). Presented at the IEEE 34th International Conference on Software Engineering (ICSE) (2012).
42. Erder, M., Pureur, P.: Continuous Architecture: Sustainable Architecture in an Agile and Cloud-Centric World. Morgan Kaufmann, Waltham, MA (2015).
43. Moe, N.B., Cruzes, D., Dybå, T., Mikkelsen, E.: Continuous software testing in a globally distributed project. In: 2015 IEEE 10th International Conference on Global Software Engineering. pp. 130–134 (2015).
44. Virmani, M.: Understanding DevOps bridging the gap from continuous integration to continuous delivery. In: Fifth International Conference on the Innovative Computing Technology (INTECH 2015). pp. 78–82 (2015).
45. Rossi, C., Shibley, E., Su, S., Beck, K., Savor, T., Stumm, M.: Continuous Deployment of Mobile Software at Facebook (Showcase). In: ACM SIGSOFT: International Symposium on the Foundations of Software Engineering (FSE 2016) (2016).
46. Duvall, P., Matyas, S., Glover, A.: Continuous integration: Improving software quality and reducing risk. Pearson Professional, Upper Saddle River, NJ (2007).

Exploring Specification Pattern based Behavioral Synthesis with Scenario Clauses *

Asteasuain Fernando[1,2], Federico Calonge[1], and Manuel Dubinsky[1]

[1] Universidad Nacional de Avellaneda, BsAs, Argentina,
fasteasuain,fcalonge,mdubinsky@undav.edu.ar,
www.undav.edu.ar
[2] Universidad Abierta Interamericana, CAETI
BsAs, Argentina

Abstract. The Software Engineering community has identified behavioral specification as one of the main challenges to be addressed for the transference of formal verification techniques such as model checking. In particular, expressivity of the specification language is a key factor, especially when dealing with open systems and controllability of events. In this work we present an extension of the FVS language to denote behavior in open systems. By relying on an existing behavioral synthesis technique based on the specification patterns we show how FVS specification can be used as input to automatically build a controller from its specification.

Keywords: Open Systems, Behavioral Specifications, Synthesis

1 Introduction

Early specification of behavior has been pinpointed by the community as one of the main problems to be addressed to consolidate the transference of software formal validation and verification techniques as model checking [5] from the academic to the industrial world [11, 10]. On the other side, the increasing demand of Open Systems [7, 10, 12] calls for the creation of tools that assists the software engineer in the complex task of specifying and describing the expected behavior of the system. There is an natural challenge involved when dealing with Open Systems since actions beyond the control of the system must be considered, in contrast to systems known as *closed* where all the events to occur are handled entirely by the system. Open Systems interact with an environment which generates events (non controllable by the system) which may impact in its behavior, which constitutes an affect known as *Controllability.*

Controllability has been addressed methodologically and algorithmically from the synthesis behavior of controllers. Synthesis behavior can be seen as an automated procedure to obtain a correct-by-construction reactive system from its temporal logic specification [12]. In the case of reactive synthesis, an implementation is typically given as an automaton that accepts input from the environment

* This work was partially funded by UNDAVCYT 2014 and UAI-CAETI

(e.g., from sensors) and produces the system's output (e.g., on actuators). By construction the input and output assignments of every infinite run of the automaton satisfy the specification it was synthesized from [12]. In order to reduce the time complexity of the algorithms involved work in [4] suggested the General Reactivity of Rank 1 (GR(1)) fragment of LTL, which has an efficient polynomial time symbolic synthesis algorithm. GR(1) is a strict assume/guarantee subset of LTL, comprised of constraints for initial states, safety propositions over the current and successor state and assertions about what should hold infinitely often also known as justice constraints. A GR(1) synthesis problem is defined as a game between a system player and an environment player [4]. Efficient symbolic algorithms for GR(1) realizability checking and controller synthesis have been presented in [4].

However, these approaches have some limitations regarding the specification language used. Most of them are based on temporal logics such as LTL (Linear Temporal Logic) and some extensions as fluents [9]. The expressivity of these notations has been challenged by the community [9, 2, 14, 6, 15] so there is a need to develop more expressive specification languages.

Given this context in this work we explore the FVS (Feather Weight Visual Scenarios) [2, 1] specification language in the context of Open Systems and behavior synthesis. FVS is a declarative language based on graphical scenarios and features a flexible and expressive notation with clear and solid language semantics. FVS expressivity is a distinguished characteristic among declarative approaches since it is able to denote ω-regular properties, being for example, more expressive than LTL (Linear Temporal Logic) [2]. In [1] all the specification patterns [8] were modeled in FVS, and their specification was compared against other notations. The results showed that FVS specification turn out to be more succinct and easier to manipulate and validate. Furthermore, a tool named GTxFVS was developed giving support to all FVS's features [3].

The first step of FVS into the world of Open Systems and Behavior Synthesis is based on a technique introduced in [12], which is guided by the usage of specification patterns [8]. The mentioned paper presents an automated, sound and complete translation of most of the the specification patterns [8] to the GR(1) form. Although at this point this open flavour of FVS works only for systems whose behavior can be described using exclusively the specification patterns, the kind of properties covered by these patterns is useful enough to denote most of the common behavior [8]. What is more, the technique described in [12] does not include three specification patterns which are addressed in this work.

Summing up, the contributions of this work can be stated as:

- We introduce a simple extension of FVS to consider Open Systems.
- We rely on an existing technique based on the specification patterns so that FVS specifications can be used to synthesise behavior and automatically build a controller from its specification.
- Our approach is shown in action by modeling an attractive case of study.
- We present a version of our tool GTxFVS including these new features.
- We incorporate three specifications patterns that were left out in the original technique.

The rest of the paper is structured as follows. Section 2 briefly introduces FVS. Section 3 describes FVS in the context of Open Systems and behavior synthesis while section 4 shows our approach in action by modeling a case of study. Section 5 details how we incorporate the three missing patterns. Finally, section 6 introduces some threats to validity and also discuss future and related work whereas section 7 presents the conclusions of this work.

2 Background

In this section we will informally describe the standing features of FVS [2]. The reader is referred to [2] for a formal characterization of the language. FVS is a graphical language based on scenarios. Scenarios are partial order of events, consisting of points, which are labeled with a logic formula expressing the possible events occurring at that point, and arrows connecting them. An arrow between two points indicates precedence of the source with respect to the destination: for instance, in Figure 1-a A-event precedes B-event. We use an abbreviation for a frequent sub-pattern: a certain point represents the next occurrence of an event after another. The abbreviation is a second (open) arrow near the destination point. For example, in Figure 1-b the scenario captures the very next B-event following an A-event, and not any other B-event. Conversely, to represent the previous occurrence of a (source) event, there is a symmetrical notation: an open arrow near the source extreme. For example, in Figure 1-c the scenario captures the immediate previous occurrence of a B-event from the occurrence of the A-event, and not any other B-event. Events labeling an arrow are interpreted as forbidden events between both points. In Figure 1-d A-event precedes B-event such that C-event does not occur between them. FVS features aliasing between points. Scenario in 1-e indicates that a point labeled with A is also labeled with *A and B*. It is worth noticing that A-event is repeated on the labeling of the second point just because of FVS formal syntaxes [2] . Finally, two special points are introduced as delimiters to denote the beginning and the end of an execution. These are shown in Figure 1-f.

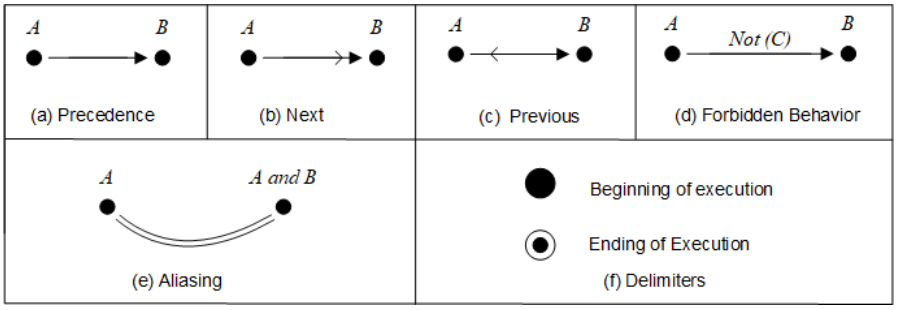


Fig. 1. FVS Basic Features

We now introduce the concept of **FVS rules**, a core concept in the language. Roughly speaking, a rule is divided into two parts: a scenario playing the role of an antecedent and at least one scenario playing the role of a consequent. The intuition is that whenever a trace matches" a given antecedent scenario, then it must also match at least one of the consequents. In other words, rules take the form of an implication: an antecedent scenario and one or more consequent scenarios. Graphically, the antecedent is shown in black, and consequents in grey. Since a rule can feature more than one consequent, elements which do not belong to the antecedent scenario are numbered to identify the consequent they belong to. Two examples are shown in Figure 2 modeling the behavior of a client-server system. The rule in left margin of Figure 2 establishes that every request received by a server must be answered, either accepting the request (consequent 1) or denying it (consequent 2). The rule at the right margin of Figure 2 dictates that every granted request must be logged due to auditing requirements.

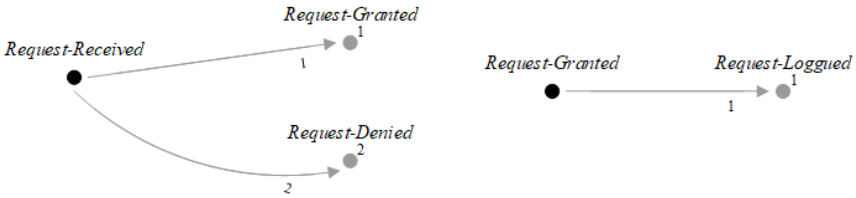


Fig. 2. FVS Rules examples

3 Open FVS and Behavior Synthesis

We now briefly introduce a simple extension to the FVS language to handle open systems. This is achieved by introducing a new type of events, which are considered as environment events not controlled by the system. Semantics of FVS (see [2]) is based on the notion of morphisms between scenarios. In particular morphisms between scenarios allow to determine whether a system trace satisfies a rule and semantics of the language is defined by the set of traces that fulfills all the specified rules. We first define the concept of scenarios considering non controllable events. The set of events is divided into controllable events and non controllable events. An FVS scenario including non controllable events in their alphabet is described by the following definition

Definition 31 (FVS Scenario) *An FVS* scenario *is a tuple* $\langle \Sigma, P, \ell, \equiv, \not\equiv, <, \gamma \rangle$, *where:*
S1: *Σ is a finite set of propositional variables standing for types of events such that $\Sigma = \Sigma_c \bigcup \Sigma_{uc}$ where Σ_c represents controllable events and Σ_{uc} non controllable events*

S2: *P is a finite set of points;*
S3: *$\ell : P \to \mathcal{PL}(\Sigma)$ is a function that labels each point with a given formula;*
S4: *$\equiv \ \subseteq P \times P$ is an equivalence relation;*
S5: *$\not\equiv \ \subseteq P \times P$ is an asymmetric relation among points;*
S6: *$<\subseteq (P \uplus \{\mathbf{0}\} \times P \uplus \{\infty\}) \setminus \{\langle \mathbf{0}, \infty \rangle\}$ is a precedence relation between points, where $\mathbf{0}$ and ∞ represent the beginning and the end of execution, respectively;*
S7: *$\gamma : (\not\equiv \cup <) \to \mathcal{PL}(\Sigma)$ assigns to each pair of points, related by precedence or separation, a formula which constrains the set of events occurrences that may occur between the pair.*

We now formally define *morphisms* between scenarios. Intuitively, we would like to obtain a matching between scenarios ,i.e., a **mapping** between their points exhibiting how a scenario "specializes" another one.Given that non controllable events values are considered as inputs to the system the usual morphism definition specified in [2] can be applied.

Definition 32 (Morphism) *Given two scenarios $\mathcal{S}_1$, $\mathcal{S}_2$ (assuming a common universe of event propositions), and f a total function between P_1 and P_2 we say that f is a morphism from $\mathcal{S}_1$ to $\mathcal{S}_2$ (denoted $f : \mathcal{S}_1 \to \mathcal{S}_2$) iff*
M1: *$\ell_2(a) \Rightarrow \ell_1(\mathsf{p})$ is a tautology for all $\mathsf{p} \in P_1$ and all $a \in P_2$ such that $a \equiv_2 f(p)$;*
M2: *$\gamma_2(f(\mathsf{p}), f(\mathsf{q})) \Rightarrow \gamma_1(\mathsf{p}, \mathsf{q})$ is a tautology for all $\mathsf{p}, \mathsf{q} \in P_1$;*
M3: *if $\mathsf{p} \equiv_1 \mathsf{q}$ then $f(\mathsf{p}) \equiv_2 f(\mathsf{q})$ for all $\mathsf{p}, \mathsf{q} \in P_1$;*
M4: *if $\mathsf{p} \not\equiv_1 \mathsf{q}$ then $f(\mathsf{p}) \not\equiv_2 f(\mathsf{q})$ for all $\mathsf{p}, \mathsf{q} \in P_1$;*
M5: *if $\mathsf{p} <_1 \mathsf{q}$ then $f(\mathsf{p}) <_2 f(\mathsf{q})$ for all $\mathsf{p}, \mathsf{q} \in P_1$.*

3.1 FVS specifications as input to a synthesis scheme

Work in [2] describes a tableau algorithm which translates FVS scenarios into Büchi automata. Also related to FVS, work in [1] shows how FVS is expressive enough to denote all the specification patterns introduced by [8]. On the other hand, work in [12] proposes a synthesis scheme for a set of LTL (linear temporal logic) formulas, the set of formulas describing the behavior of all the specification patterns. These formulas were translated into General Reactivity of Rank 1 (GR(1)) fragment of LTL (due to performance and complexity reasons) and then synthesised following an usual game based strategy.

Given this context, we translated all the FVS scenarios describing all the specification patterns. The resulting Büchi automata were then used to feed the synthesis scheme described earlier. In this way, a controller can be automatically built synthesizing the behavior of a system using as input FVS specifications. Although this open flavour of FVS works only for system whose behavior can be described using only the specification patterns, the kind of properties covered by these patterns is an attractive portion of the cake [8]. It is worth mentioning that this version of FVS is fully available in the current status of GTxFVS[3], the software tool which implements all the features of the language.

4 Case of Study

In this section we analyze the case of study introduced in [12], the "Lego Forklift" example. The behavior of the model is given in section 4.1 while the synthesis procedure is detailed in section 4.2.

4.1 The "Lego Forklift" example

The behavior of the system is described in [12]. In few words, the forklift has three sensors: one sensor to determine whether it is at a station and two distance sensors to detect obstacles and cargo. It also has three motors, to turn the left and right wheels and to lift the fork. Values read by the sensors are provided as inputs to component ForkliftController and its outputs are commands controlling the motors. The properties modeling the behavior of the system is described using assumptions and guarantees based on the scheme introduced in [12]. In particular, three guarantees and two assumptions properties are specified. The first guarantee property is a simple safety one: *if the forklift detects an obstacle, both motors are stopped*. The second one is a liveness property: *cargo will always eventually be delivered*. Finally, the third guarantee property establishes that *the forklift has to leave its pick-up station between lifting and dropping cargo*. Rules in figure 3 denote the behavior of these three properties.

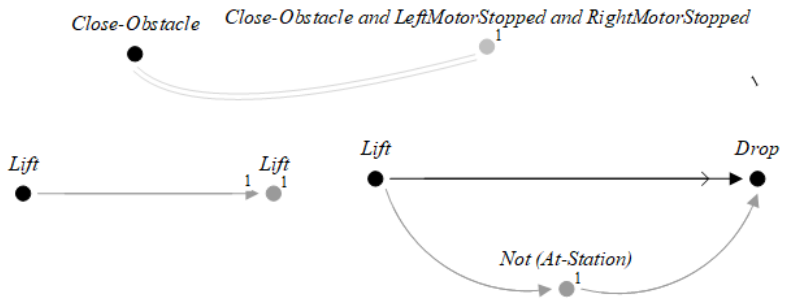


Fig. 3. Guarantees properties for the Lift Controller

The rule in the upper part of figure 3 says that every time the *Close-Obstacle* event occurs, then motors left and right should be stopped (denoted by the occurrence of the events *LeftMotorStopped* and *RightMottorStopped*). Note that these events are modeled as occurring simultaneously as it is indicated in the systems requirements specified in [12]. The second rule (in the left margin of the lower part of figure 3) simply states that *Lift* event will always eventually occurs, a classical liveness property. Finally, the rule in the right margin of the lower part of figure 3 demands the forklift to leave the station (modeled as the

negation of the *At-Station* event) between the occurrence of the *Lift* event (the forklift lifted an item) and the occurrence of the *Drop* Event (the forklift dropped an item).

The forklift specification is completed with two assumptions on the environment. One assumption is that going forward with both motors will lead to reaching a station unless the motors are not going forward anymore. As it is noted in [12] in order to satisfy this assumption an adversary environment can prevent the forklift from reaching a cargo station by presenting obstacles forcing the forklift to stop. Hence, an additional assumption for a well-behaved environment is added: in the given setting it is reasonable to expect that between two stations, the forklift may be blocked by obstacles at most twice.

Rules in figure 4 model these two assumptions. The rule in the top of the figure tackles the first assumption: once left and right motors are on, then the forklift will reach an station if *MotorStopped* event do not occur. The second rule focuses on the two times blocked maximum restriction. The rule contains four possible consequents once the forklift has left the station and the first obstacle occurs. Consequent 1 deals with the situation that no other obstacle is detected until the forklift arrives at the station. Consequent 2 do consider the occurrence of a second obstacle, but no other obstacle must occur until the station is reached. Consequents three and four follows a similar analysis but considering that the system may be stopped for some reason before reaching an station. This is why the *End* point of traces is included. So, before reaching a new station or the system is stopped the forklift may be blocked by obstacles at most twice.

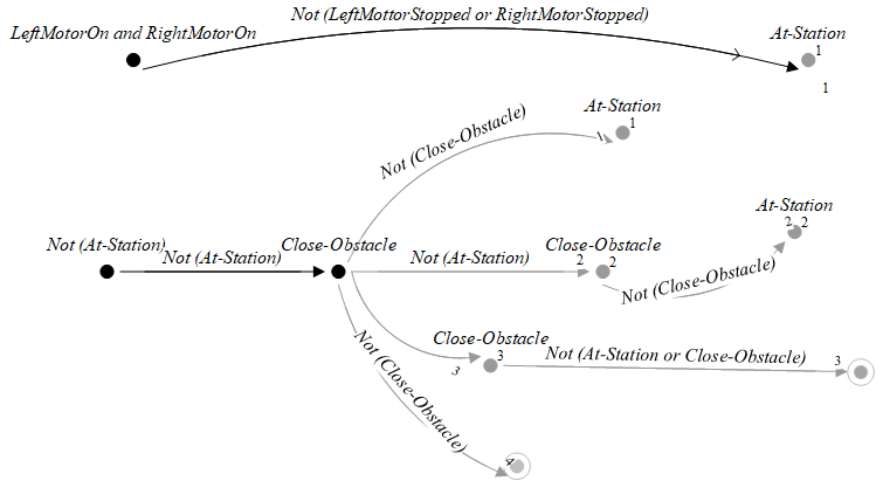


Fig. 4. Environment assumptions properties for the Lift Controller

4.2 Synthesising behavior from FVS scenarios

By relying on the technique introduced in [12] as detailed in section 3.1 we build a controller for the *ForkLift* case study taking as input FVS scenarios. In particular, three properties of the *Forklift* example matches with three specification pattern: the third guarantee property corresponds to the Existence pattern with *Between Q and R* scope, the first assumption with the Response event with *Global* scope and and second assumption is an instance of the Bounded Existence pattern with *After Q until R* scope.

5 The three missing patterns

Work in [12] proposes a synthesis scheme based on the GR(1) fragment of LTL (Linear Temporal Logic) for most of the specification patterns and scopes introduced by [8]. However, three patterns are not currently supported by their technique. These patterns are the followings ones: *Precedence pattern with After Q scope*, *Response Chain pattern (with one stimuli and two responses) with After q until r scope* and *the Constrained Chain pattern (responses s, t without z responds to the p stimuli event) with After q until r scope*. These patterns are not included since no Deterministic Büchi automata exists to describe their behavior and the mentioned technique requires this kind of automaton as input.

We propose an alternative approach to circumvent this issue. The FVS scenarios for these patterns (the FVS scenarios for all the specification patterns are shown in [1]) can be later translated into non determinist Büchi automata. By employing an advanced tool for manipulating diverse kinds of automata named GOAL [16] we translated these automata into *Deterministic Rabin automata*. Since synthesis algorithms are also incorporated into the GOAL tool using Rabin automata as input, we could incorporate these patterns so as to cover all the specification patterns. Furthermore, since our tool GTxFVS [3] can interact with the GOAL tool, all the patterns and their synthesized behavior are available for the software engineer.

Being able to handle all the specification patterns including these three patterns is important in terms of completeness and expressiveness but it also has its drawbacks. Most of the algorithms involved in automata manipulation and synthesis depend on the size of the automata and have performance and complexity issues. In addition, the synthesis procedure proposed in [12] introduces extra variables and therefore the the costs are higher.

The problem is exacerbated with these patterns using *Deterministic Rabin* automata since their size is really important. For example, the Rabin automaton for the *Response Chain pattern (with one stimuli and two responses) with After q until r scope* consists of 94 states and 1003 transitions. These automata can be simplified, but their size remains important. The simplified automaton for the latter pattern consists of 63 states and 611 transitions. Table 1 describes the automata size for these three patterns. The abbreviations s and t stands for states and transitions respectively.

Table 1. Automata size Complexity for the missing patterns

Pattern	Not Det. Buchi	Det. Rabin	Simplified Rabbin
Precedence after	5s and 13t	11s and 41t	11s and 41t
Response Chain after until	9s and 37t	94s and 1003s	63s and 611t
Constrained Chain after until	8s and 34t	47s and 611t	35t and 411t

Using the proposed technique to incorporate these patterns (using Rabin automata and the GOAL tool) we add an extra assumption for the ForkLift example and obtained a new controller that includes this new behavior: *after leaving an station, all the Drop events must be preceded in time with an Lift event*. This behavior is an instance of the *Precedence pattern with After Q scope*. This gain in expressiveness come with an cost in terms of performance due to the size of the involved automata. The time consumed to obtain this new controller took eight times more than the previous one. As it is mentioned in [8, 12] these patterns are far from being the most used ones. However, we believe it is desirable being able to express them despite the fact they are time consuming.

6 Some Threats to Validity, Future and Related Work

In this section we briefly discuss some issues that can be seen as threats to validity to the results introduced in this work.

First of all, performance and complexity of the algorithms used must be addressed. One way to optimize this problem is trying to reduce the overhead introduced in the pipe of tools used to obtain a controller using the mentioned synthesis scheme, which we aim to cover in future work. A second issue is related to FVS as an open system language specification. Synthesis scheme is built upon the work introduced in [12], which only covers properties included in the specification patterns. In this sense, to obtain a complete synthesis scheme for properties beyond the patterns is clearly an appealing challenge regarding future work. Despite these facts, we believe that the obtained results are promising enough to consider FVS as an attractive alternative to describe behavior in Open Systems.

Regarding related work several approaches can be mentioned. To begin with, there exist several specifications graphical languages based on events like FVS. For example, TimeEdit [15] or Graphical Interval Logic (GIL) [6]. However, these languages are not focused on modeling behavior in Open Systems.

Work in [9] uses the concept of fluent to relate occurrence of events and predicate about systems behavior. A fluent represents an ongoing behavior, with a set of starting and ending events. We believe there is a possible contribution combining fluents and FVS scenarios for specifying behavior in Open Systems. Finally, GR(1) synthesis has been used and extended in different contexts and for different application domains [13, 7]. However, we consider that FVS expressivity can be a distinguishable feature among these and other similar approaches.

7 Conclusions

In this work we present an open system flavour of FVS. By relying on a synthesis technique proposed by [12] we showed how FVS specification can be used as input to automatically build a controller from its specification. This work constitutes an early first step for FVS in the open systems specifications since our approach currently applies for those properties denoted by the specification patterns. As it was mentioned in Section 6 we would like to augment the kind of properties that can be expressed in future work. We also incorporated in our approach three specification patterns that were not included in [12]. This gain in expressibility come with a cost in terms of complexity and computing time. In this sense, we believe expressivity is a key factor when denoting behavior in early stages.

References

1. F. Asteasuain and V. Braberman. Specification patterns: formal and easy. *IJSEKE*, 25(04):669–700, 2015.
2. F. Asteasuain and V. Braberman. Declaratively building behavior by means of scenario clauses. *Requirements Engineering*, 22(2):239–274, 2017.
3. F. Asteasuain and F. Tarulla. Exploring architectural model checking with declarative specifications. In *CACIC*, 2017.
4. R. Bloem, B. Jobstmann, N. Piterman, A. Pnueli, and Y. Sa'Ar. Synthesis of reactive (1) designs. 2011.
5. E. M. Clarke, O. Grumberg, and D. Peled. *Model checking*. MIT press, 1999.
6. L. K. Dillon, G. Kutty, L. E. Moser, P. M. Melliar-Smith, and Y. S. Ramakrishna. A graphical interval logic for specifying concurrent systems. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 3(2):131–165, 1994.
7. N. DIppolito, V. Braberman, N. Piterman, and S. Uchitel. Synthesising non-anomalous event-based controllers for liveness goals. *ACM Tran*, 22(9), 2013.
8. M. Dwyer, M. Avrunin, and M. Corbett. Patterns in property specifications for finite-state verification. In *ICSE*, pages 411–420, 1999.
9. D. Giannakopoulou and J. Magee. Fluent model checking for event-based systems. In *ACM SIGSOFT*, volume 28, pages 257–266. ACM, 2003.
10. I. Krka, Y. Brun, G. Edwards, and N. Medvidovic. Synthesizing partial component-level behavior models from system specifications. In *ESEC-FSE*.
11. A. V. Lamsweerde. Goal-oriented requirements engineering: A guided tour. In *RE*, 2001.
12. S. Maoz and J. O. Ringert. Synthesizing a lego forklift controller in gr (1): a case study. *arXiv preprint arXiv:1602.01172*, 2016.
13. S. Maoz and Y. Saar. Assume-guarantee scenarios: Semantics and synthesis. In *MODELS*, pages 335–351. Springer, 2012.
14. P. Pelliccione, P. Inverardi, and H. Muccini. Charmy: A framework for designing and verifying architectural specifications. *IEEE TSE*, 35(3):325–346, 2009.
15. M. H. Smith, G. J. Holzmann, and K. Etessami. Events and constraints: A graphical editor for capturing logic requirements of programs. In *Re*.
16. Y.-K. Tsay, Y.-F. Chen, M.-H. Tsai, K.-N. Wu, and W.-C. Chan. Goal: A graphical tool for manipulating büchi automata and temporal formulae. In *TACAS*, pages 466–471. Springer, 2007.

TG Project Management: Seguimiento y documentación de actividades orientada al usuario desarrollador.

Valentín Peluso, Ariel Pasini[1], Nicolás Galdámez, Luciano Marrero[2]

Instituto de Investigación en Informática LIDI (III-LIDI)*
Facultad de Informática – Universidad Nacional de La Plata 50 y 120 La Plata Buenos Aires
*Centro Asociado Comisión de Investigaciones Científicas de la Pcia. de Bs. As. (CIC)

{apasini, ngaldamez, lmarrero}@lidi.info.unlp.edu.ar
valentinpeluso@gmail.com

Abstract.: Los proyectos de software actuales no experimentan las mismas necesidades que los proyectos de hace algunos años. La evolución del mercado actual requiere nuevos tiempos de respuesta ante soluciones informáticas. Realizar una buena gestión del proyecto no garantiza su éxito, pero realizar una mala gestión seguramente lleve a su fracaso. Para que un proyecto resulte exitoso es necesario que su gestión y administración sean eficientes, para ello es fundamental la planificación, la experiencia, la buena comunicación y además es importante contar con herramientas de apoyo durante su gestión. Hoy en día se dispone de una gran variedad de herramientas para la gestión de proyectos, entre ellas, algunas muy populares como Trello y Jira. Ambas herramientas al intentar cubrir todo tipo de proyecto carecen de ciertos aspectos importantes, agilidad y facilidad de uso, pero poca estructura (Trello) o viceversa (Jira). La elección de herramientas en un proyecto debe ser un equilibrio entre las características del proyecto y las características de las herramientas elegidas con el objetivo de no modificar las reglas de negocio. El propósito de TG Project Management es nutrir el mercado actual con una herramienta orientada al usuario desarrollador, mejorando así, ciertos aspectos como el seguimiento del avance del proyecto, el control de tareas realizadas por el equipo, la estimación de esfuerzos, la organización del cronograma, etc. y una interacción fluida con herramientas de desarrollo colaborativo como Github.

Keywords: Ingeniería de Software, gestión de proyectos de software, metodologías ágiles, gestores de proyectos.

[1] ORCID: 0000-0002-4752-7112
[2] ORCID: 0000-0002-7507-1562

1 Introducción

En la Ingeniería de Software la gestión de proyectos de software es una temática de estudio e innovación constante. Gestionar un proyecto involucra la aplicación del conocimiento, habilidades, herramientas y técnicas a las actividades necesarias para alcanzar los objetivos deseados. En la gestión de proyectos existen dos tipos de procesos bien distinguidos: administración general y administración de proyecto. Si bien poseen aspectos en común, no deben ser confundidos. La administración general está pensada para un sistema de gestión de una organización a largo plazo y sin un límite conocido, en cambio, la administración de proyectos es finita y con objetivos específicos, que una vez cumplidos determinan su fin [1].

Administrar un proyecto de software implica atender diferentes procesos de gestión (alcance, recursos, cronograma, etc.), para asegurar que todo se ejecute de acuerdo con lo definido y conseguir los objetivos propuestos. Lograr el éxito de un proyecto involucra poner énfasis en varias cuestiones, desde la elección del personal que participará en el proyecto, pasando por la metodología de proceso aplicada, hasta las herramientas técnicas de apoyo que se utilizaran. Contar con personal de experiencia ligado al conocimiento y a las buenas prácticas, hace que su rendimiento mejore con respecto a proyectos previos. La metodología de gestión de un proyecto es un elemento clave para el conjunto de las actividades que involucra, elegir erróneamente la metodología a utilizar es costoso y puede llevar a su fracaso. Una metodología debe reflejar el tamaño, la duración y la complejidad del proyecto, como así también, adaptarse a la industria, la cultura y al nivel de madurez de la organización.

Durante el ciclo de vida de un proyecto se pueden utilizar diversas herramientas de apoyo y para diversos objetivos. Las herramientas para la gestión del proyecto (Trello y Jira) y las herramientas para el control de versiones (GIT) son el foco de estudio para este artículo. El uso de herramientas durante la gestión de proyectos puede enriquecer y potenciar la capacidad de los usuarios, como también asistir en el avance del proyecto, gestionar tareas del personal, realizar estimaciones, mejorar la comunicación con el cliente y mantener una documentación actualizada [1], [2] y [3].

En la sección dos se describen los dos desafíos más importantes de la gestión de los proyectos a gran escala, la administración de los recursos humanos y de la distribución de tareas entre los recursos. En la siguiente sección los inconvenientes encontrados en los principales gestores de proyectos. En la sección TG Project Management, se describe la herramienta propuesta para cubrir las falencias planteadas en el capítulo anterior. Por último las conclusiones y trabajos futuros.

2 Gestión en proyectos complejos y de amplio alcance.

2.1 Contexto

Actualmente la evolución de la industria del software crece a pasos agigantados. Se prioriza la agilidad y flexibilidad en el desarrollo, el producto se construye y al mismo tiempo se modifican e introducen requerimientos, los cuales cambian para adaptarse a un contexto versátil. El cliente generalmente posee una visión clara del proyecto, pero

la innovación y la velocidad a la que evoluciona el entorno de negocio, no le permite prever como será en detalle el resultado final. Se debe actuar rápidamente y utilizar tácticas orientadas a la respuesta ágil y flexible. Los proyectos de software se vuelven cada vez más complejos y a medida que aumenta la complejidad las estrategias deben actualizarse. El grado de adaptación a estos cambios es lo que permite mejorar y crecer sin perder la calidad de un producto para que perdure en el tiempo [1], [4] y [5].

2.2 Gestión del personal del proyecto

En un proyecto de software participan distinto tipos de usuarios, algunos vinculados directamente y otros de forma indirecta. Cuanto mayor mano de obra se necesita, mayor será el control a realizar, además, no siempre es sencillo contar con personas altamente capacitadas, lo habitual es que los más experimentados formen a los más novatos, para lograr esto es fundamental la buena comunicación de todo el equipo de trabajo. Es importante que los participantes de un proyecto se interesen en el mismo y que sepan trabajar en equipo, aunque el desempeño de una persona depende de muchos factores (motivos personales, ámbito laboral, etc.), es fundamental que el personal sienta la importancia de cada una de sus actividades y el impacto de estas en el proyecto, seguramente, así se aumenta la productividad.

Es común que durante el ciclo de vida de un proyecto, cambie alguno de sus requerimientos o sufra bajas y altas de personal. Estas situaciones impactan directamente sobre el equipo de trabajo. Se hace necesario tener estrategias y herramientas de gestión para absorber tales cambios con el menor costo posible [1] y [7].

Los grandes proyectos de software involucran un gran número de integrantes, lo que dificulta la organización, coordinación y asignación del trabajo a realizar. Es por esto por lo que la estrategia más utilizada es el trabajo en equipos organizados de forma autónoma para cumplir un objetivo dentro del proyecto, aun así, es importante contar con herramientas y estrategias de apoyo para gestionar la comunicación y la convivencia de los grupos, disminuyendo así la probabilidad de riesgos durante el proyecto.

2.3 Gestión de las tareas

En algunas ocasiones se puede confundir el concepto de tarea con el de proyecto, ya que ambos poseen atributos en común, personas, recursos escasos (tiempo, dinero, etc.), necesitan ser planificados y requieren de continuo control. En general las tareas son repetitivas y se mantienen en el tiempo mientras que los proyectos son únicos y temporales.

Es importante que un proyecto posea un director de proyecto, quién será el responsable del control y administración de todas las actividades durante todo el proceso de gestión. El director del proyecto deberá ser el integrador y facilitador, tanto en lo que respecta al proyecto como al equipo humano. Independientemente de la metodología de procesos elegida, tradicional o ágil, la dirección de los equipos de

trabajo suele ser piramidal. Cada equipo cuenta con un líder interno, quién conoce mejor el proyecto, las reglas de negocios y además es el responsable de la interacción con el resto del personal externo al equipo.

En proyectos de gran escala, es fundamental la supervisión, comunicación y articulación de los líderes de equipos, tanto con sus dirigidos como con sus superiores y pares. Contar con estrategias y herramientas de apoyo para cada tarea involucrada en la dirección de un equipo es esencial para su éxito [4] y [5].

3 Gestores de proyectos

Existen numerosas posibilidades al momento de elegir una herramienta para gestionar proyectos de software, pero no todas se adaptan a todos los tipos de proyectos, dicha elección es una balanza entre las características del proyecto y las características ofrecidas por la herramienta.

Las organizaciones deben enfrentar algunos inconvenientes al momento de gestionar proyectos grandes y complejos. Algunos de ellos son:

Configuración por defecto: con la ambición de abarcar todo tipo de proyectos, algunas herramientas poseen configuraciones muy genéricas.

Control del proyecto: en la medida que un proyecto aumenta en complejidad y tamaño, es común realizar divisiones de tareas en equipos de trabajo y aplicar diferentes estrategias de gestión y control. En algunas ocasiones existe un alto acoplamiento entre los equipos de trabajo, lo que agrega complejidad en el uso de la herramienta para supervisar el trabajo conjunto y le quita agilidad a la gestión del proyecto.

Equipos y roles: un punto a tener en cuenta es la división y administración del personal en equipos, similar al problema anterior pero enfocado a la autogestión de sus participantes. Disponer de equipos con muchos integrantes dificulta la organización y control de tareas.

Proceso de desarrollo, planificación y verificación: la alta competitividad del mercado hace que las organizaciones intenten lograr la mayor agilidad en la implementación de los cambios con el objetivo de alcanzar sus metas. Más allá de la moda del desarrollo ágil, en algunas ocasiones es necesario realizar un “mix” con las metodologías tradicionales. Esto es un reto al momento de elegir una herramienta que permita mantener la satisfacción de todos los stakeholders.

Búsqueda de historias de usuario: en los proyectos complejos y de gran escala, el número de historias de usuario es alto, por lo tanto, se necesita algún método para agilizar la búsqueda de alguna de ellas. Seguramente existan muchas historias con características en común en su resolución, lo cual, al momento de resolver una historia, se necesite consultar por otras.

La mayoría de los gestores de proyectos son genéricos y admiten configuraciones para cualquier tipo de proyecto. Pretenden abarcar la mayor cantidad de tipos de proyectos posibles. En muchas ocasiones se dejan de lado cuestiones importantes

como el proceso de licitación, la verificación del software, la agilidad en ciertos puntos del desarrollo, etc. [1], [8] y [9],

4 TG Project Management (TG)

TG Project Management, es una nueva propuesta en herramientas de software para la gestión de proyectos complejos que tiene su foco en la eficiencia del proceso de gestión y en la posibilidad de acotar la complejidad de las cuestiones anteriormente descriptas [1] y [9].

4.1 Configuraciones por defecto

TG adopta las mejores prácticas de las metodologías ágiles (Scrum y Kanban), aplicando características de cada metodología según sea conveniente. Con esto mejora la agilidad en el proceso de gestión al eliminar la necesidad de definir una configuración determinada [8] y [10].

TG posee métodos para organizar y administrar el trabajo a través de la definición de roles, equipos y módulos de trabajo.

TG permite la definición de requerimientos e historias de usuarios asociadas a los requerimientos.

TG implementa el modelo de iteraciones (Sprints) de Scrum, permitiendo que cada equipo de trabajo organice y administre el desarrollo de sus historias independientemente de los demás.

TG es una herramienta orientada al usuario desarrollador. Posee características esenciales que aumentan la agilidad durante el proceso de desarrollo.

4.2 Creación de un proyecto

Crear un proyecto en TG es un proceso simple y ágil, simplemente con definir un nombre para identificar al mismo se tiene un nuevo proyecto con las configuraciones por defecto definidas. Figura1.

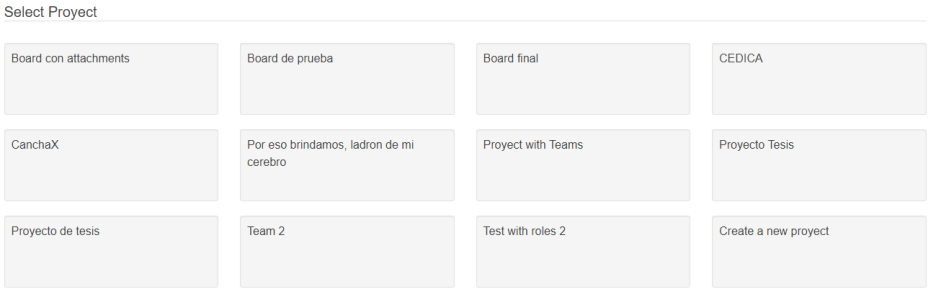


Fig. 1. Listado de proyectos creados.

4.3 Gestión del trabajo y del personal

TG es apto para la gestión de grandes proyectos de software, lo cual brinda la posibilidad de dividir los proyectos en módulos y el personal en equipos. TG permite la creación de equipos en cualquier etapa del proyecto. Una vez definido el equipo, se le pueden asignar miembros y a cada miembro el rol (master, user, owner y other) que cumplirá en el equipo (Figura 2). Hay que tener en cuenta que un mismo usuario no puede pertenecer a dos equipos simultáneamente. Antes de asignar un usuario a un equipo, TG solicita un correo electrónico para realizar la verificación de que el usuario este registrado en Trello y en Github.

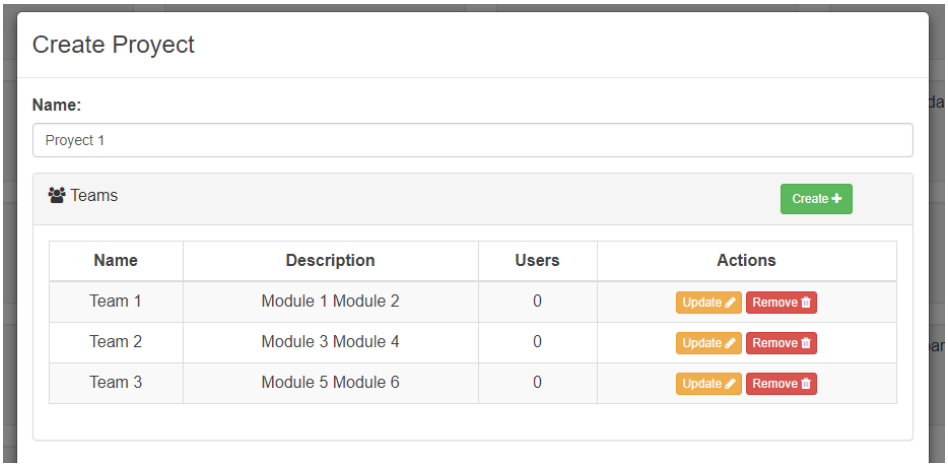


Fig. 2. Creación de equipos para un proyecto en particular.

4.4 Gestión del desarrollo

Una vez creado el proyecto y haber asignado el personal que participará del mismo, TG establece el modelo de proceso de desarrollo con el cual se trabajará. Se optó por el modelo en etapas de la Ingeniería de Software: requisitos, diseño, implementación, verificación y mantenimiento. TG traduce estás etapas en: crear un requerimiento, crear las historias de usuarios asociadas al requerimiento, asignar responsables, asignar historias de usuario a una iteración (Sprint), desarrollar la historia de usuario, vincular la historia de usuario con Github (a través de un pull request), verificar la historia de usuario y validar la solución propuesta.

TG organiza estas etapas en cuatro grandes secciones o vistas, independizando algunos procesos con el objetivo de reducir la complejidad de la herramienta. Figura3.



Fig. 3. Organización de la gestión del desarrollo

Pila de producto (product backlog): desde aquí se crean, modifican y eliminan requerimientos de usuarios. Esta posibilidad solo la tienen aquellos usuarios con roles "master" u "owner". Para cada requerimiento se puede especificar: título, descripción, prioridad, estimación del esfuerzo, razón de existencia, origen y dependencias (Figura 4).

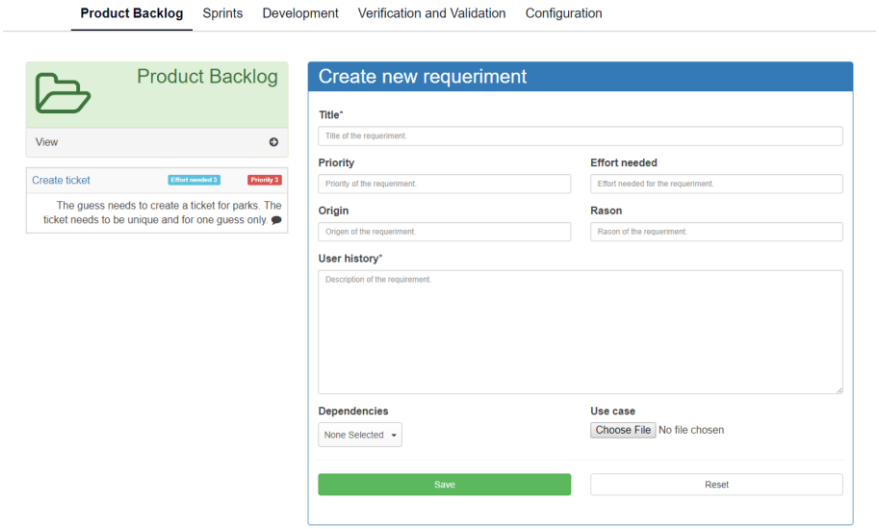


Fig. 4. Pila de producto.

Iteraciones (sprints): desde aquí se gestionan las historias de usuario. Esta posibilidad solo la tienen los usuarios con rol "master". Para cada historia se puede indicar: título, descripción, reglas de negocio, responsable, cantidad de días estimados, reporte de estado, prioridad, relaciones con otras historias, archivos adjuntos e información adicional para la historia. Además, se cuenta con información importante, como las wikis, las cuales brindan información de apoyo al desarrollador en el objetivo de las historias y la pila de producto, pero desde una perspectiva del "master". Figura 5.

Además, en esta sección TG lista los requerimientos de la pila de producto, de esta manera se consigue agilizar el proceso de identificación de un requerimiento en particular. Las historias de usuario son un elemento importante de comunicación en las metodologías ágiles, son el medio para brindar la información que el desarrollador necesita para incorporar o modificar una funcionalidad del producto de software. En TG todas las historias de usuario son creadas a partir de un requerimiento de la pila de producto.

En TG todas las historias de usuario pasan por los siguientes estados: "Not started", "Ready for dev", "Dev completed", "Ready for test" y "Closed". Estos estados son utilizados para la gestión del desarrollo y marcan la etapa en la que se encuentra una historia.

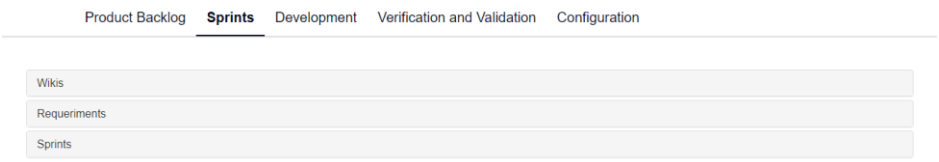


Fig. 5. Gestión de las iteraciones (Sprints)

Desarrollo (Development): desde aquí se realiza la gestión del proceso de desarrollo en la etapa de implementación. Esta solapa está destinada a que cada integrante del equipo sea el responsable de administrar y gestionar el desarrollo de las historias asignadas. Esta vista es accedida solamente desde el rol "member".

Verificación y Validación (Verification and Validation): desde aquí se puede decir que la historia se encuentra aceptada desde un punto de vista de la calidad de la solución propuesta, pero aún debe pasar por un ambiente de prueba. Desde esta vista TG le permite al "master" seleccionar aquellas historias de la iteración actual y aceptarlas o rechazarlas en base a la validación que han recibido las mismas. Validar una historia, consiste en tratar de descubrir defectos funcionales, comprobando que la solución propuesta cumple con las expectativas del cliente. Si una historia es rechazada, significa, que no cumple con las expectativas del cliente o que no se encuentra acorde a alguna regla definida, es aquí, que el "master" debe regresar la historia para desarrollar una nueva solución y reparar los errores encontrados. Una vez que la historia ha sido aceptada, TG asume que se ha generado un incremento dentro del sistema, esto es que el requerimiento al cual se encuentra asociado la historia de usuario avanza y además la iteración actual posee una historia más resuelta. Cuando se llega a este punto se considera finalizado el proceso de gestión del desarrollo de una historia. TG almacena todas las historias finalizadas para futuras búsquedas.

4.5 Búsqueda de una historia de usuario

La búsqueda de una historia de usuario es importante debido a que en algunas ocasiones suelen compartir ciertas características (reglas de negocio, implementación, problemas, etc.) y en su desarrollo puede resultar muy útil referenciar a alguna historia que ya se haya realizado. TG posee un potente buscador de historias de usuario. La principal característica es que se apoya en la relación entre historias cuando las mismas son creadas. Si esta opción se utiliza adecuadamente, simplemente con navegar a través de las historias se obtendría rápidamente lo buscado. No obstante, no es la única forma de buscar historias, TG almacena las iteraciones (Sprints) finalizadas, esto permite acotar la búsqueda, visualizando únicamente una iteración en particular. Otra posibilidad, es buscar a partir de un requerimiento, en TG todas las historias nacen a partir de un requerimiento, TG mantiene esta relación permitiendo acotar la búsqueda a las historias derivadas de dicho requerimiento.

Contar con más de un proceso para realizar búsquedas de historias, añaden a TG una característica de agilidad para el desarrollador al momento de necesitar encontrar una historia de usuario en particular.

5 Conclusiones

TG Proyect Management se suma como una opción competitiva, al momento de seleccionar una herramienta para la gestión de proyectos de software amplios y complejos. TG se presenta como una herramienta de apoyo, que prioriza la agilidad y flexibilidad durante el desarrollo, en lugar de definir procesos largos y estructurados. TG permite gestionar y administrar todo lo relacionado al proyecto de software, controlar el desarrollo, gestionar requerimientos, historias de usuario, equipos, actividades, etc.

Aunque los proyectos de software actuales aumentan su complejidad a medida que avanzan en su ciclo de vida, TG se encuentra diseñado para adaptarse fácilmente al impacto del incremento de la complejidad.

La principal característica y distinción de TG (con respecto a Trello y Jira) es permitir la gestión del proyecto desde su inicio hasta su finalización, sin tener que preocuparse por el momento del ciclo de vida en el que se encuentra, la cantidad de integrantes o la magnitud del mismo. TG se adapta al proyecto, no es necesario adaptar el proyecto a TG.

6 Trabajos Futuros

Como trabajo futuro, se propone ampliar las características de TG para beneficiar la agilidad y el avance del proyecto. Alguna de estas características son:

- Incorporar gráficos estadísticos para el progreso del proyecto e iteraciones.
- Definir estrategias para identificar, analizar y definir riesgos para el proyecto. Mejorando así los tiempos estimados y la eficiencia de la gestión.
- Incorporar análisis de costos.
- Incorporar estrategias para que el usuario desarrollador participe en diversas tareas.
- Vincular TG con una herramienta de correo como por ejemplo Gmail.

7 Referencias

1. Pablo Lledó y Gustavo Rivarola. (2007). *Gestión de proyectos / Pablo Lledó y Gustavo.* Buenos Aires: Prentice Hall - Pearson Education.
2. Schaull, S. F. (2011). EL “DESARROLLO DE SOFTWARE” COMO “INGENIERÍA DE SOFTWARE”. *Ing. USBMed*, 4.

3. Alexander Menzinsky, Gertrudis López, Juan Palacio. (2016). *Scrum Manager.* Iubaris Info 4 Media SL.
4. Schwaber, K. (1995). Scrum. *SCRUM Development Process* , 23.
5. Lledó, P. &. (2007). *Gestión de Proyectos - Cómo dirigir proyectos exitosos, coordinar los recursos humanos y administrar los riesgos.* California: Pearson. Prentice Hall.
6. Almunia, P. (22 de Febrero de 2016). *itmplatform*. Recuperado el 25 de Junio de 2017, de itmplatform: http://www.itmplatform.com
7. Pfleeger, S. L. (2002). *Ingeniería de software teoría y práctica.* Buenos Aires: Pearson Educación.
8. Figuerola, N. (2015). Cómo seleccionar una Metodología de Project Management. *Project Management*, 7
9. Pressman, P. R. (2003). *Ingeniería del software - Un enfoque práctico.* Mexico: Mc Graw Hill.
10. Highsmith, M. F. (2001). The Agile Manifesto. *The Agile Manifesto*, 7.

Análisis de consumo de energía en aplicaciones 3D para dispositivos móviles

Federico Cristina, Sebastián Dapoto, Pablo Thomas, Patricia Pesado

Instituto de Investigación en Informática LIDI,
Universidad Nacional de La Plata – Argentina
Centro Asociado Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

{fcristina, sdapoto, pthomas, ppesado}@lidi.info.unlp.edu.ar

Resumen. Actualmente, un requerimiento no funcional crítico en aplicaciones sobre dispositivos móviles es el consumo de energía. Si bien otras restricciones, como el poder de procesamiento y el tamaño de la memoria, han evolucionado notablemente a lo largo de los años, el aumento en la autonomía de las baterías no ha sido realmente pronunciado. Bajo este escenario, las aplicaciones en general, y especialmente las que incluyen entornos tridimensionales (3D) son cada vez más complejas, requieren un mayor poder de cómputo y por consiguiente generan un mayor consumo de energía. Por este motivo es importante tener en cuenta durante el desarrollo de una aplicación 3D, el eventual consumo de energía que ésta pueda implicar al ser utilizada por un lapso de tiempo determinado. El presente trabajo propone un conjunto de pruebas para evaluar la incidencia de las principales características de las aplicaciones 3D sobre el consumo de energía en dispositivos móviles.

Palabras Clave: Unity engine, dispositivos móviles, aplicaciones 3D, consumo de energía

1 Introducción

En la actualidad los dispositivos móviles forman parte de la vida cotidiana, tanto en el ámbito personal como en el profesional. En los últimos años, estos dispositivos se han vuelto mucho más sofisticados, permitiendo ejecutar aplicaciones complejas y con exigentes requerimientos de hardware.

Sin embargo, en lo que respecta a la autonomía de las baterías de los dispositivos móviles, los fabricantes han logrado paulatinamente aumentar la capacidad de dichas baterías, pero sin resultados notables.

La mayoría de los teléfonos inteligentes actuales tiene una autonomía de batería de a lo sumo dos días. Esto se debe a que todos los componentes del dispositivo móvil dependen del suministro de energía de la batería, especialmente el procesador y la pantalla, creando un compromiso entre el poder de procesamiento, la visualización y

la duración de la batería. Como consecuencia de ésto, si una aplicación realiza un procesamiento intensivo, limitará el tiempo que el usuario puede usar el dispositivo.

Este mayor consumo plantea problemas para la evolución de la computación móvil, ya que los desarrolladores no pueden utilizar todo el potencial de la tecnología actual sin sacrificar el tiempo de vida útil de la batería.

El principal problema está dado porque la vida útil de las baterías de litio está determinada por la cantidad de ciclos de carga y descarga. Es decir, mientras más intensivamente se utilice el dispositivo, más rápido se deteriorará la batería. Por consiguiente, un usuario podría decidir desinstalar una aplicación cuyo consumo de energía sea considerado como excesivo.

Por todo esto, es imperioso desarrollar aplicaciones móviles considerando el consumo de energía que puedan producir. Esto toma mayor relevancia en las aplicaciones móviles 3D, dado que éstas son implícitamente demandantes de poder de cómputo y por consiguiente de consumo de energía.

A fin de lograr este objetivo, el presente trabajo propone una evaluación que permite aislar, analizar y dimensionar la incidencia de las principales características de las aplicaciones 3D sobre el consumo de energía en dispositivos móviles.

Esta evaluación brinda un soporte al ingeniero de software de aplicaciones móviles 3D, posibilitando la identificación de los factores que impactan mayormente en el consumo de energía, permitiendo ajustarlos hasta lograr minimizar el consumo.

El resto del trabajo se organiza del siguiente modo: la sección 2 describe la motivación; la sección 3 presenta en detalle la metodología de evaluación; la sección 4 exhibe la experimentación realizada y los resultados obtenidos; finalmente en la sección 5 se exponen las conclusiones y trabajos futuros.

2 Motivación

El análisis que se propone en este trabajo surge de la metodología de estudio propuesta y resultados obtenidos en [1], en el cual se evaluó la performance del framework Unity [2] para desarrollo de aplicaciones móviles 3D, considerando un conjunto de características inherentes a aplicaciones 3D y evaluando cada una de ellas de forma independiente. Así, fue posible individualizar los principales factores de incidencia en la degradación de la fluidez y visualización de las aplicaciones móviles 3D.

De la misma forma, es posible extender el análisis realizado en [1] hacia el campo del consumo de energía. Es decir, del mismo modo que se buscó determinar cuáles son las características que más afectan a la performance de las aplicaciones móviles 3D, se propone ahora evaluar si existe una correlación de estas características con el consumo de energía; analizándolas de manera aislada.

Por consiguiente, el conjunto de pruebas es el mismo que el realizado en [1], pero centralizado en el consumo de energía. Las pruebas a las que se hace referencia son:

1. Renderización de objetos simples (Basic Mesh Rendering)
2. Renderización de objetos complejos (Complex Mesh Rendering)
3. Luces y sombras (Lights & Shadows)

4. Texturas (Textures)
5. Sistemas de partículas (Particle Systems)
6. Física (Physics)

Para realizar la evaluación de consumo de energía en aplicaciones móviles 3D es necesario disponer las herramientas de software/hardware a fin de llevar a cabo la serie de pruebas. Estas herramientas son:

- Una aplicación 3D ad-hoc que abarque la ejecución independiente de cada una de las pruebas previamente detalladas.
- Una herramienta de profiling, la cual se encarga de medir el consumo de energía del dispositivo en tiempo real.
- Dispositivos móviles sobre los cuales realizar las pruebas.

Si bien existen trabajos previos relacionados con el consumo de energía en dispositivos móviles [3][4], consideramos novedoso el análisis del consumo de energía en aplicaciones 3D para dispositivos móviles de manera independiente para cada una de las características especificadas.

3 Metodología de evaluación

El principal objetivo del análisis propuesto es determinar cuáles son las características de una aplicación 3D que generan un mayor consumo de energía en un dispositivo móvil.

Las características que tienen relación directa con procesamiento intensivo tendrían también relación directa con el alto consumo de energía. A fin de poder evaluar aisladamente cada una de esas características, la metodología de evaluación adopta entonces el conjunto de pruebas definido en [1].

La aplicación 3D ad-hoc implementada es una ampliación de la aplicación desarrollada sobre Unity en [1], especialmente modificada para los casos de prueba a realizar.

Los cambios realizados en la aplicación incluyen una configuración especial para cada una de las pruebas definidas. Para el caso particular de este trabajo de análisis se utilizan cubos como objetos para la prueba de objetos simples, mientras que para la prueba de renderización de un objeto complejo se toma parte del modelo del edificio de la Facultad de Informática (la planta baja) que es utilizado en la aplicación InfoUNLP3D [5][6]. Dicho objeto contiene aproximadamente 200.000 polígonos. Para las pruebas de texturas se aplica transparencia tipo vidrio a los elementos. En cuanto al sistema de partículas, se genera uno similar al utilizado por el robot en R-Info3D [7] al momento de ejecutar una instrucción de reposicionamiento (Pos).

Además, la aplicación contiene dos parámetros numéricos que permiten configurar la prueba a realizar. El primero permite especificar la cantidad de objetos a utilizar en la prueba, o, en el caso de la prueba 2, la distancia de acercamiento al objeto complejo. El segundo permite configurar la cantidad de cuadros por segundo (frames per second - FPS) a utilizar en la prueba. Un valor igual a -1 deja este parámetro libre.

En este último caso, los FPS dependerán del poder de cómputo del dispositivo utilizado.

La herramienta de profiling seleccionada para realizar las mediciones de consumo de energía se denomina Trepn Profiler [8][9]. Esta herramienta permite medir en tiempo real el consumo de energía de una aplicación en particular. Trepn Profiler fue desarrollada por la empresa Qualcomm Technologies Inc., y asegura una medición precisa de la energía consumida en dispositivos que cuenten con procesadores desarrollados por Qualcomm. Los dispositivos para ejecutar las pruebas fueron seleccionados teniendo en cuenta esta restricción. Además, para contar con un espectro abarcativo de hardware, fueron elegidos equipos con diferente capacidad de procesamiento.

La metodología de evaluación implica relevar durante un período de tiempo prefijado el consumo de energía bajo la ejecución de cada una de las distintas pruebas. A diferencia de la metodología utilizada en [1], el número de objetos en pantalla es constante durante el tiempo de ejecución de cada prueba. Una vez transcurrido el tiempo prefijado, se toma la mediana del consumo relevado a lo largo de la prueba.

A fin de homogeneizar las pruebas y minimizar errores en las mediciones, todas las ejecuciones a realizar sobre todos los dispositivos deben respetar las siguientes definiciones:

- Al comenzar con el conjunto de pruebas, los dispositivos a utilizar deben tener la carga de la batería mayor al 95%.
- Las ejecuciones deben realizarse con el brillo de la pantalla del dispositivo al mínimo.
- Los dispositivos deben estar configurados en “Modo avión”, para evitar cualquier tipo de recepción de señal que pudiera incidir en las mediciones.
- No deben existir aplicaciones ejecutándose en segundo plano, dado que pueden consumir tiempo de procesamiento y de esa forma afectar las mediciones.
- Se debe prefijar el mismo número de cuadros por segundo para todas las pruebas en todos los dispositivos, dado que el FPS utilizado tiene una incidencia directa sobre el consumo, tal como se verifica posteriormente en la sección 4.
- La aplicación Trepn debe estar configurada sólo para relevar y guardar los datos de consumo, evitando mostrar datos en pantalla. La visualización de información en tiempo real puede influir en la medición del consumo de la prueba que está siendo ejecutada.

4 Pruebas realizadas y resultados obtenidos

El conjunto de pruebas se efectuó sobre tres dispositivos móviles con poder de cómputo diferente. Los teléfonos inteligentes utilizados fueron los listados en la tabla 1.

Todas las pruebas se realizaron limitando el número de cuadros por segundo a diez (10 FPS). A su vez, el número de objetos fijado para cada prueba se estableció con el

fin de garantizar que todos los dispositivos usados, incluso los de menor poder de procesamiento, puedan ejecutar fluidamente el conjunto de pruebas.

Tabla 1. Dispositivos utilizados en las pruebas.

Id	**Marca y modelo del dispositivo**	**Sistema Operativo**	**Características generales**
D1	Samsung Galaxy J7 Prime	Android 6.0.1	Smartphone. Octa-core, 1.6 GHz. 3 GB RAM.
D2	Samsung Galaxy A5 (2017)	Android 6.0.1	Smartphone. Octa-core, 1.9 GHz. 3 GB RAM.
D3	Samsung Galaxy S8	Android 7.0	Smartphone. Octa-core, 4 x 2.3 GHz, 4 x 1.7 GHz. 4 GB RAM.

Además, las pruebas se realizaron 30 veces sobre cada uno de los dispositivos móviles, a fin de validar la consistencia de la medición. Luego se obtuvo la mediana de consumo de cada prueba sobre cada dispositivo. A modo de ejemplo, en la figura 1 se puede observar la aplicación ejecutando la prueba número 2, en la que se visualiza un conjunto de objetos simples (cubos) girando sobre su eje y permitiendo la proyección de luces y sombras.

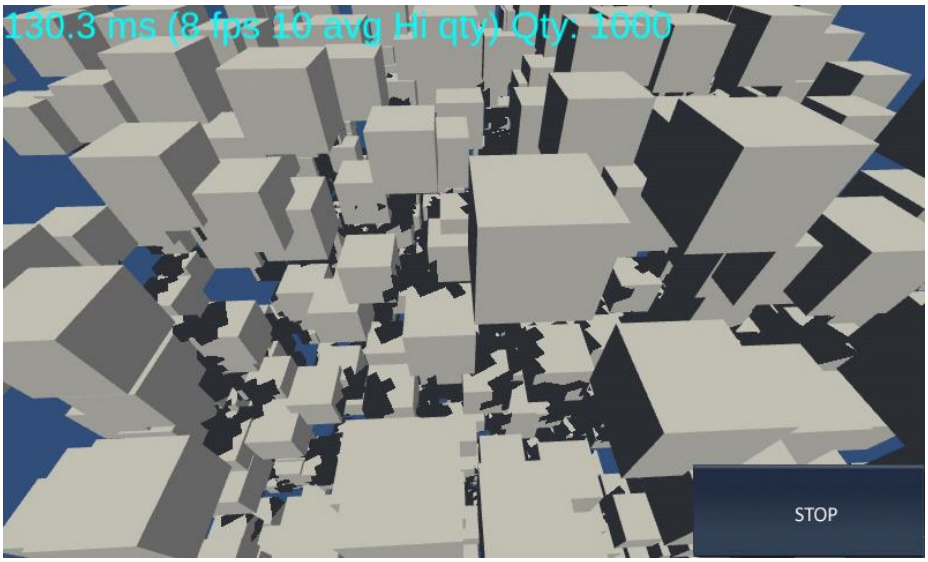


Fig. 1. Prueba número 2: luces y sombras.

Con la finalidad de obtener resultados normalizados, es decir, independientes del dispositivo móvil en el que se ejecutan las pruebas, se adoptó un enfoque de consumo

de energía referencial a la prueba más simple. Es por esto que el análisis se basa en tomar como referencia de consumo a la prueba número 1. Esta prueba se establece con un valor base de consumo igual a 1, y las restantes pruebas se evalúan como el delta (Δ) de consumo con respecto a dicha prueba.

Las figuras 2, 3 y 4 presentan para cada dispositivo en cada una de las seis pruebas realizadas: la mediana del consumo de energía en mWh, el delta y la gráfica de delta. Tal como se explicó previamente, para las pruebas 2 a 6 se calcula el delta de consumo de energía basado en el consumo obtenido en la prueba 1.

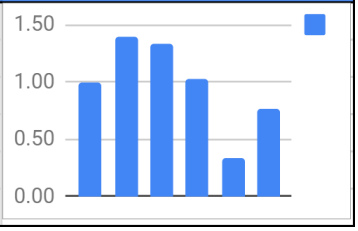
	D1 (Samsung J7)	mWh	Δ	Gráfica Δ
OBJ. SIMPLE	Prueba 1. 1200 Objetos	**217**	1.00	
OBJ. COMPLEJO	Prueba 2. 200 Distancia	**303**	1.40	
LUCES/SOMBRAS	Prueba 3. 1000 Objetos	**290**	1.34	
TEXTURAS	Prueba 4. 1200 Objetos	**223**	1.03	
SIST.PARTICULAS	Prueba 5. 8 Objetos	**74**	0.34	
FISICA	Prueba 6. 180 Objetos	**166**	0.76	

Fig. 2. Dispositivo D1. Mediana muestral de las pruebas ejecutadas, fijadas a 10 FPS.

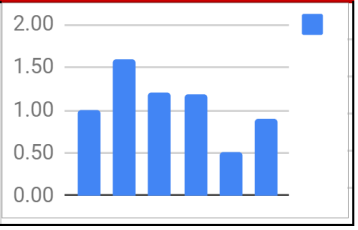
	D2 (Samsung A5)	mWh	Δ	Gráfica Δ
OBJ. SIMPLE	Prueba 1. 1200 Objetos	**342**	1.00	
OBJ. COMPLEJO	Prueba 2. 200 Distancia	**543**	1.59	
LUCES/SOMBRAS	Prueba 3. 1000 Objetos	**414**	1.21	
TEXTURAS	Prueba 4. 1200 Objetos	**403**	1.18	
SIST.PARTICULAS	Prueba 5. 8 Objetos	**176**	0.51	
FISICA	Prueba 6. 180 Objetos	**309**	0.90	

Fig. 3. Dispositivo D2. Mediana muestral de las pruebas ejecutadas, fijadas a 10 FPS.

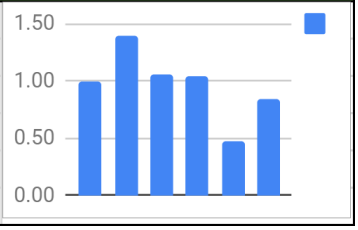
	D3 (Samsung S8)	mWh	Δ	Gráfica Δ
OBJ. SIMPLE	Prueba 1. 1200 Objetos	**421**	1.00	
OBJ. COMPLEJO	Prueba 2. 200 Distancia	**585**	1.39	
LUCES/SOMBRAS	Prueba 3. 1000 Objetos	**448**	1.06	
TEXTURAS	Prueba 4. 1200 Objetos	**438**	1.04	
SIST.PARTICULAS	Prueba 5. 8 Objetos	**201**	0.48	
FISICA	Prueba 6. 180 Objetos	**356**	0.85	

Fig. 4. Dispositivo D3. Mediana muestral de las pruebas ejecutadas, fijadas a 10 FPS.

La figura 5 permite contrastar visualmente el consumo de cada una de las características evaluadas, en todos los casos en función de la prueba 1. Aquí es posible notar la similitud en el consumo para cada una de las características evaluadas a lo largo de todos los equipos utilizados. Como puede observarse, el mayor consumo se da en la prueba 2, la cual implica renderizar un volumen considerable de polígonos en pantalla. Como contrapartida, la prueba 5 es la que genera menor consumo de energía, debido al tipo de cálculos requeridos para la renderización de sistemas de partículas.

A fin de confirmar la incidencia del FPS en el consumo, se realizó un análisis adicional sobre la prueba 1. Este análisis consistió en mantener fijo el número de objetos en escena, pero incrementar el número de cuadros por segundo desde 10 hasta 60, en intervalos de 10, relevando el consumo en cada paso.

La figura 6 muestra los resultados obtenidos, los cuales demuestran que efectivamente a mayor cantidad de FPS existe un mayor consumo. Adicionalmente, esto permite estimar los valores de consumo de energía de las características evaluadas para un número mayor de FPS.

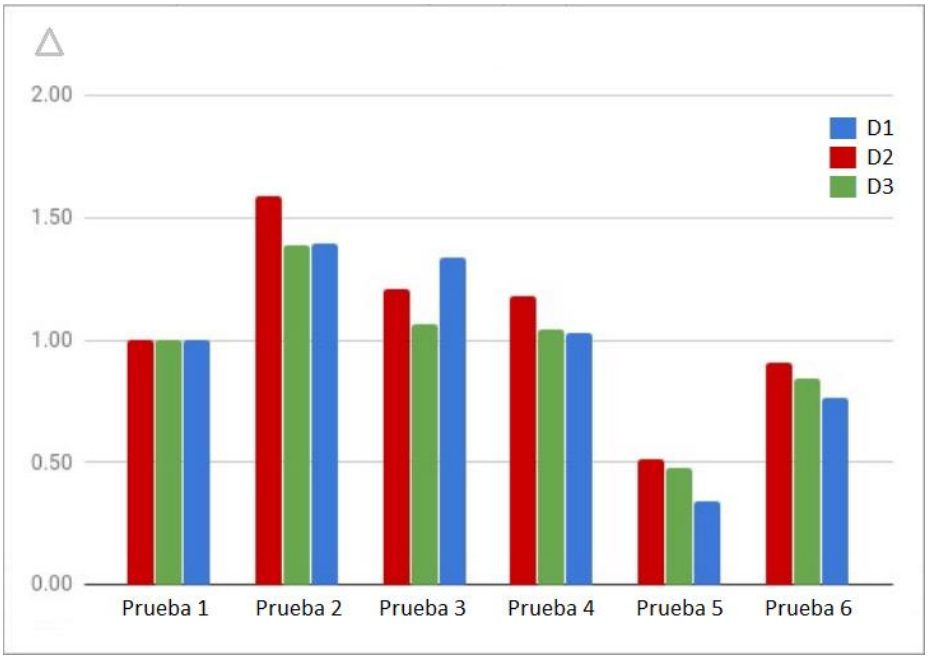


Fig. 5. FPS. Comparación de las pruebas realizadas sobre los dispositivos. El patrón de consumo de las características evaluadas es similar para los equipos utilizados.

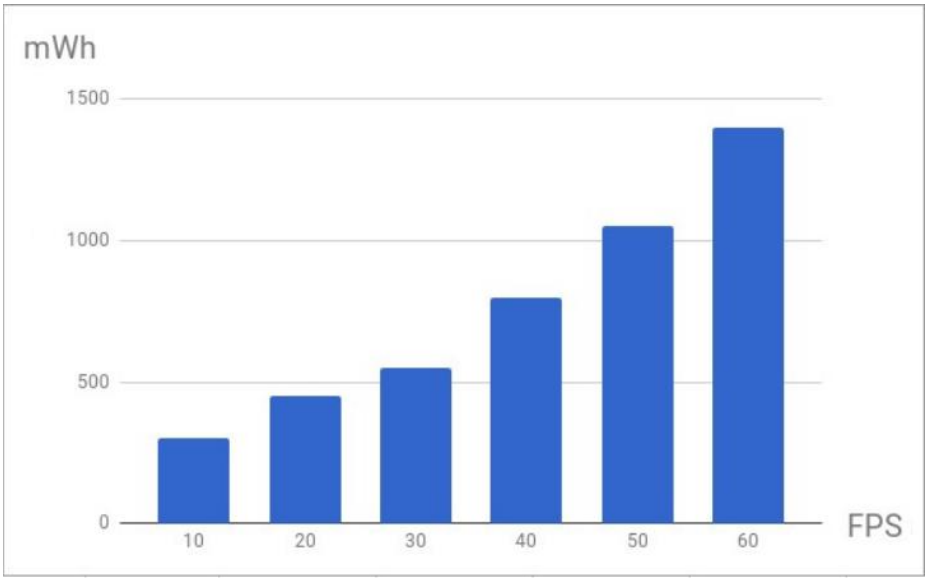


Fig. 6. Prueba 1. Consumo según FPS.

Cabe destacar, que si bien la incidencia de las características analizadas sobre el consumo varía de acuerdo al software y al hardware sobre el cual se ejecuta la aplicación, las pruebas realizadas demuestran que existe un patrón común respecto al consumo de energía en relación al tipo de requerimiento visual.

5 Conclusiones y trabajo a futuro

El presente trabajo propone un conjunto de actividades para evaluar el consumo de energía de aplicaciones 3D sobre dispositivos móviles. La evaluación propuesta permite determinar la incidencia de cada una de las características principales de las aplicaciones 3D sobre el consumo de energía.

Se desarrolló un prototipo que permite ejecutar una serie de pruebas independientes para cada una de las características típicas de una aplicación tridimensional. Luego, midiendo el consumo de cada prueba por separado, se logró determinar cuales son las características que consumen más energía para la experimentación realizada.

Por lo tanto, un ingeniero de software puede desarrollar una aplicación móvil 3D considerando el posible consumo de energía de su aplicación en función de la complejidad de la escena en lo referido a: cantidad de polígonos, texturas, sombras, entre otras características visuales.

A futuro se prevé realizar la misma experimentación sobre otros engines de desarrollo de aplicaciones 3D, como por ejemplo Unreal Engine [10] o bien CryEngine [11], y así poder comparar consumos de energía entre los distintos frameworks, para cada una de las características evaluadas.

Referencias

1. Cristina, F.; Dapoto, S.; Thomas, P.; Pesado, P. “Performance evaluation of a 3D engine for mobile devices”. Computer Science – CACIC 2017. Communications in Computer and Information Science, vol 790. A. De Giusti, Springer International Publishing. 2018. ISBN: 978-3-319-75213-6, 978-3-319-75214-3, pages 155-163.
2. Unity. https://unity3d.com.
3. Riaz, M. "Energy consumption in hand-held mobile communication devices: A comparative study". International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). 2018. ISBN: 978-1-5386-1370-2/18.
4. Corbalan L., Fernández J., Cuitiño A., Delia L., Cáseres G., Thomas, P.; Pesado, P. "Development Frameworks for Mobile Devices: A Comparative Study about Energy Consumption". Association for Computing Machinery (ACM). 2018. ISBN 978-1-4503-5712-8/18/05.
5. Cristina, F.; Dapoto, S.; Thomas, P.; Pesado, P. “InfoUNLP3D: An interactive experience for freshman students”. Computer Science & Technology Series - XXII Argentine Congress of Computer Science. Selected Papers CACIC 2016. EDULP. 2017. ISBN: 978-987-4127-28-0, pages 249-256.
6. InfoUNLP3D. Google Play Store Download Link. https://play.google.com/store/apps/details?id=ar.edu.unlp.info.infounlp3d.
7. Cristina, F.; Dapoto, S.; Thomas, P.; Pesado, P. “3D Mobile Prototype for Basic Algorithms Learning”. Computer Science & Technology Series - XXI Argentine Congress Of Computer Science. Selected Papers CACIC 2015. EDULP. 2016. ISBN: 978-987-4127-00-6, pages 239-247.
8. Trepn Power Profiler. https://developer.qualcomm.com/software/trepn-power-profiler.
9. Trepn Power Profiler. Google Play Store Download Link. https://play.google.com/store/apps/details?id=com.quicinc.trepn.
10. Unreal Engine. https://www.unrealengine.com/.
11. CryEngine. https://www.cryengine.com/.

Mobile Application Development Approaches: A Comparative Analysis on the Use of Storage Space

Juan Fernández Sosa[1], Pablo Thomas[1], Lisandro Delía[1], Germán Cáseres[1], Leonardo Corbalán[1], Fernando Tesone[1], Alfonso Cuitiño[1], Patricia Pesado[1]

[1] Computer Science Research Institute LIDI (III-LIDI)*
School of Computer Science, National University of La Plata,
La Plata, Buenos Aires, Argentina
*Partner Center of the Scientific Research Agency of the Province of Buenos Aires (CICPBA)
{ jfernandez, pthomas, ldelia, gcaseres, corbalan, ftesone,
acuitino,ppesado}@lidi.info.unlp.edu.ar

Summary. The purpose of software development is meeting both functional and non-functional requirements. In mobile device applications, non-functional requirements are more relevant due to the restrictions inherent to these devices. One of these restrictions is the availability of limited storage space. Therefore, the size of a mobile application affects user preference for use. In this article, we assess how the choice of a mobile application development approach affects the final size of the application; we focus our analysis on text-, audio- and video-based applications.

Keywords: Mobile devices, multi-platform mobile applications, native mobile applications, application size.

1 Introduction

Application development for mobile devices poses a number of challenges specific to this activity that were not present in traditional software development [1]. The diversity of platforms, programming languages, and development tools, as well as device heterogeneity as regards computation power, storage and battery life, are just some of the aspects that Software Engineers have to consider.

In many cases, the success of an application for mobile devices depends on its popularity. To maximize its presence in the market, it should be possible to run it on different platforms [2]. Currently, the universe of operating systems for mobile devices is led by the Android (74.24%) and iOS (20.83%) platforms [3].

In recent years, various methodologies for developing mobile applications have been proposed and studied – the native approach and several multi-platform development approaches (web, hybrid, interpreted and cross-compilation). The former requires the development of specific applications for each platform, with parallel development projects, using specific programming languages and tools for each platform. On the other hand, multi-platform approaches allow generating applications that can be run in more than one platform and produced within a single development project.

In addition to the choice of development approach, there are other difficulties to be considered in relation to the limitations posed by mobile devices. Battery life, computation power and storage space are non-functional requirements that significantly affect the decision of the end users to install or keep a mobile application in their smartphones.

In [4], the authors of this paper have analyzed the advantages and disadvantages of the multi-platform development approaches mentioned above, from the point of view of the Software Engineer. Similarly, the authors of [2] and [5] have carried out, respectively, a comparative analysis of performance and battery consumption for native developments and various multi-platform approaches.

Available storage space varies significantly among mobile devices. The operating system and pre-installed applications (also known as bloatware) take up a large portion of this space. This limits the possibilities for use [6] and, as a result, users are usually reluctant to install new apps, or they even stop using those that require a lot of space.

The size of the apps for mobile devices has increased with time. In Android, it quintupled between 2012 and 2017 [7]. In iOS, the 10 most-installed applications in the US increased their size 12 times (1,100%) between 2012 and 2017 [8]. This increase is largely due to market maturity – as time goes by, users require new functionalities for their apps.

For all these reasons, it is essential that developers consider a limited use of storage space to attract more users.

In this paper, a comparative analysis of the storage space used by mobile applications developed using the native approach and the different multi-platform approaches described in [9] is presented.

In Section 2, the different approaches for developing mobile apps are discussed and described; Section 3 details the experiments carried out to compare the storage space used by applications developed with these approaches. In Section 4, the results obtained are presented and discussed. Finally, the conclusions and future lines of work are presented.

2. Types of Applications for Mobile Devices

In recent years, the mobile device market, especially that of smartphones, has seen a remarkable growth [10]. As regards operating systems, Android and iOS are the strongest in the market. Each has its own development infrastructure. The main challenge for mobile device application developers is being able to offer solutions for all market platforms; in these cases, development costs are so high that sometimes are hard to afford [11].

An appropriate solution to this problem is creating and maintaining a single application that is compatible with all platforms. The goal of multi-platform development is maintaining a single source code for several platforms. This results in a significant reduction of effort and costs.

In the following sections, different approaches are presented for the development of applications for mobile devices:

2.1 Native Applications

Native applications are developed to be run on a specific platform, considering the type of device and the operating system and its version. The source code is compiled to obtain executable code, similar to the process used for traditional desktop applications. When the application is ready for distribution, it is published in the app store specific for each platform. These stores have an audit process to check if the application meets the requirements of the platform on which it is going to be run. Finally, the application becomes available for download by end users.

One of the characteristics of a native application is that it allows unlimited interaction with all the functions and features offered by the device (GPS, camera, accelerometer, calendar, etc). Additionally, Internet access may not be a requirement to run this type of applications. They are fast and can be run in the background, and they issue an alert when there is an event that requires user intervention.

This development approach has a high cost, since each platform requires the use of a specific programming language. Therefore, if the goal of a project is to encompass several platforms, a different application must be generated for each of them. This means that the coding, testing, maintenance and going live processes must be carried out more than once.

2.2 Web Applications

Web applications for mobiles are designed to be executed in the web browser of the device. They are developed using standard technologies such as HTML, CSS and JavaScript.

One of the advantages of this approach is that no specific component needs to be installed in the device, and no third-party approval is required before publication and distribution. Only Internet access is required. Additionally, updates are pushed directly to the device, since changes are applied on a server and enabled for immediate access by the users. In summary, they are easy and quick to implement. However, the greatest advantage of mobile web applications is that they are fully platform-independent. There is no need to adapt to a specific operating system, only a web browser is needed.

On the other hand, this approach can reduce execution speed, which can result in a somewhat less satisfactory user experience, and interfaces are more limited than those offered by native applications. Performance can also be affected due to connectivity issues, among others. Finally, some limitations could also be observed in relation to access to specific features offered by the device [12].

2.3 Hybrid Applications

Hybrid applications use web technologies (HTML, JavaScript and CSS), but are not run by a browser. Instead, they are run on a web container of the device that provides access to device-specific features through an API.

Hybrid applications offer great advantages, such as code reuse for the different platforms, access to device hardware, and distribution through application stores [13].

Hybrid applications have two disadvantages when compared to native applications. The first of these is that user experience suffers from not using the native

components in the interfaces. The second disadvantage is that these apps may be slower due to the additional load associated to the web container where they are run.

One of the most popular frameworks used in this approach is Apache Cordova [14].

2.4 Interpreted Applications

Interpreted applications are built from a single project that is mostly translated to native code, with the rest being interpreted at runtime. Their implementation is non-platform dependent and uses several technologies and languages, such as Java, Ruby, XML, and so forth.

Unlike the web and hybrid multi-platform development approaches, with the interpreted applications approach native interfaces are obtained, which is one of the main advantages of this type of applications.

Some of the most popular interpreted development environments for these applications are Appcelerator Titanium [15] and NativeScript [16].

2.5 Applications Generated by Cross-Compilation

These applications are compiled natively by creating a specific version for each target platform. Some examples of development environments used to generate applications by cross-compilation are Xamarin [17] and Corona [18].

Xamarin allows compiling fully native applications for iOS and Android by sharing the same base code written in C#. Integrated with Microsoft Visual Studio, it also allows generating applications for Windows Phone.

Xamarin allows sharing the entire business logics code, but user interfaces must be programmed separately for each target platform.

Corona is a multi-platform framework that allows developers build both general-purpose applications and games for the main platforms. A single base code is used, which is then published for the different platforms. Unlike Xamarin, no specialized rewriting or projects are required. Programming is done with Lua, which is a simple scripting language.

3. Experiment

The different approaches for mobile development, native and multi-platform, use different techniques to build the applications and the code, which can affect the final size of the application.

The experiments presented here are aimed at assessing the impact that choosing one development approach has on the final size of the applications.

3.1 Designing the Experiment

To carry out the experiment, development frameworks were selected based on the different approaches discussed in previous sections. Tests were designed to assess application size using the latest stable versions of these frameworks at the time of the experiments.

The frameworks used, and their versions, are:

1. Android SDK API 25 (Java, native)
2. Apache Cordova, version 7 (multi-platform, hybrid)
3. Appcelerator Titanium, version 5 (multi-platform, interpreted)
4. NativeScript, version 3 (multi-platform, interpreted)
5. Xamarin, version 6 (multi-platform, cross-compilation)
6. Corona, version 2016 (multi-platform, cross-compilation)

When assessing the use of storage space, the various functionalities offered should be analyzed since, depending on the framework development strategy, there is a chance that more libraries, modules or plug-ins are added to the generated APK. An APK, or file with .apk extension, contains an application for the Android operating system. Its name is an abbreviation of the term **A**ndroid **P**ac**K**age.

To carry out the task, three applications with different functionalities were built.

The first experiment consisted in showing the text "Hello, World!" on the screen until the user closes the application. This trivial functionality allows calculating how many bytes are added by each framework.

Experiments #2 and #3 were designed to play multimedia files (audio and video, respectively), since there is a chance that other tools or libraries are added to that end, increasing the final size of the application.

The app that plays audio uses a 1.32 MB audio file, while the app that plays video uses an 89.2 MB video file. In both cases, the file is played automatically when the app is started.

Thus, there are 18 test cases: 3 applications (text, audio and video) for each of the 6 frameworks.

In all cases, when building the application special care was taken to check that no additional files were added, since the frameworks could potentially add them.

All tests were carried out for the Android operating system compiled for version 7.1, since Android currently represents the lion's share of the global market [3].

The experiments are independent from the mobile device used, since the results considered the size of the APK generated by each framework.

The source code for all developments is available for public access [19].

4 Results obtained

For all tests, the applications were generated following the standard procedure recommended by the documentation for each framework.

Below, the results obtained for each experiment are discussed.

4.1.1 Use of Storage Space in Experiment #1

The text-based application developed using the native approach turned out to be the smallest one. Cordova, hybrid approach, produced an APK only 18% larger than that of the native approach.

The application generated by cross-compilation with Xamarin had a size of 4.08 MB (176% larger than the native approach), and the one generated with Corona was 6.51 MB (340% larger than the native one).

Finally, the applications created with interpreted approach frameworks were the largest of the lot.

Table 1 shows the end sizes of the developed applications in ascending order of storage size used by each.

Framework	Development approach	Size in MB
Android with Java	Native	1.48
Cordova	Hybrid	1.74
Xamarin	Generated by cross-compilation	4.08
Corona	Generated by cross-compilation	6.51
Titanium	Interpreted	8.54
NativeScript	Interpreted	12.49

Table 1. Application sizes; Experiment #1

4.1.2 Use of Storage Space in Experiment #2

Audio-playing applications were generated following the standard procedure recommended by the documentation for each framework.

Table 2 displays the results obtained. The third column on the table details the total size of each of the generated applications. The fourth column shows the difference between total size and the size of the audio file, since this file has to be packed together with the code for each application. This allows seeing more clearly the size added by each framework.

For this experiment, the native approach was the best in terms of final application size. The APK file generated with the hybrid approach is just 32% larger than that generated with the native approach.

This difference increases with cross-compilation frameworks. The application generated with Xamarin is 196% larger than the natively generated application. The application produced by Corona was 388% larger than the one created using the native approach.

As in the previous experiment, the largest applications in MB are those produced with the frameworks used for generating interpreted applications. In this case, Titanium and NativeScript generated applications that were 564% and 1430% larger than those produced with the native approach.

Framework	Development approach	Size in MB	Size excluding audio file
Android with Java	Native	2.7	1.38
Cordova	Hybrid	3.14	1.82
Xamarin	Generated by cross-compilation	5.4	4.08
Corona	Generated by cross-compilation	8.05	6.73
Titanium	Interpreted	10.48	9.16
NativeScript	Interpreted	22.43	21.11

Table 2. Application sizes; Experiment #2

4.1.3 Use of Storage Space in Experiment #3

In the case of the applications that played video, the procedures recommended by their corresponding documentation were also followed.

Table 3 displays the results obtained. The same as in Table 2, the third column shows the total size of each application, and the fourth column displays the difference between total size and the size of the video file used. For result analysis, the last column was used.

In this experiment, application sizes followed the same order as that observed in experiments #1 and #2.

The native approach was once again the option that resulted in the smallest APK. The second place was taken by the hybrid approach, with Cordova. It was followed by cross-compilation frameworks, with Corona and Xamarin, respectively.

Lastly, the frameworks that resulted in the largest applications were those based on the interpreted approach.

Framework	Development approach	Size in MB	Size excluding video file
Android with Java	Native	90.24	1.04
Cordova	Hybrid	91.97	2.77
Xamarin	Generated by cross-compilation	94.21	5.01
Corona	Generated by cross-compilation	95.78	6.58
Titanium	Interpreted	98.43	9.23
NativeScript	Interpreted	101.67	12.47

Table 3. Application sizes; Experiment #3

4.1.4 General Result Analysis

Figure 1 displays the results for each of the 18 applications built, sorted by application final size.

As it can be seen, the order of the results obtained is the same in all three experiments.

The difference between the best and the worst framework is percentually significant in experiments #1 and #2, not so much in experiment #3. This is due, possibly, to the fact that the greatest impact of the framework on application size occurs when packing the basic components of the app. That is, the libraries, interpreters and other resources that the framework needs to include within the generated application for it to work.

A native application will be run directly by the underlying operating system, which means that it does not need to pack an interpreter or other tools within the application.

A hybrid application developed with web tools can be designed to be run on the web browser provided by the operating system, or it could include its own browser within the application, which would affect application size. A similar situation occurs with interpreted applications, which are run partly by the operating

system and partly by an interpreter. This interpreter can be either internal or provided by the platform.

Thus, the applications created with the native and hybrid approaches ended up being the smallest ones. The second place was taken by cross-compilation frameworks. Finally, interpreted frameworks generated the largest applications; this is due to the fact that both NativeScript and Titanium embed a JavaScript interpreter in the application.

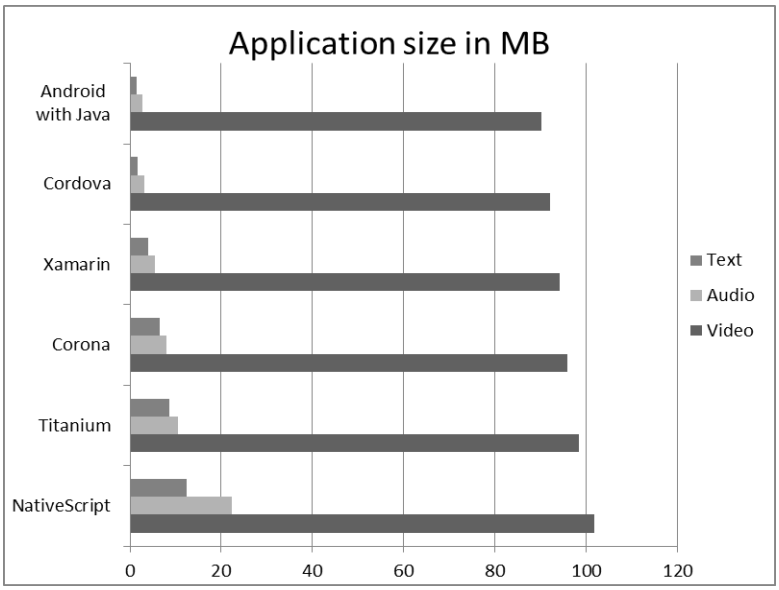


Figure 1. Summary of storage space measurements

5 Conclusions

In mobile application development, non-functional requirements are more relevant due to the restrictions imposed by devices: battery life, connectivity, limited development times, and significant fragmentation among the different platforms.

On the other hand, several mobile application development approaches are available. In this article, the native, interpreted, hybrid and cross-compilation approaches were analyzed.

A comparative study of the storage space used by applications for mobile devices, generated using different development approaches, was presented.

Three experiments were carried out – the first one consisted in displaying a text message on the screen, while the other two were designed to play multimedia files (audio and video, respectively).

It should be noted that the results presented in this paper are linked to the development framework versions used for the experiments and, therefore, could change in the future as these frameworks evolve.

The Android SDK framework corresponding to the native approach and the Apache Cordova framework corresponding to the hybrid approach yielded the best results as regards storage space usage in mobile devices.

Cross-compilation frameworks came in second.

The worst results were yielded by the frameworks that used the interpreted approach, since they generated the largest applications. This is mainly due to the fact that this development approach translates part of the code to native code and interprets the rest at runtime, which means that an interpreter has to be embedded within the application package.

In short, the experiments carried out help find out which framework would be more convenient, if size is a priority, for developing mobile applications similar to those presented in this article.

6 Future Work

In the future, we plan on expanding the number of frameworks tested, as well as considering any new versions of the frameworks used here.
On the other hand, iOS application size will also be studied, since iOS is the second most widely used operating system in the market.

References

1. Mona Erfani Joorabchi, Ali Mesbah, Philippe Kruchten. Real Challenges in Mobile App Development, ACM / IEEE International Symposium on Empirical Software Engineering and Measurement, Baltimore, Maryland, US, October 2013.
2. L. Delía, N. Galdamez, L. Corbalan, P. Pesado and P. Thomas, "Approaches to mobile application development: Comparative performance analysis," *2017 Computing Conference*, London, 2017.
3. http://gs.statcounter.com/os-market-share/mobile/worldwide [Last access: May 2018]
4. Delia, L.; Galdamez, N.; Thomas, P.; Corbalan, L.; Pesado, P., Multiplatform mobile application development analysis, Research Challenges in Information Science (RCIS), 2015 IEEE 9th International Conference on, Athens, Greece, 2015.
5. Corbalan L.; Fernandez Sosa J.; Cuitiño A.; Delia L.; Caseres G.; Thomas P.; Pesado P., Development Frameworks for Mobile Devices: A Comparative Study about Energy Consumption (ICSE), MobileSoft 2018 5th IEEE/ACM International Conference on Mobile Software Engineering and Systems on, Gothenburg Sweden, 2018.
6. K. Vandenbroucke, D. Ferreira, J. Goncalves, V. Kostakos y K. D. Moor, «Mobile cloud storage: a contextual experience,» Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI '14), pp. 101-110, 2014.
7. S. Tolomei, «Shrinking APKs, growing installs,» November 20, 2017. [Online]. Available: https://medium.com/googleplaydev/shrinking-apks-growing-installs-5d3fcba23ce2. [Last access: May 2018].

8. https://sensortower.com/blog/ios-app-size-growth [Last access: May 2018].
9. Spyros Xanthopoulos, Stelios Xinogalos, A Comparative Analysis of Cross-platform Development Approaches for Mobile Applications, BCI' 2013, Greece, 2013.
10. http://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/worldwide/#monthly-201403-201803 [Last access: May 2018].
11. Raj, C. R., & Tolety, S. B. (2012, December). A study on approaches to build cross-platform mobile applications and criteria to select appropriate approach. In India Conference (INDICON), 2012 Annual IEEE (pp. 625-629). IEEE
12. Tracy, K. W. (2012). Mobile application development experiences on Apple's iOS and Android OS. *Ieee Potentials*, 31(4), 30-34.
13. Delia, L., Galdamez, N., Thomas, P., Corbalan, L., & Pesado, P. (2015, May). Multi-platform mobile application development analysis. In *Research Challenges in Information Science (RCIS)*, 2015 IEEE 9th International Conference on (pp. 181-186). IEEE.
14. http://cordova.apache.org [Last access: May 2018].
15. http://www.appcelerator.com [Last access: May 2018].
16. https://www.nativescript.org/ [Last access: May 2018].
17. https://xamarin.com [Last access: May 2018].
18. https://coronalabs.com/ [Last access: May 2018].
19. https://gitlab.com/iii-lidi/papers/apps-size.git

Usabilidad de Framework Web: identificación de problemas y propuesta de evaluación

Marcela A. Constanzo y Sandra I. Casas,

GISP - Instituto de Tecnología Aplicada
Universidad Nacional de la Patagonia Austral
Río Gallegos, Argentina, mconst1978@hotmail.com

Resumen. Los frameworks web son artefactos que pueden ser reusados, son herramientas para construir rápidamente aplicaciones web a través de diferentes mecanismos. Deben contar con la documentación y el soporte adecuado para facilitar su uso. La usabilidad es uno de los requisitos de calidad más importantes en el desarrollo de productos "usables", indica la facilidad con la que un usuario puede usar una aplicación de software. Existen actualmente modelos de evaluación de usabilidad para aplicaciones móviles, sitio, aplicaciones web y escritorio, pero no para framework web. La evaluación de usabilidad aplicada a los frameworks web, tiene como objetivo mejorar la calidad del framework, ayudar a los usuarios a compararlos y seleccionarlos, identificando debilidades y realizando mejoras. Este trabajo propone una lista de verificación, apoyada en los estudios cualitativos desarrollados, tales como observaciones de 17 frameworks web más populares y análisis de los foros publicados.

Palabras Claves: Usabilidad, Framework web, Documentación, Foros

1 Introducción

Un framework web, es un conjunto de componentes con un diseño reutilizable que facilita y agiliza el desarrollo de aplicaciones Web. Se basan principalmente en el Modelo Vista Controlador (MVC), ofreciendo mecanismos para mejorar la integración con otras herramientas para la implementación de las capas de negocio y presentación [1].

Los frameworks son particularmente difíciles de aprender y comprender por los usuarios principiantes, especialmente si no son acompañados con la documentación apropiada [2] [3]. La documentación de buena calidad es un factor de éxito fundamental para la reutilización del framework ya que ayuda a la comprensión, guía a los desarrolladores en el proceso de personalización y explica sus principios de diseño y detalles y evita errores [2] [4].

La complejidad resulta de los diferentes públicos que lo utilizan, desarrolladores de aplicaciones, desarrolladores de frameworks, encargados de proyectos, de los diferentes estilos de documentos y diferentes notaciones. La identificación y clasificación de los problemas en el uso de estas herramientas, provenientes de los

usuarios de frameworks, puede aportar información para la propuesta de métodos de evaluación de usabilidad.

La usabilidad es una propiedad inherente a la calidad del software debido a que expresa la relación entre el software y su dominio de aplicación [3] [5]. La usabilidad es una de las características principales que utiliza el Modelo de Calidad de ISO/IEC 25010 y puede ser valorada en término de 4 subcaracterísticas, Comprensibilidad, Aprendibilidad, Operabilidad y Atractividad [6] [7]. La usabilidad aplicada a los frameworks web se refiere a la capacidad de un framework de ser entendido, aprendido, usado y protegido de errores por el usuario, es de interés tanto para desarrolladores de aplicaciones web como para los desarrolladores de frameworks web.

La evaluación de la usabilidad ha aportado diversos métodos y modelos principalmente enfocados a aplicaciones web, escritorio y móviles. Sin embargo en el contexto de herramientas de desarrollo de software, particularmente para el caso de frameworks web, no se encuentran claramente sistematizaciones de métodos, modelos, lineamientos y guías de usabilidad.

El presente trabajo expone resultados de estudios realizados a los recursos disponibles en los sitios oficiales de un grupo de frameworks web, que permitieron identificar los aspectos que presentan mayor dificultad en su uso y que características debe cumplir la documentación que acompaña a los mismos con el objeto de confeccionar una lista de verificación para evaluar la usabilidad de los frameworks web.

Este estudio continúa de la siguiente manera, en la Sección 2 se describe el análisis de documentación y recursos seleccionados y sus resultados, en la Sección 3 el análisis de los foros y sus resultados, en la Sección 4 se presenta la implementación de la lista de verificación en cinco sitios web orientada a mejorar la usabilidad de los frameworks web y en la Sección 5 se detallan las conclusiones y futuros trabajos.

2 Análisis de los recursos y soportes de documentación

Con el objeto de identificar qué tipo de documentación es la más utilizada y cuáles son las características más relevantes que permiten a los usuarios facilitar el uso de frameworks web, se realizó un estudio exploratorio y descriptivo sobre el contenido de los distintos tipos de soporte y recursos disponibles y publicados en los sitios oficiales de los frameworks web populares entre los desarrolladores. Se buscó analizar todo tipo de soporte y recurso que se provea y sirva para el uso del framework, independientemente de su formato (documento textual/ visual/ etc.).

Se comenzó con la selección de los frameworks, utilizando los rankings disponibles en la web, tal y como se describe en [8]. Posteriormente se procedió a realizar un examen exhaustivo de los sitios oficiales de los frameworks web identificando y analizando los recursos y soportes disponibles. Se consideró soporte a todo tipo de recurso, documento, libro, tutorial, get started, video, ejemplos, como así también todo servicio que preste el sitio para que el usuario pueda evacuar sus dudas o plantear consultas, tal como blogs, foros, github, chats, lista de mails. Se

identificaron los componentes prioritarios de cada uno de los recursos y las características comunes a todos los frameworks.

Las Tabla 1 y Tabla 2, aportan fundamentalmente dos datos: a) la cantidad de recursos de documentación por framework y b) la cantidad de frameworks que usan los distintos Recursos.

Tabla 1. Cantidad de Recursos disponibles por Framework web.

FRAMEWORK	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17
RECURSOS	6	10	6	8	9	7	9	9	6	7	6	7	8	7	6	6	4

F1	Laravel	F4	Ruby on Rails	F7	Symfony	F10	Zend	F13	React	F16	Vue
F2	Yii	F5	AngularJS	F8	Ember.js	F11	Backbone.js	F14	Aurelia.js	F17	Mercury
F3	Codelgniter	F6	Meteor	F9	CakePHP	F12	Knockout.js	F15	Polymer		

Tabla 2. Cantidad de Framework web por Recursos.

RECURSOS	F	B	C	L	O	G	M	I	T	U	E	V	A	H
FRAMEWORK	7	12	3	5	5	14	1	6	9	1	15	10	17	16

F	Foro	O	Comunidad	G	GetStarted	U	Guías	A	API
B	Blog	M	Manuales	I	Libros	E	Ejemplos	H	GitHub
L	Lista de Mails	C	Chat	T	Tutoriales	V	Videos		

Este trabajo se centra en la usabilidad y se basa en la Norma ISO/IEC 25010/11, adaptándola particularmente a los frameworks web, así la Usabilidad de un framework web es la capacidad del framework para ser entendido, aprendido, usado, cuando se usa bajo determinadas condiciones. Se valorará en base a las siguientes características: *comprensibilidad* (capacidad del framework que permite al usuario entender si es adecuado para sus necesidades, bajo condiciones de uso), *aprendibilidad* (se refiere a la facilidad de aprender la funcionalidad básica, para que el usuario realice la tarea), *operabilidad* (capacidad del framework para permitir al desarrollador operar con él y controlarlo) y *protección contra errores de usuario* (capacidad del framework para proteger a los usuarios de cometer / incurrir errores).

Los recursos identificados se pueden relacionar con la usabilidad del framework, ejemplos, tutoriales, get started, videos, con la aprendibilidad, debido a que esta característica se refiere a la facilidad de aprender la funcionalidad básica, para que el usuario realice la tarea correctamente y estos soportes, correctamente documentados, permiten eso. Los foros, blogs, github, chats, comunidades, presentan relación directa con la operabilidad, esta capacidad de usabilidad que permite al usuario operarlo y controlarlo con facilidad, este recurso ayuda, mediante la interacción de los usuarios, a operar los diferentes componentes que lo integran y evitar o resolver los errores que puedan presentarse; también se asocia con la aprendibilidad ya que permiten aprender las funcionalidades que ofrece el framework.

Las APIs, tutoriales y github son los recursos que permiten comprender si el framework se adapta a las necesidades que el usuario necesita, en el caso de las APIs por ser código que forma parte del mismo, y tutoriales y github por tener información que explica el contenido y la estructura del frameworks y permite conocer de una manera correcta los elementos que lo componen. Los foros, blogs, get started, manuales, libros, tutoriales, guías, ejemplos, videos, APIs y github se pueden asociar

con la protección contra errores porque son recursos que permiten cierta protección a los usuarios sobre los errores que puedan presentarse al usar el framework.

Del estudio realizado y presentado precedentemente se desprende que los recursos de documentación que facilitan a los desarrolladores el uso de los frameworks son variados en tipo y formato.

Sobre los recursos:

- El promedio de recursos disponibles en los sitios web oficiales por framework es de 7,4. Mediana y Moda =8 y una desviación Estándar: 1,37, es decir que cada valor observado (total de recursos por framework) difiere en promedio un 1,37 de la media ponderada de 7,4. De un total de 122 recursos disponibles, 48 son de Interacción, 49 de Texto, 10 Audiovisual y 15 Audiovisual/Texto.

Se pudo identificar características de relevancia en cada uno de los recursos seleccionados y analizados:

- Para el caso de las APIs, se determinó que los comunes a la mayoría de ellos son los Paquetes, Clases, Métodos e Interfaces. Para los get started se puede concluir que abordan temas resumidos como Instalación, Configuración, Conexión a la base de datos y muestran código para desarrollar Primeras aplicaciones. Los Tutoriales, se explayan sobre temas como Plantillas Estáticas, Plantillas dinámicas, Enlace de datos (conexión a la base de datos), Enrutamiento, Vistas, Formularios, Controlador de eventos, Modelos, Configuración, Aplicaciones y Seguridad, en formato Html con texto, código fuente e imágenes.
- Los videos muestran una introducción al mismo, utilidades y en muchos casos relatan experiencias en relación a su uso y los ejemplos, demos en algunos de los casos, muestran aplicaciones desarrolladas utilizando el framework, y comentan código para realizar acciones básicas con el mismo. La plataforma de desarrollo colaborativa github permite almacenar código de forma pública y es una manera de facilitar a los usuarios el uso del framework.

Sobre frameworks y recursos:

- La documentación de las APIs, se encuentra disponible en el 100% de los sitios oficiales de los frameworks web.
- Entre los recursos más utilizados por los usuarios de los frameworks web encontramos la plataforma github, los ejemplos ya sean código o aplicaciones desarrolladas con el framework y el get started.
- Los ejemplos se encuentran disponibles en 15 de los 17 sitios, son prácticos para los usuarios ya que les permite la visualización concreta de aplicaciones creadas utilizando el framework.
- El get started o getting started es un recurso disponible en el 82% de los sitios pero solo brinda las pautas para comenzar a utilizar el framework.
- En un 76 % se encuentran los blogs y en un 78% los videos, recurso muy utilizado por los usuarios para visualizar, de manera práctica, actividades concretas.
- A diferencia de los get started, los tutoriales se encuentran disponibles en el 53% de los sitios examinados, su contenido debe detallar información básica que sirva a los usuarios para poder conocer y usar el framework. Existen recursos presentes en menor medida, como manuales y guías, sustituidos en mayor medida por los tutoriales, por contener información concreta y resumida.

En relación a las características de usabilidad y la documentación clasificada:

- Los recursos se pueden asociar a las características de usabilidad para frameworks web, Aprendibilidad, Operabilidad, Comprensibilidad y la Protección de Errores. La Aprendibilidad está cubierta por los 3 tipos de recursos, texto, interacción y audiovisual, la Operabilidad por los recursos de tipo texto e interacción, la Comprensibilidad por el texto y la interacción y la protección de errores por los tres tipos.

3 Análisis de Foros

En este análisis el enfoque es particular sobre los foros, un foro es una aplicación web, generalmente organizada en categorías, que da soporte a discusiones u opiniones en línea. Son contenedores en los que se abren temas de discusión y en los cuales los usuarios cumplen el rol más importante, por la interacción y el dinamismo, ya que participan con consultas, preguntas y también responden con sus opiniones. Esta interacción de los usuarios permite obtener un resumen de los temas consultados por los usuarios del framework web, categorizarlos y organizarlos de acuerdo a su importancia en relación a la usabilidad.

Se trabajó sobre los frameworks seleccionados anteriormente, para identificar y clasificar las entradas a los foros que sean de interés para la usabilidad de los frameworks web, que como primera etapa se identificaron aquellos que cuentan con foros actualizados en sus sitios oficiales, 6 frameworks, resultando los mismos Knockout.js, Yii, Codelgniter, Meteor, React y Laravel.

El proceso para el estudio de los foros consistió en la realización de un análisis sobre cada uno, para conocer su organización, la estructura que presentan, secciones, y la cantidad de entradas a la fecha establecida. Una primera selección para identificar aquellas secciones que son de interés y que podrían brindar información sobre APIs, procesos de instalación y configuración del framework, plantillas, testing, errores y preguntas, y posteriormente descartar aquellas que no aporten información de relevancia para el análisis.

La selección de las categorías para la clasificación está relacionada al diseño del framework web y sus componentes, en aquellos que directamente pueden influir en la usabilidad, asociada a las características de comprensibilidad, aprendibilidad y operabilidad y protección de errores – Tabla 3.

El procedimiento consistió en extraer todas las entradas de cada una de las secciones de los foros, y clasificarlas de acuerdo a las categorías especificadas en la Tabla 3. Se clasificaron las 20104 entradas en cada una de las categorías existentes. Posteriormente se analizaron y clasificaron las que corresponden a la formulación de preguntas. Respecto de las 8389 entradas no categorizadas, se realizó un filtro por temática y se agregaron nuevas categorías tales como base de datos (BD); sintaxis (SX), versiones (V) y servidor (S) relacionada al localhost y servidores existentes. 4) Sobre el remanente se procedió a realizar un análisis detallado para poder clasificar, en relación a las palabras claves definidas, aquellas entradas relacionadas con estas, reduciendo la cantidad S/C a 6189.

Por último se confeccionó una nueva tabla resumen agregando las últimas categorías creadas. Las entradas que no pudieron ser categorizadas son descartadas del presente estudio.

La Tabla 3 expone el resumen final de los foros y su clasificación, en relación a la información establecida, incluyendo la búsqueda por palabras claves con el fin de reducir las entradas que no pudieran ser clasificadas en el primer paso.

Del total de entradas se identificaron un grupo de ellas que refieren a preguntas concretas sobre ciertos temas relacionados con las categorías definidas o no, pero solicitan ayuda o asistencia para realizar alguna función o acción con un componente del framework, esto también se expresa en la Tabla 3.

Tabla 3. Clasificación final de las entradas de los foros y resumen de preguntas

	A	*C*	*Es/c*	*EA*	*EC*	*EI*	*EP*	*ET*	*I*	*P*	*T*	*BD*	*SX*	*V*	*S*	*PRE*	*S/C*	*TOTAL*
ENT.	1454	411	3445	385	60	137	62	32	184	461	352	382	1078	398	342	4732	6189	20104
PRE	400	149	256	5	1	0	4	3	58	139	0	249	829	221	163			4732

Clasificación	Abrev.	Clasificación	Abrev.	Clasificación	Abrev.
Errores (general)	E S/C	Errores de APIs	EA	Instalación	I
Errores de Instal.	EI	Errores de Plantilla	EP	Configuración	C
Errores de Config.	EC	Errores de Testing	ET	Plantillas	P
APIs	A	Testing	T	Preguntas	PRE
				Sin clasificación	S/C

El estudio de los foros de 6 frameworks web, sobre un total de 20104 entradas, indica que el principal problema de los usuarios son los errores, principalmente los errores relacionados con el código, la conexión a las bases de datos y el uso de las APIs. Esto podría estar orientado a la escasa documentación existente en los sitios oficiales en relación a estos temas o a cómo realizar determinadas tareas tales como la instalación o configuración de los componentes y sobre todo la implementación de APIs.

La revisión realizada sobre los recursos existentes tales como tutoriales, ejemplos, get started, videos, documentación de APIs, no cuenta con información sobre los errores que puedan surgir con el uso del framework, por eso se indica que la documentación es escasa, insuficiente, incompleta y se complementa con el análisis sobre los foros.

4 Lista de verificación para evaluar y mejorar la usabilidad

Se presenta primeramente el diseño de un esquema de usabilidad basado en el enfoque [9] que permite valorar y mejorar la usabilidad de los frameworks web, adaptándolo a la propuesta de la lista de verificación. Este esquema analiza el sitio web oficial del framework a evaluar, los recursos de documentación disponibles, y la interacción con los desarrolladores de aplicaciones web para alcanzar sus objetivos. El framework tiene definidos atributos de usabilidad, y el método asociado para evaluar la usabilidad, la lista de verificación, basada en la documentación que acompaña al framework. Los componentes se encuentran definidos dentro de un contexto específico de uso tal y como se observa en la Figura 1.

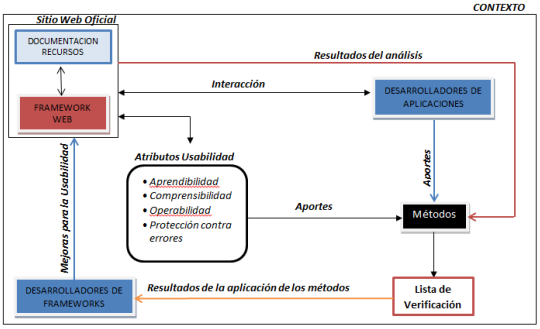


Figura 1 – Esquema de usabilidad adaptado para Frameworks web

La lista de verificación tiene el objetivo de valorar el sitio web oficial de un frameworks web. Esta lista es una herramienta simple de aplicar, de manera rápida permite obtener información objetiva sobre las fortalezas y debilidades del sitio.

La información recabada es de tipo cualitativa, por lo que se hará una equivalencia numérica para poder definir un umbral de evaluación sobre la cantidad de recursos de documentación presentes en el framework, entre otros aspectos, tal como se observa en la Figura 2.

Dado que la conversión entre datos cualitativos y cuantitativos no es única, se trabajará con la siguiente escala (o referencia): 0: No, 1: Si con restricciones, 2: Si aceptable y 3: Si ideal.

Estos umbrales de "calidad", se establecen a criterio del equipo de trabajo y son comunes dentro de análisis similares como en [10] y [11]. La figura 2 muestra la lista de verificación.

Calificación de la evaluación: para cada pregunta con respuesta (0) no se asignarán puntos. Si la respuesta es (1), se aplicarán 4 puntos, si la respuesta es (2) se asignarán 7 puntos y si la respuesta es (3), se asignarán 10 puntos.

Las respuestas 0 y 3 son los dos extremos posibles, nada de información y por otro lado la información completa y deseable. Las respuestas 1 y 2 deberían seguir la siguiente lógica:

- Dos respuestas de tipo 1 son mejores que una respuesta de tipo 0 y una respuesta de tipo 2.
- Una respuesta de tipo 1 y una respuesta de tipo 2 son mejores que una respuesta de tipo 0 y una respuesta de tipo 3.

Una de las posibilidades numéricas de esta lógica es la detallada, es decir, puntajes de 0, 4, 7 o 10 de acuerdo al tipo de respuesta.

Interpretación de la calificación: lo deseable es que el puntaje total sea 180. Sin embargo, más importante que el número alcanzado en esta evaluación, es la posibilidad de identificar potenciales problemas que pueden influir en la usabilidad de los frameworks web.

La lista de verificación presentada precedentemente permite verificar si ésta es adecuada o si es recomendable revisar la documentación del framework si recurre en alguna de las siguientes consideraciones:

1) Si alguna de las dos preguntas consideradas fundamentales es valorada en 0.

Independientemente de la calificación asignada a cada respuesta existen items considerados fundamentales, esto significa que la respuesta asignada a los mismos no puede ser igual a 0, estos puntos fundamentales son:

- ¿Las APIs que implementa el framework están correctamente documentadas, incluyen información sobre paquetes, interfaces, clases y métodos?
- ¿La documentación que acompaña al framework está actualizada y acorde con las versiones del mismo?

Orientadas al sitio web del framework	**0**	**1**	**2**	**3**
¿Provee ejemplos de aplicaciones concretas realizadas con el framework?				
¿Incluye ejemplos claros y concretos de diferentes operaciones a realizar y para diferentes niveles de usuarios?				
¿Se vincula con una plataforma de desarrollo colaborativo, como por ejemplo GitHub?				
¿Provee un foro activo?				
¿El nivel de acceso de los recursos es directo o como opción de menú principal?				
Orientadas a la comunidad				
El framework ¿cuenta con una comunidad activa de soporte?				
¿El foro presenta una estructura organizada?				
¿Los usuarios del foro se encuentran clasificados de acuerdo a su expertiz?				
¿Existe una sub-comunidad de usuarios experimentados?				
Orientadas a reducir errores				
¿Las APIs que implementa el framework están correctamente documentadas, incluyen información sobre paquetes, interfaces, clases y métodos?				
¿Se provee un documentación (catálogos, lineamientos, etc) de errores?				
¿Se incluye en la documentación de las diferentes operaciones a realizar por el usuario del framework (instalación, configuración, implementación de APIs, conexión a bases de datos), los errores más comunes detectados?				
¿Se incluye documentación de cada componente que forma parte de él?				
¿Se proveen ejemplos de código sobre funcionalidades básicas?				
¿El recurso get started detalla en su contenido los items instalación, configuración, conexión a B.D. y primeras aplicaciones?				
¿El tutorial documenta en su contenido los items plantillas estáticas, plantillas dinámicas, enlace de datos (conexión a la base de datos), enrutamiento, vistas, formularios, controlador de eventos, modelos, configuración, aplicaciones y seguridad?				
¿El foro incorpora en su estructura las categorías o subcategorías relacionadas con archivos, pruebas, instalación y configuración.				
¿La documentación que acompaña al framework está actualizada y acorde con las versiones del mismo?				

Figura 2 – Esquema de usabilidad adaptado para Frameworks web

En relación a las APIs, es de vital importancia que las mismas se encuentren documentadas en el sitio, ya que forman parte del framework y se debe conocer la mayor cantidad de información sobre ellas para usar correctamente el mismo.

También es considerado imprescindible que la documentación que acompañe al frameworks este actualizada y acorde a las versiones porque de no ser así esto llevaría a errores que provocarían el uso incorrecto del mismo por cuestiones externas al usuario.

2) Si presenta 4 o más respuestas con calificación igual a 0, puesto que esos 4 o más ítems sin calificar no pueden ser “compensados” por la calificación perfecta de otros ítems.

3) Si el promedio de la suma de las valoraciones es inferior al 70% del caso ideal.

Se estableció un umbral de 70% para el análisis de calidad teniendo en cuenta lo siguiente: luego de evaluar los dos ítems fundamentales, los 16 ítems restantes para ser considerados válidos deben contener (caso mínimo válido) 5 respuestas de tipo 1, 5 respuestas de tipo 2 y 6 respuestas de tipo 3. Esto da un total de 115 puntos, lo que equivale a un 71.875%, sin embargo, para no utilizar un umbral tan "complejo", se redondeó a 70%, valoración igual o superior a 112 puntos (70% de 160 puntos).

4.1 Aplicación

Como aplicación de la lista de chequeo propuesta se seleccionaron 5 frameworks web de similares características, que cuentan con un sitio web oficial. Los frameworks elegidos fueron: Solar, KumbiaPHP, Phalcon, Django, Web2py. Se procedió al

análisis del sitio oficial de cada uno de ellos, –Figura 1- una vez finalizado el mismo, se obtuvieron los siguientes valores:

FRAMEWORK SOLAR VALORACION/PUNTAJE = 56.

RECOMENDACIONES = Se aconseja revisar la documentación que acompaña al framework, disponible en el sitio oficial, por contener 8 respuestas con puntaje = 0. Se debería revisar los siguientes ítems, tal como incluir ejemplos, mantener un foro activo y bien estructurado, proveer un catálogo de errores y un tutorial completo.

FRAMEWORK KUMBIA VALORACION/PUNTAJE = 113.

RECOMENDACIONES = Se considera que cuenta con documentación aceptable para el uso del framework, sin embargo, se identifican debilidades en la documentación, tal como mejorar la información sobre las APIs y completar el tutorial con la información mínima indispensable.

FRAMEWORK PHALCON VALORACION/PUNTAJE = 94.

RECOMENDACIONES = Se aconseja revisar la documentación que acompaña al framework, por obtener una valoración inferior al valor asignado (112) 70%. Se debería revisar al menos algunos de los ítems, principalmente en los orientados a la comunidad y a la reducción de errores.

FRAMEWORK DJANGO VALORACION/PUNTAJE = 72.

RECOMENDACIONES = Se aconseja revisar la documentación que acompaña al framework, disponible en el sitio oficial, por contener 7 respuestas con puntaje = 0. Se debería revisar los siguientes ítems, tal como incluir ejemplos, mantener un foro activo y bien estructurado y proveer un catálogo de errores.

FRAMEWORK WEB2PY VALORACION/PUNTAJE = 59.

RECOMENDACIONES = Se aconseja revisar la documentación que acompaña al framework, disponible en el sitio oficial, por contener 8 respuestas con puntaje = 0. Se debería revisar los siguientes ítems, tal como incluir vinculación con una plataforma colaborativa, mantener un foro activo y bien estructurado y proveer un catálogo de errores.

5 Conclusiones

En la actualidad los frameworks web son ampliamente usados por los desarrolladores, dado que permiten agilizar el proceso mediante la incorporación de funcionalidades ya desarrolladas y probadas, por tanto, utilizar un framework que cumple criterios de usabilidad, contribuirá a facilitar su uso y ahorrar costos. La importancia de la usabilidad, ha originado numerosos estudios, métodos y modelos para evaluar la usabilidad, con especial énfasis en aplicaciones de usuario final (web-escritorios, móvil, etc.) [12] [13] [14]. Se encuentran algunas propuestas para herramientas de desarrollo como APIs y componentes [15] [16], para frameworks web son escasos los aportes.

Este trabajo de investigación permitió identificar los recursos de documentación más utilizados y las características más relevantes de cada uno de ellos, asociando los atributos de usabilidad de los frameworks web, identificados mediante un análisis exploratorio sobre un conjunto de 17 frameworks web más populares entre los desarrolladores, seleccionados de rankings disponibles en la web.

Mediante el estudio de los foros, se identificaron los aspectos que presentan mayor dificultad y los errores más comunes consultados por los usuarios de los frameworks web, categorizados y organizados de acuerdo a su importancia en relación a la usabilidad (comprensibilidad, aprendibilidad, operabilidad y protección de errores).

Se propone un método para evaluar y mejorar la usabilidad de los frameworks web, que permite de manera rápida y sencilla detectar cuáles son las fortalezas y debilidades que presenta la documentación que acompaña al framework web y que se encuentra disponible en el sitio oficial del mismo. La lista de verificación permite realizar una evaluación de la usabilidad de los frameworks web, en función a estos recursos.

Las restricciones del presente trabajo están asociadas a que el estudio se basa en los frameworks para el desarrollo de aplicaciones web y que cuenten con un sitio web oficial. El estudio se centra en la documentación que acompaña al framework web, sin analizar su diseño ni la complejidad interna del mismo.

En relación a los trabajos futuros los mismos se orientan a profundizar el estudio realizado a los foros, concentrándose en los errores, con el fin de proponer documentación específica para el mismo, realizar un estudio sobre la estructura de los frameworks web para ampliar la evaluación de usabilidad e identificar los atributos asociados a la complejidad de los mismos y ampliar el método propuesto para evaluar otros factores.

Referencias

1. Gutiérrez, J. J.:¿Qué es un framework web? Escuela Técnica de Ingeniería Informática – Departamento de Lenguajes y Sistemas Informáticos, sitio web: http://www.lsi.us.es/~javierj/investigacion_ficheros/Framework.pdf (2018)
2. Taligent, P.: Building Object-Oriented Framework, Addison-Wesle (1994)
3. Greg, B., Pierre, D.: Documenting frameworks, In Building Application Frameworks – Object-Oriented Foundations of Framework Desing (2006)
4. Mohamed, F., Douglas, S.: Object-oriented application frameworks, Communications of the ACM (1997)
5. Bertoa, M.F., Vallecillo, A.: Medidas de Usabilidad de Componentes Software, IEEE, Latin América Transactions, Vol. 4, Nro. 2 (2006)
6. International Standard ISO/IEC 25010 – First editons (2011)
7. Alfonso, P.L.: Revisión de modelos para evaluar la calidad de productos Web, Experimentación en portales bancarios del NEA, Universidad Nacional de la Plata – Facultad de Informática (2012)
8. Constanzo, M. A., Casas, S.: Usability Evaluation of Web Support Frameworks, XLII Conferencia Latinoamericana de Informática. Simposio Latinoamericano de Ingeniería de Software. ISBN: 978-1-5090-1632-7 (USB) (2016)
9. Fernandez A., Insfran E., Abrahao S.: Modelo de Usabilidad Web alineado con el SQuaRE para procesos de desarrollo dirigido por modelos (2010)
10. Araujo M.: Indicadores de Calidad: Criterios para la definición de umbrales – Gobierno de Chile – Superintendencia de Salud – Intendencia de Prestadores (2010)
11. Leyva Parra G., Páez Domínguez O., Sainz López M.,: Un umbral empírico y otras recomendaciones para el reporte de la confianza del consumidor en México, vol. 7, nro. 1 (2016)
12. Nielsen J., Usabilidad. Diseño de Páginas Web, Prentice Hall (2000)
13. Ivory M. Y., Hearst, M. A., The state of the art in automating usability evaluation of user interfaces, Journal ACM Computing Surveys (CSUR), vol. 33, no. 4, pp. 470−516 (2001)
14. Zhang D., Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications, vol. 18, Issue 3 (2005)
15. Bertoa M. F., Vallecillo A., Medidas de Usabilidad de Componentes Software, IEEE LATIN AMERICA TRANSACTIONS, VOL. 4, NO. 2 (2006)
16. Minhaz F., Zibran Farjana Z., Eishita Chanchal K., Factors Affecting the Usability of APIs, Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada S7N 5C9 (2011)

SGR: Sistema para la Gestión de Requerimientos

Rípodas Alejandra, Aguirre Verónica, Delía Lisandro, Marrero Luciano
{aripodas, vaguirre, ldelia, lmarrero}@lidi.info.unlp.edu.ar

Instituto de Investigación en Informática LIDI
Facultad de Informática UNLP

Abstract. La ingeniería de requerimientos es una etapa clave en el ciclo de vida del software. Actualmente existen en el mercado diversas herramientas que dan soporte a dicha etapa. Al analizar y comparar un conjunto de herramientas, se identificaron las características que consideramos debe tener un producto de software para la gestión de requerimientos. Del análisis comparativo se concluye que no existe una herramienta, que cumpla con la totalidad de las características consideradas. Es por esto que se decidió desarrollar un producto de software SGR (Sistema para la Gestión de Requerimientos) que brinde soporte a la fase de ingeniería de requerimientos y que reúna todas las características consideradas. A través de la creación de SGR se pretende aportar una herramienta colaborativa que fomente la participación del cliente para logra obtener un producto final exitoso y que genere el SRS basado en el estándar IEEE 830-1998.

Keywords: Requerimiento, Ingeniería de Requerimientos, Herramienta CASE

1. Introducción

Los requerimientos reflejan las necesidades de los clientes. Son un elemento clave en el desarrollo de un proyecto de software, ya que marcan el punto de partida para actividades como planeación, estimación de esfuerzo, definición de recursos, costos, etc. Inician cuando comienza un proyecto y pueden variar a lo largo de este.

La extracción y el análisis de los requerimientos sirve para dos propósitos diferentes, pero relacionados. Por una parte, la extracción nos permite escribir un documento de definición de requerimientos; escrita en términos que el cliente puede entender. Representa una comprensión entre el cliente y el desarrollador de lo que se necesita o desea, y por lo general es escrito en forma conjunta entre el cliente y el desarrollador. Por otra parte, la especificación de requerimientos reitera la definición en los términos técnicos apropiados para el desarrollo del diseño de un sistema [1].

La especificación de requerimientos deber ser clara, completa y estar libre de ambigüedades. Es por esto que, se considera que la mejor forma de realizarla es mediante el uso del método estructurado. Este método propone la definición y el uso de plantillas para asegurar que los requerimientos estén expresados de forma uniforme. De esta manera se minimizan los errores relacionados con la interpretación durante las distintas etapas de un proyecto. Por otro lado, la especificación

estructurada de requerimientos guía a analistas inexpertos en el desarrollo de la especificación de requerimientos.

En [1] a través de una investigación empírica sobre errores de software, se observa que el 48% de los defectos encontrados en proyectos de software de mediana escala eran “atribuidos a especificaciones funcionales o requerimientos incorrectos o mal interpretados”. Asimismo, en [1] se afirma que “encontrar y corregir los defectos en la fase de desarrollo del software es más costoso que hacerlo en la fase de requerimientos”.

Con el paso de los años el uso de herramientas para dar soporte a la gestión de requerimientos se convirtió en un factor importante en el desarrollo del software. Estas herramientas suelen ser herramientas CASE (Computer-Aided Software Engineering), y tienen como objetivo hacer que el desarrollo de un proyecto sea un proceso eficientemente organizado. Existe una gran cantidad de estas herramientas, las cuales permiten controlar proyectos complejos y brindan funcionalidades para especificar, organizar, almacenar y gestionar requerimientos. De esta manera se intenta reducir errores que a largo plazo puedan perjudicar el correcto desarrollo de un proyecto [2].

Con el objetivo de detectar un conjunto mínimo y deseable de características a cumplir por una herramienta CASE, se analizaron y compararon una serie de herramientas para la gestión de requerimientos. Del análisis comparativo se concluye que no existe una herramienta entre las analizadas, que cumpla con la totalidad de los aspectos esperados.

Es por esto que se decidió desarrollar un producto de software que reúna las características evaluadas brindando soporte a una de las actividades más importantes en el desarrollo de un proyecto, como es la fase de Ingeniería de Requerimientos.

En la sección 2 se muestra un análisis de un conjunto de herramientas de gestión de requerimientos y se establecen características que consideramos debe contener una herramienta CASE para la gestión de requerimientos. En la sección 3 se presentan SGR (Sistema para la Gestión de Requerimientos) y sus funcionalidades más importantes. Por último se presentan conclusiones y trabajos futuros.

2. Análisis de herramientas para la gestión de requerimientos

Los autores de [3] determinaron las características que debería tener una herramienta para la gestión de requerimientos en base al documento ISO/IEC TR 24766: 2009. Este establece un total de 157 capacidades, las cuales fueron agrupadas en 6 categorías principales: obtención de requerimientos, análisis de requerimientos, especificación de requerimientos, verificación y validación de requerimientos, administración de requerimientos y otras capacidades.

Tomando como referencia las categorías definidas anteriormente se evaluaron tres herramientas del ámbito académico (REM [4], Heler [5] y Let’s Req! [6]) y siete del ámbito comercial (Visure Requirements [7], DOORSNG [8], Rommana [9], TraceCloud [10], Psoda [11], TrackStudio [12] Y Redmine [13]). Como resultado del análisis realizado se identificaron las características mínimas y deseables que debería

tener una herramienta CASE para la gestión de requerimientos. Estas características son:

- **Diferencia entre requerimientos:** brindar la posibilidad de definir requerimientos funcionales y no funcionales de forma separada.
- **Objetivos de proyecto:** crear objetivos del proyecto y asignarlos a los requerimientos para establecer sus fines.
- **Roles y permisos:** permitir el acceso al sistema mediante autenticación, restringiendo las funcionalidades que se pueden realizar.
- **Versiones de requerimientos:** asignar números de versiones a los requerimientos para indicar los cambios que transitó.
- **Estado**: asignar el estado en que se encuentra tanto un requerimiento como un proyecto.
- **Stakeholder**: definir personas que influyan directa o indirectamente en la gestión de requerimientos.
- **Prioridad:** asignar prioridad para indicar qué requerimientos son más importantes.
- **Estabilidad**: definir y asignar estabilidad a requerimientos funcionales y no funcionales para indicar el nivel de modificación.
- **Historial de cambios de requerimientos:** llevar un registro de los cambios de cada requerimiento para realizar su seguimiento.
- **Documentación externa:** permitir el almacenamiento de documentación externa en diferentes formatos (pdf, doc, odt, jpg, etc.), de esta manera se complementa la información tanto de proyectos como de requerimientos funcionales y no funcionales.
- **Generación automática del IEEE Std. 830-1998:** generar automáticamente documentos basados en el estándar de la IEEE 830-1998 en formato PDF para contar con los datos sin necesidad de acceder al sistema.
- **Trazabilidad:** permitir identificar el origen y las relaciones entre los requerimientos con el fin de mostrar la matriz de trazabilidad correspondiente.
- **Trabajo colaborativo:** permitir que los diferentes usuarios de un proyecto puedan trabajar de forma colaborativa para agilizar la gestión de requerimientos.
- **Comunicación interna:** brindar un canal de mensajería interna para realizar la comunicación a través de la herramienta.
- **Notificaciones por email:** realizar el envío automático de notificaciones por email para mantener actualizados a los miembros del proyecto.
- **Claridad y facilidad de uso:** brindar una interfaz intuitiva para agilizar su utilización.
- **Interfaz web adaptable:** brindar una interfaz web adaptable para satisfacer la necesidad de utilizar la herramienta desde diferentes dispositivos como Tablets, Smartphones, etc.
- **Multiplataforma:** permitir utilizar la herramienta desde distintas plataformas independientemente del sistema operativo del usuario.

Luego, se realizó un cuadro comparativo (Ver Tabla 1) en el que se puede visualizar el comportamiento de cada herramienta con las características

anteriormente mencionadas. En Tabla 1 se puede observar que ninguna de las herramientas estudiadas satisface la totalidad éstas características.

A partir de esta comparación, se decidió realizar un sistema de software denominado SGR que reúna todas las características que consideramos necesarias y deseables que debe tener un sistema para la etapa de gestión de requerimientos.

Tabla 1. Cuadro comparativo de herramientas seleccionadas.

Características \ Herramientas	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
Diferencia entre requerimientos	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Objetivos de proyecto	✓	✓	✓	✓	✓	-	-	✓	-	-
Roles y permisos	-	✓	-	✓	-	✓	✓	✓	✓	✓
Versiones de requerimientos	✓	✓	✓	-	✓	✓	-	-	✓	-
Estado	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Stakeholders	✓	-	✓	✓	✓	-	-	-	-	-
Prioridad	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Estabilidad	✓	-	✓	-	-	✓	✓	✓	-	-
Historial de cambios de requerimientos	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Documentación externa	-	✓	-	-	-	✓	✓	✓	✓	✓
Generación automática del IEEE Std. 830-1998	-	-	✓	-	-	-	-	-	-	-
Trazabilidad	✓	✓	✓	✓	✓	✓	✓	✓	-	✓
Trabajo colaborativo	-	✓	-	✓	✓	✓	✓	✓	✓	✓
Comunicación interna	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Notificaciones por email	-	✓	-	✓	-	✓	✓	✓	✓	✓
Claridad y facilidad de uso	✓	✓	✓	-	✓	-	-	-	-	✓
Interfaz web adaptable	-	-	-	-	-	-	-	-	-	✓
Multiplataforma	-	-	-	✓	✓	✓	✓	✓	✓	✓

H1: REM, H2: Visure Requirements, H3: Heler, H4: DOORSNG, H5: Let's Req!, H6: Rommana, H7: TraceCloud, H8: Psoda, H9: TrackStudio, H10: Redmine

✓= Cumple con la funcionalidad, - = No cumple

3. SGR (Sistema para la Gestión de Requerimientos)

En esta sección se presenta SGR, una herramienta CASE para la etapa de elicitación de requerimientos. Además se expone un caso de estudio donde se la utilizó y por último, se muestra una comparación entre SGR y las herramientas analizadas en la sección 2.

3.1 Funcionalidades destacadas

3.1.1 Administración de Proyectos

SGR permite la creación de proyectos compuestos por: requerimientos funcionales, no funcionales, objetivos, usuarios asignados, archivos adjuntos, atributos, stakeholders y versiones. Además, permite mantener historial de cambios e indicar el estado del proyecto en un momento determinado (Figura 1).

Es posible realizar filtros y establecer órdenes para mejorar la visualización de los datos.

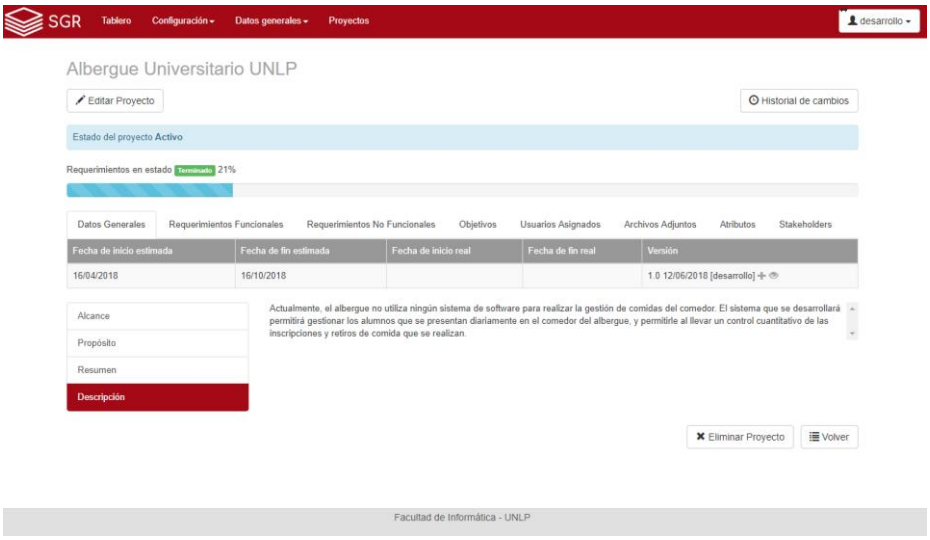


Figura 1. Interfaz de un proyecto.

3.1.2 Administración de Requerimientos

Los requerimientos deben ser creados con el mayor detalle posible, con el fin de generar documentos completos y consistentes. Éstos están compuestos por estado, prioridad y estabilidad. Además se permite asignarles stakeholders, los cuales no pueden interactuar directamente con el sistema, pero son asociados como fuente de información de los requerimientos. También permite realizar el seguimiento de los requerimientos a través de su historial de cambios, administrar versiones, asignar objetivos y adjuntar archivos. Estas características enriquecen su especificación.

3.1.3 Trabajo Colaborativo

Es importante mantener informados a los usuarios acerca de los cambios en la información para favorecer la comunicación y obtener mejores resultados en el desarrollo de un proyecto de software.

Para ello se realizan dos tipos de notificaciones: una en la pantalla inicial del sistema denominada *Tablero de usuario* y otra a través del envío de emails.

En el *Tablero del usuario* se muestra: un listado con los proyectos próximos a concluir, los proyectos retrasados y un listado con los proyectos que fueron actualizados recientemente.

Las notificaciones por email se generan por dos motivos: cambio de estado de un requerimiento por parte de un cliente o por actualización en un proyecto.

Además, SGR permite que los usuarios asignados a un proyecto comenten tanto el proyecto como los requerimientos pertenecientes a este, logrando de esta manera, una comunicación fluida entre ellos.

3.1.4 Generación automática del IEEE Std. 830-1998

En algunas ocasiones es conveniente que los usuarios puedan disponer de los datos sin necesidad de recurrir al sistema. Es por esto que SGR provee la funcionalidad de generar y exportar automáticamente documentos de especificación de requerimientos SRS basado en el estándar IEEE Std. 830-1998. Según IEEE, un buen Documento de Requerimientos, pese a no ser obligatorio que siga estrictamente la organización y el formato dados en el estándar 830, deberá incluir, de una forma o de otra, toda la información presentada en dicho estándar. Este tiene como objetivo ayudar a los clientes a describir claramente lo que se desea obtener mediante un determinado software, ayudar a los desarrolladores a entender qué necesita exactamente el cliente y servir de base para desarrollos de estándares de SRS particulares para cada organización [14].

3.1.5 Roles y Usuarios

La gestión de requerimientos involucra, generalmente, a varias personas que se relacionan con el proyecto de diferentes formas. Debido a esto, el sistema diferencia entre diversos roles de usuarios.

Para brindar privacidad de los datos almacenados, los usuarios que quieran operar con SGR deben autenticarse. Cada usuario autenticado puede realizar diferentes tareas dependiendo de los permisos que tenga asignado dentro del proyecto.

3.2 Caso de estudio

Con el objetivo de evaluar el comportamiento y desempeño de SGR, actualmente se lo está utilizando en un proyecto llevado a cabo en el Instituto de Investigación en Informática - LIDI (III-LIDI). El proyecto surge como respuesta a la solicitud del Albergue Universitario de la UNLP. Esta institución se encarga de dar hospedaje y

alimentación (almuerzo y cena) de forma gratuita a más de 130 jóvenes estudiantes de carreras de grado de la Universidad. La problemática presentada se basó puntualmente en la necesidad de llevar a cabo la gestión de comidas brindadas por el comedor del albergue. Actualmente no utilizan ningún sistema informático para esto, por lo que se propuso desarrollar un sistema de software para la gestión del comedor.

Para llevar a cabo la etapa de elicitación de requerimientos se realizaron entrevistas y con la información obtenida, se procedió a la carga de los datos del proyecto en el sistema.

A partir de este caso de estudio, se concluyó que a través de SGR es posible llevar adelante la etapa de elicitación de requerimientos de una manera ordenada y guiada, como también exportar el documento SRS basado en IEEE Std. 830-1998, mantener la información centralizada y realizar el trabajo de forma colaborativa. Además fue posible fomentar la participación de los clientes permitiendo negociar los requerimientos, logrando de esta manera alinear las necesidades del cliente con el software resultante. Cabe destacar que las funcionalidades de exportar el documento SRS basado en IEEE Std. 830-1998 y el trabajo colaborativo no se encuentran satisfechas de forma simultánea en las herramientas analizadas.

3.3 Tecnología utilizada

Para el desarrollo de SGR se utilizó el framework Symfony [15], el cual se basa en PHP [16], diseñado para optimizar el desarrollo de aplicaciones web. Al ser un sistema web permite que el sistema sea multiplataforma. La decisión de utilizar el framework Symfony se debió a que cuenta con una amplia y completa documentación y además cuenta con una gran comunidad de colaboradores que lo apoyan.

Como sistema de gestión de base de datos se utilizó MYSQL [17] ya que se considera una de las bases de datos de uso libre más utilizada en entornos web. Para realizar la interacción con la base de datos se utilizó Doctrine [18], una librería que permite realizar consultas para modelos orientados a objetos.

Con respecto a la interfaz, se utilizó Twig [19] y Bootstrap [20]. Twig es un motor de creación de plantillas que facilita la implementación de las vistas haciendo el trabajo más simple y ordenado. Bootstrap es un conjunto de herramientas de uso libre pensado para sitios web adaptables.

La figura 2 muestra la arquitectura utilizada para SGR.

Una funcionalidad importante de SGR es el envío automático de notificaciones a los integrantes de un proyecto para mantenerlos actualizados sobre nuevos cambios, para esto se utilizó Swiftmailer [21] debido a que brinda una solución simple y flexible.

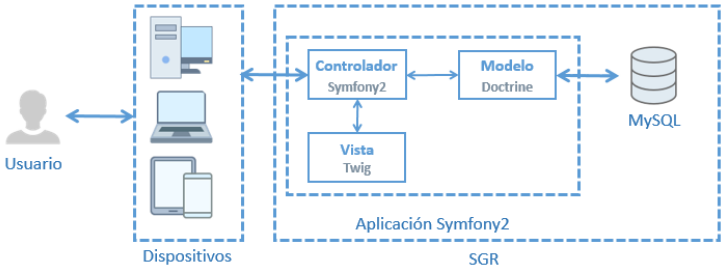


Figura 2. Modelo cliente-servidor de SGR.

3.4 Interfaz de usuario

SGR cuenta con una interfaz de usuario amigable, ya que, tratándose de un sistema para la gestión de requerimientos, se espera que el usuario pueda utilizarla de manera intuitiva y sin grandes dificultades. Al ser una herramienta web se amplía su utilización para cualquier usuario que posea un dispositivo con acceso a un navegador web y conexión a internet. El sistema está desarrollado utilizando diseño web adaptable, para cumplir con los objetivos de una interfaz amigable y de fácil utilización en cualquier dispositivo electrónico como Smartphones, Tablets, Notebook, etc.

Las figuras 3 y 4 muestran la interfaz móvil del detalle de un proyecto y un requerimiento funcional respectivamente.

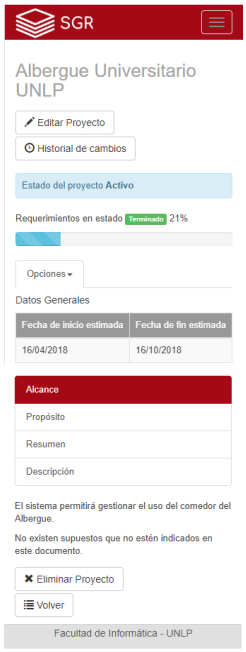


Figura 3. Interfaz de Proyecto.

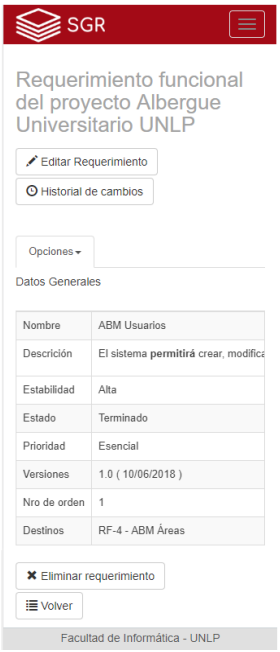


Figura 4. Interfaz de Requerimiento.

3.5 Objetivos alcanzados

SGR fue incluido en la comparación de las herramientas evaluadas en la sección 2 (ver Tabla 1). En consecuencia, se deprende la Tabla 2 la cual muestra la cantidad de características mencionadas en este artículo, que satisface cada herramienta. Se puede observar que SGR cumple con la totalidad de éstas ya que se puso énfasis en las debilidades de las otras herramientas como lo son la generación de documentos basados en el estándar IEEE 830-1998 y el diseño web adaptable, entre otros.

Tabla 2. Resultados obtenidos.

Herramientas	Características Satisfechas/Total
REM	11-18
VISURE	13-18
HELER	12-18
DOORSNG	12-18
LETS REQ!	12-18
ROMMANA	13-18
TRACECLOUD	12-18
PSODA	13-18
TRACKSTUDIO	11-18
REDMINE	13-18
SGR	18-18

4. Conclusiones

Siendo la elicitación de requerimientos una etapa clave en el ciclo de vida del software, encontrar y corregir los defectos en la fase de desarrollo es más costoso que hacerlo en la fase de requerimientos. Por lo tanto, la especificación de requerimientos debe ser clara, completa y estar libre de ambigüedades.

Con el avance de la tecnología informática, la utilización de las herramientas CASE destinadas a la Ingeniería de Software ha tomado gran importancia. Uno de los principales objetivos de estas herramientas es facilitar y mejorar las actividades involucradas en la gestión de requerimientos.

Se realizó un análisis comparativo de un conjunto de herramientas del ámbito académico y comercial. A partir de este análisis se determinó que no existe una herramienta que cumpla con la totalidad de las características mínimas y deseables que debería contener una herramienta CASE para la etapa de gestión de requerimientos. Es por esto que se decidió desarrollar SGR, un sistema que permita: gestionar los requerimientos de manera detallada, trabajar colaborativamente, facilitar la comunicación entre los usuarios del sistema, almacenar toda la información de manera centralizada, brindar seguridad y privacidad de los datos a partir de definición de roles y permisos, permitir realizar trazabilidad de los requerimientos, permitir su utilización desde cualquier plataforma, brindar interfaces amigables y web adaptables.

Se presentaron las funcionalidades más importantes de SGR y a través del proyecto desarrollado para el Albergue Universitario UNLP se pudo observar que a partir de su utilización se llevó adelante la etapa de elicitación de requerimientos de manera ordenada y guiada; se generó el documentos SRS basado en IEEE Std 890-1998 que

están siendo utilizados actualmente por los desarrolladores para la implementación del sistema con resultados y feedback positivos.

5. Trabajos futuros

Se propone agregar nuevas funcionalidades a SGR: diccionario de frases del dominio, notificaciones a través del navegador, calendario con fechas importantes para los proyectos y generación de diversos reportes con el fin de mejorar la experiencia de los usuarios.

Además ampliar las pruebas de campo en asignaturas relacionadas con la Ingeniería de Software de la Facultad de Informática (UNLP) para realizar mejoras o agregar funcionalidades a partir de la retroalimentación de éstas.

6. Bibliografía

1. Pfleeger, Shari Lawrence. Ingeniería de software: teoría y práctica (2002).
2. Bárbara A. Mcdonald Landazuri. Definición de Perfiles en Herramientas de Gestión de Requisitos (2005).
3. IEEE Software Editado por Christof Ebert. Requirements Engineering Tools. Páginas 86-91. Volumen 28 Julio/Agosto (2011).
4. Durán Amador. Un Entorno Metodológico de ingeniería de requisitos para sistemas de información. Tesis doctoral. Universidad de Sevilla (2010).
5. Mauro Callegas Cuervo, Luz Yadira Castillo Estupoñán, Ruby Mónica Fernández Álvarez. HELER: una herramienta para la ingeniería de requisitos automatizada. Págs. 184-200. Colombia: Entramado (2010).
6. Pablo Torrecilla González, María N. Moreno García. Proyecto TCUE Universidades de Castilla y León. (2011).
7. Visure Requirements www.visuresolutions.es (4-6-2018)
8. Doorsng www.doorsng.com (4-6-2018)
9. Rommana https://rommanasoftware.com/ (7-6-2018)
10. Tracecloud https://www.tracecloud.com/ (14-6-2018)
11. Psoda https://www.psoda.com (14-6-2018)
12. Trackstudio http://www.trackstudio.com/ (14-6-2018)
13. Redmine https://www.redmine.org/ (14-6-2018)
14. Especificación de Requisitos según el estándar de IEEE 830, IEEE Std. 830- 1998. (1998).
15. Symfony https://symfony.com (27-6-2018)
16. PHP http://php.net/ (27-6-2018)
17. MySql https://www.mysql.com/ (27-6-2018)
18. Doctrine https://symfony.com/doc/2.8/doctrine.html (27-6-2018)
19. Twig https://twig.symfony.com/ (27-6-2018)
20. Bootstrap http://getbootstrap.com/ (27-6-2018)
21. Swifmailer http://swiftmailer.org/ (27-6-2018)

Hacia una mejora de calidad en Scrum. Integrando documentación requerida por IRAM-ISO 9001:2015.

Julieta Calabrese [1], Silvia Esponda [2], Marcos Boracchia [3], Patricia Pesado [4]

Instituto de Investigación en Informática LIDI (III-LIDI) *
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 - La Plata - Buenos Aires
*Centro Asociado Comisión de Investigaciones Científicas de la Pcia. de Bs. As. (CIC)

{jcalabrese, sesponda, marcosb, ppesado}@lidi.info.unlp.edu.ar

Abstract. En el camino de obtener una buena gestión de la calidad de procesos en las pequeñas y medianas empresas desarrolladoras de software, se evidencia la falta de documentación en metodologías muy usadas en la actualidad (como Scrum). Se propone compatibilizar la documentación de Scrum y los estándares IEEE recomendados en las etapas del proceso de desarrollo con la documentación requerida por la norma, definiendo un único documento.

Keywords: Scrum – Calidad – IRAM-ISO 9001 – Documentación

1. Introducción

En la actualidad, la gestión de la calidad es una parte fundamental dentro de cualquier organización ya que dicha gestión puede llegar a ser una ventaja competitiva que fortalezca a la organización en el momento de prestar un mejor servicio o tener un producto que cumpla las exigencias del cliente. Puntualmente, la industria del software es una de las industrias de más rápido crecimiento en las últimas décadas, y aunque la producción de software tiene a su disposición un conjunto de normas específicas destinadas a evaluar diferentes aspectos del proceso y/o producto, muy a menudo es necesario establecer un Sistema de Gestión de la Calidad (SGC) de acuerdo a la norma IRAM-ISO 9001.

La norma ISO 9001:2015 “Sistemas de gestión de la calidad. Requisitos” [1] está basada en los retos a los que se enfrentan las empresas de cualquier tamaño y sector hoy en día. Además, se centra en la eficacia del sistema de gestión para dar cumplimiento a los requisitos del cliente. La norma promueve la adopción de un enfoque basado en procesos y **exige cierta información documentada que la organización determine como necesaria para la eficacia del SGC**.

Por otro lado, las metodologías ágiles representan una alternativa muy usada en la actualidad para el proceso de desarrollo de sistemas de software, haciendo hincapié en la relación con el cliente y en el desarrollo incremental del producto. Estas metodologías ofrecen entregas frecuentes de software funcional, permiten cambios de requisitos y poseen participación

[1] ORCID: 0000-0003-4845-0556
[2] ORCID: 0000-0002-1986-4867
[3] ORCID: 0000-0002-4568-3883
[4] ORCID: 0000-0003-0000-3482

directa del cliente durante toda la etapa de desarrollo. Una de las metodologías más utilizadas en la actualidad es Scrum [2], la cual se define como un proceso iterativo incremental y empírico para administrar y controlar el trabajo de desarrollo, y una de las más utilizadas en PyMEs desarrolladoras de software.

En el camino de obtener una buena gestión de la calidad de procesos en las pequeñas y medianas empresas desarrolladoras de software, surge la necesidad de compatibilizar la documentación utilizada por la metodología y los estándares recomendados para cada etapa del proceso de desarrollo, con la documentación requerida por la norma.
La documentación definida por Scrum para sus desarrollos es insuficiente para satisfacer los requisitos del estándar IRAM-ISO 9001:2015, por lo cual es necesario definir documentación adicional basada en estándares, presentado como un nuevo documento único.

En la siguiente sección se describe en profundidad la estructura de la norma IRAM-ISO 9001:2015 junto con una breve descripción de las directrices brindadas por la norma IRAM-ISO/IEC 90003:2006. En la tercera sección se presenta Scrum, la metodología ágil de desarrollo estudiada, indicando las falencias que presenta en lo que a documentación respecta. En la cuarta sección se describe el documento integrador propuesto, con el fin de satisfacer la documentación requerida por la norma IRAM-ISO 9001:2015. En este trabajo se tendrán en cuenta puntualmente las secciones 6, 7 y 8 de la norma, las cuales están enfocadas en el proceso a certificar. Por último, se exponen las conclusiones y trabajos futuros.

2. IRAM-ISO 9001:2015

2.1 Iniciar el camino hacia la Gestión de Calidad

Una organización que se propone aplicar e implementar un SGC bajo la norma ISO 9001 se enfrenta con diferentes barreras. La visualización del punto de partida es crítica para un alto porcentaje de estas, sobre todo si partimos de la base de que no cuentan con personal capacitado para su ejecución y los costos relacionados a consultorías.

La documentación es un aspecto sumamente importante, tanto en el desarrollo de la aplicación como en el mantenimiento. Constituye un elemento indispensable para que los usuarios puedan obtener el máximo rendimiento de las prestaciones que ofrece el producto, como así también para apoyar a los stakeholders a la hora de entender sus características y funciones. Una buena documentación debe ser específica, concisa, y relevante.

2.2 IRAM-ISO 9001:2015

En 2015 se publicó una nueva versión de ISO 9001 denominada “ISO 9001:2015 ‘Sistemas de gestión de la calidad. Requisitos’”, la cual posee una reestructuración basada en la IRAM-ISO 9001:2008. La revisión surgió por la necesidad de adaptar la norma a los tiempos actuales de las organizaciones. Posee por la siguiente estructura:

1. Alcance
2. Referencias Normativas
3. Términos y Definiciones
4. Contexto de la Organización

5. Liderazgo
6. Planificación
7. Soporte
8. Operación
9. Evaluación del Desempeño
10. Mejora

Con esta nueva estructura, algunos requisitos se han visto modificados, han sido eliminados o se han añadido como nuevos. Por ejemplo, la gestión de documentos y registros, que hasta el momento se localizaban en el apartado 4.2, pasa a ser el apartado 7.5 y se adquiere el nuevo término de **información documentada**.

En la nueva versión, el enfoque basado en procesos se centra en el incremento de la gestión y control de las interacciones existentes entre procesos y jerarquías funcionales de cada organización. Además, se ha trabajado para que el lenguaje empleado sea simple y comprensible por cualquier persona. Ya no sólo se habla de clientes, sino que en ella se determinan los requisitos necesarios para cumplir con las exigencias de las partes interesadas.

En ISO 9001:2015 existe un apartado que hace hincapié en el **"**Contexto de la organización". Este punto trata lo imperativo de tener en cuenta y analizar el entorno socioeconómico de la empresa y los vínculos existentes con los stakeholders (internos y externos). Este análisis facilita la identificación de problemas y necesidades que puedan generar un impacto en la planificación del Sistema de Gestión de la Calidad.

Este es uno de los cambios más importantes, ya que el concepto es transversal a la totalidad del sistema de gestión, es decir, influye en el establecimiento, implantación, mantenimiento y mejora.

Los riesgos son un punto crítico que se ha agregado a la norma, recomendando su gestión a lo largo de todo el proceso. Para administrar los riesgos se puede adoptar la metodología que más se ajuste a las necesidades de la empresa.

2.3 IRAM-ISO/IEC 90003:2006

Actualmente, existen un conjunto de directrices que proporcionan una guía para las organizaciones en la aplicación de ISO 9001 a la hora de adquirir, proveer, desarrollar, operar y mantener software y servicios de soporte relacionados. Estas directrices conforman la norma IRAM-ISO/IEC 90003 [3], cuya última versión publicada es del año 2006 basada en la IRAM-ISO 9001:2000. La norma IRAM-ISO/IEC 90003:2006 es utilizada como guía para certificaciones de procesos de software bajo la IRAM-ISO 9001:2008.

Al haberse realizado una actualización de la norma IRAM-ISO 9001 en el año 2015, se produjo la necesidad de adaptar las directrices brindadas por IRAM-ISO/IEC 90003:2006 a la nueva versión de la norma. Bajo este contexto, "IRAM-ISO/IEC 90003 Ingeniería de software – Guía para la aplicación de la ISO 9001:2015 al software de computadora" se encuentra en la etapa DIS (Proyecto de Norma Internacional), a la espera de ser aprobada y publicada.

3. Proceso de desarrollo de software bajo IRAM-ISO 9001:2015

Las versiones anteriores de IRAM-ISO 9001 requerían un manual de la calidad. La versión actual del estándar ya no estipula esto. Desaparece la obligación específica de tener que desarrollar manuales, procedimientos documentados exigibles en las auditorías de certificación, instrucciones de trabajo, registros, etc. Sin embargo, requiere que las organizaciones determinen el nivel de información documentada que es **necesaria** para controlar el SGC. Además, se hace hincapié en controlar su acceso, lo que refleja la importancia de la seguridad de la información.

A la hora de iniciar el camino a la certificación de la norma, una gran cantidad de procesos de desarrollo de software cuentan con **escasa documentación, incapaz de satisfacer la requerida por la norma**. Como se mencionó anteriormente, se hará hincapié en el proceso de desarrollo ágil Scrum.

Scrum

Scrum [6] es un marco de trabajo ágil para desarrollo de software que tiene entre sus objetivos el control continuo sobre el software, en el cual el cliente establece las prioridades y el equipo Scrum se autoorganiza para determinar la mejor forma de entregar resultados. El trabajo se organiza en ciclos llamados *sprints* que son iteraciones de corta duración, típicamente de 2 a 4 semanas. Durante cada sprint, el equipo selecciona un conjunto de requisitos de una lista priorizada, de manera que las funciones desarrolladas al principio del proyecto son las de más alto valor. Al final de cada sprint se entrega un producto de software ejecutable en el ambiente requerido por el cliente.

El enfoque Scrum propone software funcional por sobre la excesiva documentación, utilizando una *pila del producto* la cual es el corazón de Scrum. La pila conforma el listado donde se detallan y priorizan cada uno de los requisitos, y está en constante evolución. Es abierta a todos los roles, pero es el propietario del producto el responsable de dar la prioridad a los requisitos y es quien decide sobre ésta.

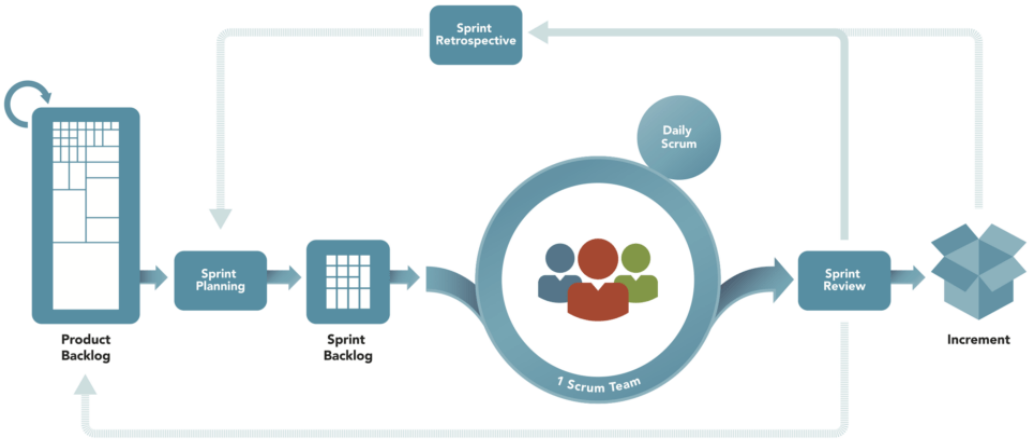


Figura 1. Metodología Scrum (figura extraída de scrum.org)

4. Integración

En base a la escasa documentación de los procesos de desarrollo de software más usados en la actualidad (puntualmente Scrum [7]), la cual por sí sola no satisface la exigida por la norma, se presenta un único documento que integra un conjunto de ítems basado en estándares conocidos. La integración busca satisfacer la documentación necesaria (en la etapa de análisis de requisitos y planificación) definida en los requisitos del estándar IRAM-ISO 9001:2015 a la hora de crear un producto de software utilizando la metodología ágil de desarrollo Scrum.

Primera parte: Documentación existente

- **Documentación basada en estándares**

  Existe un conjunto de documentos destinados a especificar diferentes atributos funcionales y no funcionales del producto a desarrollar, los cuales se encuentran definidos por el estándar IEEE para las etapas de especificación de requisitos y planificación.

Documento	Definido por
Especificación de Requisitos de Software (SRS)	IEEE-Std830 [4]
Plan de Gestión de Proyecto (PGP)	IEEE-Std1058 [5]

Tabla 1. Documentación basada en estándares

Especificación de Requisitos de Software (SRS)

La ingeniería de requisitos es un proceso especializado que se realiza en un dominio para documentar las características que debe cumplir un producto software y transformarlas en una especificación. El documento de especificación de requisitos debe cubrir la representación y comprensión del ambiente específico y debe contener todas las funciones esenciales (funcionalidad delimitada) del software de manera rastreable, no ambigua, con independencia de los requisitos no funcionales y de las restricciones de diseño.

IRAM-ISO 9001:2015, en su apartado 7.5, establece que debe quedar documentada la información que la organización determina como necesaria para la eficacia del Sistema de Gestión de la Calidad.

En diferentes organizaciones, este documento se genera con diversas estructuras, aunque existen estándares internacionales que especifican su estructura, tales como IEEE Std 830-1998. Dicho estándar es una guía de lo que debería contener este documento, dando pautas generales de contenido para cada sección. Las secciones principales del documento se presentan en la *Tabla 2*:

1. Introducción
2. Descripción General
3. Requisitos Específicos
4. Apéndices

Tabla 2. Especificación de Requisitos de Software

Plan de Gestión de Proyecto (PGP)

En el Plan de Gestión del Proyecto se define la ejecución, supervisión, y cierre del proyecto. Se documentan las salidas de los procesos de planificación. El estándar correspondiente no especifica las técnicas exactas que pueden ser usadas en la creación de los planes de proyectos ni ofrece ejemplos, sino que cada organización que utilice este estándar debería desarrollar un conjunto de prácticas y procedimientos para proporcionar una guía detallada para la preparación y actualización de los planes de gestión de los proyectos de software. Estas prácticas y procedimientos deberían tener en cuenta los factores del entorno, organizacionales y políticos que pueden influenciar en la aplicación de este estándar. Las secciones principales del documento se presentan en la *Tabla 3*:

1. Visión General del Proyecto
2. Referencias
3. Definiciones
4. Organización del Proyecto
5. Planes de gestión del proceso
6. Planes de Proceso Técnicos
7. Planes de Procesos de Soporte
8. Planes Adicionales

Tabla 3. Plan de Gestión de Proyecto

- **Documentación de Scrum**

Por su parte, la metodología ágil Scrum utiliza para su desarrollo, a modo de documentación, la denominada *Pila de Producto* la cual incluye una serie de elementos (principalmente funcionalidades nuevas del cliente) pero también objetivos de mejora de ingeniería, trabajo de investigación o exploratorio, y posiblemente fallos conocidos. Adicionalmente, Scrum utiliza una *Pila de Sprint* la cual se compone de las funcionalidades a desarrollar para un sprint específico. Dicha pila no será de interés para este trabajo.

Una práctica muy conocida es articular los requisitos utilizando **historias de usuario**, las cuales brindan descripciones claras y concisas de la funcionalidad en términos del valor que aporta al usuario final del producto. Una historia de usuario es redactada con un formato específico: “Como ROL quiero FUNCIONALIDAD para poder BENEFICIO”.

La siguiente tabla define los documentos utilizados en la metodología:

Documento	Definido por
Pila de Producto	Scrum
Pila de Sprint	Scrum

Tabla 4. Documentación de Scrum

Segunda parte: Documentación existente vs. IRAM-ISO 9001:2015

La pila de producto utilizada por Scrum es insuficiente para lograr los requisitos de documentación de la IRAM-ISO 9001:2015. Con el fin de satisfacer dicha documentación, se analizó la manera de integrar ítems de los estándares utilizados en la actualidad a modo de complementar lo solicitado en la norma. Cabe destacar que para el análisis se tuvieron en cuenta las secciones 6, 7 y 8 de la norma, las cuales hacen referencia puntualmente al proceso. Adicionalmente, se tuvo presente a la norma IRAM-ISO/IEC 90003:2006, que si bien no es utilizada para certificar un proceso de desarrollo bajo IRAM-ISO 9001:2015, pudo ser tomada de referencia a modo de guía en algunas secciones.

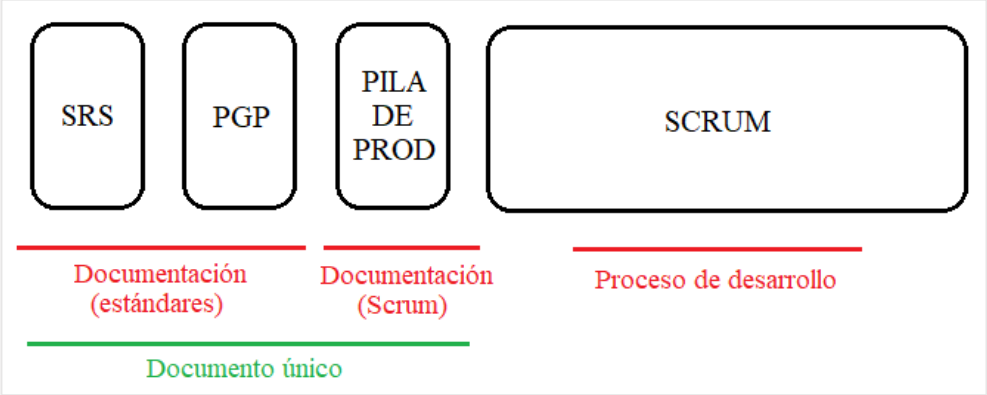


Figura 2. Generación de Documento único

El documento único representado en la *Figura 2* está conformado por ítems de la documentación existente a fin de satisfacer las secciones de la norma que hacen hincapié en el proceso de desarrollo.

La siguiente tabla muestra los ítems de la IRAM-ISO 9001:2015 en donde se requiere documentación específica, junto con los ítems de la documentación estudiada que satisfacen lo pedido.

IRAM-ISO 9001:2015	Documentación sugerida
6.1. Acciones para abordar riesgos y oportunidades.	PGP (5.4): Plan de administración de riesgos.
6.2. Objetivos de la calidad y planificación para lograrlos.	PGP (5.3.4): Plan de control de calidad.
7.1.5. Recursos de seguimiento y medición.	PGP (7.2): Plan de Pruebas.
7.2. Competencia.	PGP (4.3): Roles y Responsabilidades.
8.2.3. Revisión de los requisitos para los productos y servicios.	(Scrum) Pila de producto - Acta de Reunión. (SRS) Sección 3: Requisitos Específicos.
8.3.2. Planificación del diseño y desarrollo.	PGP (5.2): Plan de trabajo.
8.3.3. Entradas para el diseño y desarrollo.	(Scrum) Pila de producto.

	PGP (5.1.2): Plan de adquisición de recursos.
8.3.4. Controles del diseño y desarrollo.	PGP (5.3): Plan de Control.
8.3.5. Salidas del diseño y desarrollo.	
8.4.1. Generalidades (Control de los procesos, productos y servicios suministrados externamente)	PGP (5.1.2): Plan de adquisición de recursos.
8.5.6. Control de los cambios.	PGP (7): Planes de procesos de apoyo.
8.6. Liberación de los productos y servicios.	PGP (5.5): Plan de liberación de proyecto.

Tabla 5. Integración

Tercera parte: Documento único para secciones 6, 7 y 8 de IRAM-ISO 9001:2015

Con el fin de satisfacer la documentación necesaria en las secciones mencionadas anteriormente de la norma IRAM-ISO 9001:2015, se crea un único documento teniendo como base lo presentado en la *Tabla 5*.

1. Gestión de Riesgos
2. Control de Calidad
3. Roles y Responsabilidades
4. Historias de Usuario (Pila de Producto) – Requisitos específicos
5. Plan de Trabajo
   5.1. Actividades
   5.2. Asignación de esfuerzo
6. Plan de Control
   6.1. Control de requisitos
   6.2. Control de calendario
   6.3. Recolección de métricas
7. Adquisición de recursos
8. Planes de Procesos de Apoyo
   8.1. Verificación y Validación (Pruebas)
   8.2. Documentación
   8.3. Revisiones y Auditorías
   8.4. Resolución de Problemas
   8.5. Administración de Terceros
   8.6. Mejora de Procesos
9. Plan de Liberación del Proyecto

Tabla 6. Documento único

A continuación, se describen los ítems del documento:

1. Gestión de Riesgos
Se detalla cada riesgo que pueda presentarse, indicando su estrategia de mitigación y su plan de contingencia. Se sugiere indicar si es un riesgo que afecta al proceso, al producto y/o al negocio, indicando además la probabilidad de ocurrencia y el grado de impacto.

2. Control de Calidad

Se detallan los objetivos de la calidad, detallando las acciones para lograrlos y su forma de control si se cumplieron con éxito.

3. Roles y Responsabilidades
Se detallan los roles y responsabilidades que se asignan a los individuos responsables para cada función del proyecto.

4. Historias de Usuario – Requisitos Específicos
Se especifican todos los requisitos del sistema en formato de historias de usuario. Debe contener una lista detallada y completa de los requisitos que debe cumplir el sistema a desarrollar. El nivel de detalle de estos debe ser el suficiente para que el equipo de desarrollo pueda diseñar un sistema que los satisfaga, y los encargados de las pruebas puedan determinar si éstos se satisfacen. Deben incluirse requisitos no funcionales tales como seguridad, portabilidad, fiabilidad, disponibilidad, rendimiento y mantenibilidad.

5. Plan de Trabajo
5.1 Actividades
Se especifican las actividades, detallando cada tarea, si es apropiado.
5.2 Asignación de esfuerzo
Se especifican los recursos con los que se cuenta, asignando a cada uno el rol que le corresponde expresando el esfuerzo en horas para cada actividad. Se puede incorporar un diagrama de PERT o Camino Critico.

6. Plan de Control
6.1. Control de requisitos
Se definen Actas de Reunión en cada reunión de retrospectiva del Sprint, detallando los resultados de los Casos de Prueba de cada historia de usuario de la pila de producto.
6.2. Control de calendario
En el Acta de Reunión debe quedar plasmado el cumplimiento de los tiempos estimados para cada Sprint.
6.3. Recolección de métricas
Se presentan informes del Product Owner en cada retrospectiva del Sprint, en los cuales se detalle el estado de todo el proyecto (cantidad de historias aprobadas en base a las planificadas, cantidad de historias pendientes a realizar, porcentaje de conformidad/no conformidad del sprint, entre otros).

7. Adquisición de recursos
En el caso de contar con recursos externos para llevar adelante el proyecto (hardware, software, contratos de servicio y servicios administrativos, etc.) se establece una tabla control detallando cada uno, forma de adquisición, fecha de cumplimiento, entre otros.

8. Planes de Procesos de Apoyo
8.1. Verificación y Validación (Pruebas)
Se especifican los casos de pruebas necesarios en cada una de las historias de usuario de manera de evidenciar su seguimiento.
8.2. Documentación
Se describen todos los documentos externos que contribuyan al proyecto.
8.3. Revisiones y Auditorías
Se describe la planificación de las revisiones y auditorías, en el caso de aspirar a una

certificación.

8.4. Resolución de Problemas

El Scrum Master es el encargado de lidiar con la resolución de problemas. En el caso de ocurrencia, se describen los pasos a seguir para la resolución de los problemas que ocurran a lo largo del desarrollo.

8.5. Administración de terceros

Se describe cómo se seleccionarán y administrarán las empresas contratadas en caso de que algún producto del proyecto sea realizado por otra empresa.

8.6. Mejora de Procesos

Se definen en las Actas de Reunión de cada Sprint, los resultados de los Casos de Prueba de cada Historia de la Pila y sus posibles cambios.

9. Plan de Liberación del Proyecto

Se presentan informes del Product Owner al finalizar el proyecto, en los cuales se describa la no conformidad y que acciones se deciden llevar a cabo en el caso de ocurrencia.

5. Conclusiones

Se presentó una descripción de Scrum y de la norma IRAM-ISO 9001:2015 evidenciando las falencias de documentación exigidas por la norma por parte de la metodología ágil, lo que dificulta el trabajo a la hora de aspirar a una certificación IRAM-ISO 9001.
A priori, Scrum no satisfaría los requisitos de la norma, por lo que se diseñó un único documento compatibilizando la exigua documentación de Scrum y los estándares IEEE recomendados en las etapas del proceso de desarrollo, con la documentación requerida por la norma.

Dicho prototipo ha sido aplicado en la cátedra de Ingeniería de Software II de la Facultad de Informática, donde se desarrolla un proyecto bajo la metodología Scrum.

6. Bibliografía

[1] IRAM-ISO 9001:2015 "Sistemas de gestión de la calidad. Requisitos".

[2] "Q-Scrum: una fusión de Scrum y el estándar ISO/IEC 29110". http://sedici.unlp.edu.ar/handle/10915/32421

[3] IRAM-ISO/IEC 90003:2006 "Directrices para la aplicación de la norma IRAM-ISO 9001:2000 al software".

[4] IEEE Std 830-1998. "Especificaciones de los requisitos del Software".

[5] IEEE Std 1058-1998. "Plan de Gestión de Proyecto".

[6] "La guía de Scrum". https://www.scrumguides.org/docs/scrumguide/v1/scrum-guide-es.pdf

[7] "Documentación de Scrum". http://metodologiascrum.readthedocs.io/en/latest/Scrum.html

Workshop Arquitectura, Redes y Sistemas Operativos

Coordinadores

Jorge Ardenghi (UNS) - Carlos Buckle (UNPSJB) - Orlando Micolini (UNC) - Marcelo Arroyo (UNRC)

Temas

Administración y monitoreo de redes y servicios
Análisis y detección de vulnerabilidades
Arquitecturas de computadoras
Arquitecturas orientadas a servicios (SOA)
Calidad de Servicio (QoS, SLAs)
Comunicaciones unificadas
Criptografía
Infraestructura para firma digital y certificados digitales
Infraestructura para Grid
Middleware
Redes activas
Redes de comunicaciones
Redes de nueva generación
Redes heterogéneas
Redes inalámbricas
Redes móviles
Seguridad Informática y autenticación, privacidad
Servicios de integración (Web Services o Net)
Sistemas embebidos
Sistemas operativos
Sistemas P2P

Hub Of Things: Concentrador para Internet de las Cosas

Ricardo Brea[1], Daniel Skrie[1], Marisa Panizzi[1], Rodolfo Bertone[2]

[1] Escuela de Sistemas. Universidad Argentina John F. Kennedy.
Bartolomé Mitre 1411, Ciudad Autónoma de Buenos Aires (C1037ABA), Argentina
[2] Instituto de Investigaciones en Informática LIDI. Facultad de Informática. UNLP – CIC
brea.ricardo@gmail.com; dskrie@yahoo.com.ar; marisapanizzi@outlook.com;
rbertone@lidi.unlp.edu.ar

Resumen. Desde hace algunos años se observa un notorio incremento en la cantidad de dispositivos electrónicos conectados a internet, monitoreados y controlados en forma remota. La diversidad tecnológica, por un lado, y la cantidad de dispositivos, por otro, dificulta su integración o, al menos, una integración ordenada. Además, las plataformas basadas en microcontroladores o de procesamiento reducido, como Arduino, no brindan una conexión con niveles aceptables de seguridad. La privacidad constituye otra dificultad, ya que los usuarios desconocen si los distintos proveedores de soluciones IoT (Internet of Things) utilizan sus datos o los venden a terceros. La latencia también es otro factor determinante al diseñar una solución IoT, muchos de los proveedores de soluciones en la nube no tienen servidores locales, lo que degrada el tiempo de reacción ante determinado evento. El Hub Of Things (*HoT)* es una solución que permite integrar localmente los dispositivos *IoT*, brindando una alternativa para solucionar los problemas de diversidad de tecnologías, con una interface homogénea y segura. Facilitando la mediación entre los dispositivos y el usuario.

Palabras Clave: Internet de las Cosas, Fog computing, Seguridad, MQTT.

1 Introducción

El término “Internet de las Cosas” o IoT por sus siglas en inglés, es un concepto acuñado en 1999 por Kevin Ashton investigador de MIT [1]. Originalmente los datos disponibles en internet eran agregados o generados por humanos. Con la incorporación de sensores y conectividad más accesible, los dispositivos o “cosas” son capaces de generar y almacenar datos. Los dispositivos pueden ser diversos teléfonos inteligentes, bandas inteligentes (*SmartBands*), estaciones meteorológicas, automóviles, etc.

El IoT genera grandes posibilidades de crecimiento no solo en el ámbito hogareño sino también en el ámbito industrial, con lo que se denomina Industria 4.0; por ejemplo en la planificación urbana con el concepto de ciudad inteligente o Smart Cities.

En la Tabla 1, se representan las distintas proyecciones de crecimiento del IoT, son diversas, pero promedian 25,7 billones de dispositivos para 2020.

Tabla 1. Pronósticos de crecimiento del IoT

Fuente	Billones de dispositivos
Cisco [2]	**26.3**
Ericsson [3]	**28**
Gartner [4]	**20.8**
Goldman Sachs [5]	**28**

Con un pronóstico de crecimiento tan optimista, es importante detectar que problemas tiene el IoT en la actualidad con el propósito de mejorarlos. *Acquity Group* (División de *Accenture* dedicada a la estrategia digital y marketing) realizó en el año 2014 una encuesta a 2000 consumidores de los Estados Unidos, que conociendo los dispositivos *IoT*, no los adquirirían por los motivos que se presentan en la Figura 1.



Figura 1. Motivos de Rechazo del IoT

La falta de valor agregado es una falla que se puede deber a varios motivos, uno de esos motivos es la difícil integración con soluciones similares o existentes.

La preocupación por la privacidad es un elemento inherente a las soluciones en la nube. Los proveedores de servicios en la nube mitigan este problema con un contrato de privacidad.

La preocupación por los costos se presenta junto con varios elementos, desconocimiento de las tecnologías e incertidumbre en los costos de servicios necesarios para implementar las soluciones y que forman parte del costo fijo.

Hay aspectos técnicos que impiden la implementación del IoT, como la latencia o el tiempo de respuesta de la solución. Cuando se opta por tener una solución en la nube hay numerosos factores que inciden en la latencia desde que el dato es generado en el dispositivo hasta que se trasmite al proveedor del servicio. Estos factores pueden ser la calidad de la conexión a internet o la distancia entre el dispositivo y el centro de datos del proveedor. Para reducir la latencia inherente de las soluciones en la nube surge el concepto de *Fog Computing* [1] o *Edge Computing* que se refiere a:

- Analizar los datos más sensitivos en el *Edge*, el borde; donde se generan los datos, sin necesidad de enviar grandes volúmenes de datos a la red.
- Reaccionar a la información generada en el rango de milisegundos.
- Enviar la información a la nube para su análisis y almacenamiento a largo plazo.

2 Desarrollo

El objetivo del ***Hub Of Things*** es proveer una plataforma segura, homogénea, extensible y abierta, que opera en el *Fog Computing* y puede ser utilizada de base en nuevas soluciones *IoT* para el control y monitoreo de dispositivos conectados a internet. El ***Hub Of Things*** plantea una solución de bajo costo a los motivos de rechazo al IoT.

En su esencia el ***Hub Of Things*** es un IoT *Gateway*, una solución de software y hardware que sirve de intermediario entre los dispositivos y otros sistemas, ya sea que estén implementados localmente (*On-Premise*) o en la Nube. HoT incrementa la seguridad, la estabilidad y reduce la latencia Además, al brindar una capa de abstracción reduce el acoplamiento con servicios *IoT* en la nube y permite adaptar otros protocolos de comunicación como *LoRa*, *BLE*, *Modbus, etc.* a un único protocolo, en este caso *MQTT*. En la Figura 2**,** se presenta la arquitectura interna del sistema y la relación entre sus componentes.

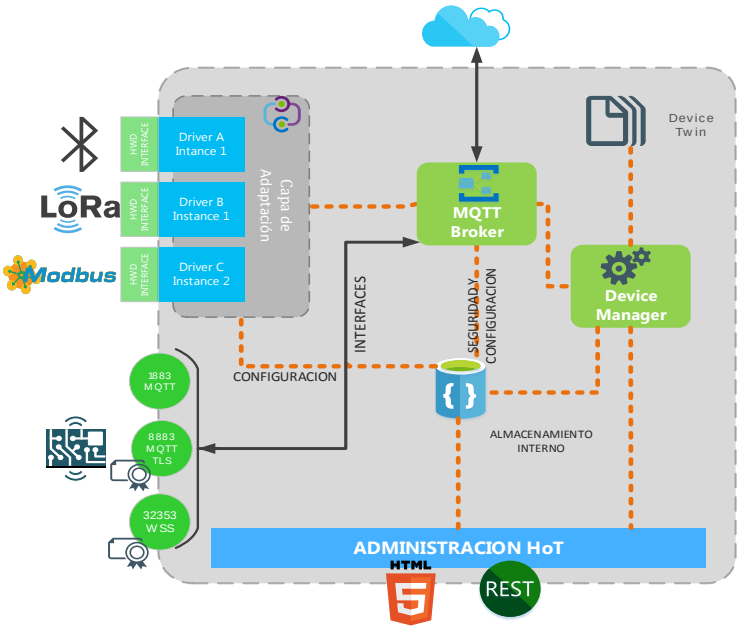


Figura 2. Arquitectura interna del HoT

- **Capa de adaptación**: el objetivo de esta capa es proveer una interface de programación, ***API***, que permita adaptar cualquier protocolo al sistema de mensajes similar a ***MQTT***, es decir en tuplas *<<TOPIC, PAYLOAD>>*.
- **MQTT Broker**: es el componente central de la solución permitiendo conectar dispositivos y clientes en forma homogénea. Implementa el protocolo MQTT sobre TCP y *WebSockets* lo que facilita la comunicación con clientes JavaScript.
- **Device Manager**: el objetivo del *Device Manager* es mantener el estado de los dispositivos utilizando el concepto de Gemelo Digital o *Digital Twins*.
- **API de dispositivos:** el HoT provee una API REST que permite administrar la definición de los dispositivos.
- **Interface HTML**: la capa de presentación se encuentra programada en *HTML5* y *JavaScript* utilizando el *framework Angular* para aplicaciones de una sola página o "*SPA*". El diseño de la página seguirá la directiva *Material Desing* de Google [2], que se asemeja a los controles del sistema operativo *Android*. Dichas directivas existen para diseñar interfaces de usuario intuitivas, homogéneas y fáciles de utilizar. Otras características de la interface HTML son; Interface *Responsive*, adaptable a varios tipos de dispositivos; Gráficos SVG, vectoriales que consumen menos espacio y son visibles en cualquier resolución [3].

2.1 MQTT.

(*Message Queue Telemetry Transport*) o MQTT un protocolo usado para la comunicación *machine-to-machine* (*M2M*) en *IoT*. Este protocolo está orientado a la comunicación de sensores y esta optimizado para no agregarle mucha información adicional a cada mensaje que envía y puede ser utilizado en la mayoría de los dispositivos embebidos y de pocos recursos. *MQTT* sigue una topología de estrella, con un nodo central que hace de servidor o *broker* y distribuye los mensajes utilizando el patrón *publiser-subscriber* [4].

Cada mensaje MQTT se compone de un tópico o *topic* y de una carga o *payload*. El tópico identifica el origen de los datos y la carga almacena el o los valores. El formato del tópico es de niveles separados por el carácter *'/'*. En la Figura 3**,** se muestra la interacción de los distintos actores dentro de la arquitectura del protocolo MQTT.

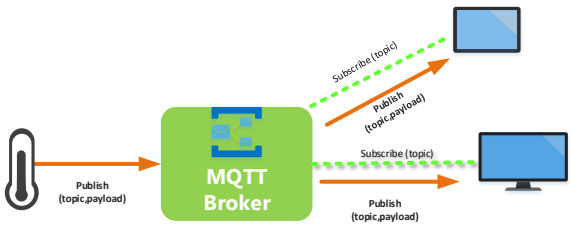


Figura 3. Arquitectura MQTT

2.2 Organización de los dispositivos

El HoT provee una estructura para organizar los dispositivos a fin de complementar al protocolo MQTT. El sistema almacena una definición por cada dispositivo conectado a fin de aplicar el concepto de *Digital Twin*, almacenando metadatos y estado actual del dispositivo. Este concepto es reciente y varias plataformas lo denominan de diferente manera. *AWS* lo denomina *Device Shadow*, *Azure* lo denomina *Device Twin*.

HoT utiliza la naturaleza jerárquica de los tópicos de MQTT para agrupar los dispositivos y sus sensores.

La definición de un dispositivo en HoT tiene los siguientes elementos:

- Información esencial:
  - Identificador de dispositivo: identificador único de dispositivo, dicho identificador se utiliza para autenticar los dispositivos en el sistema.
  - Secreto: se utiliza para autenticar al dispositivo
- Metadatos: información complementaria del dispositivo utilizada para su visualización como puede ser nombre, ícono y ubicación.
- Estado: utilizado por el *Device Manager* para monitorear los dispositivos:
  - Tópico testigo: tópico de MQTT al cual el *Device Manager* se subscribe para actualizar los demás atributos del estado
  - Fecha hora de último reporte
  - Tiempo máximo de inactividad
  - Activo
- Atributos: valores enviados desde los sensores conectados al dispositivo, dichos valores también pueden tener una estructura jerárquica.
- Comandos: valores enviados al dispositivo

2.3 Seguridad

La seguridad es un elemento fundamental del sistema y cubre los siguientes aspectos básicos:

- Recursos: se denomina recurso a los elementos resguardados por la seguridad. En este caso las interfaces REST, HTML y MQTT
- Autenticación: identificar los usuarios que se conectan al sistema
- Autorización: nivel de acceso al sistema de un usuario autenticado
- Intercambio seguro de datos: el sistema encripta los datos transmitidos desde y hacia los usuarios.

Autenticación: cualquier usuario que desee conectarse a alguna de las interfaces debe estar autenticado. Las contraseñas se validan contra la colección de usuarios.

Los dispositivos que se conecten mediante alguna de las interfaces MQTT también deben especificar credenciales. Dichas credenciales son el id de dispositivo y su secreto.

Autorización: la autorización se realiza utilizando roles en el caso de las interfaces HTML Y REST y una lista de control de acceso o ACL para la interface MQTT.

El sistema permite definir usuarios con un nivel de autorización *Administrador* o V*isualizador* para las interfaces HTML y REST. El nivel de acceso a la interface MQTT acceso utilizando una lista de control de acceso o ACL (*Access Control List*).

Cada registro de la ACL tendrá como atributos:

- Usuario: identificador de usuario
- Tópico: cadena representado el tópico al cual el usuario puede acceder, se puede hacer uso de los comodines admitidos por MQTT**.**
- Tipo de acceso: LECTURA | ESCRITURA | COMPLETO

En el caso de los dispositivos la autorización, se realiza de una forma más simple, sin necesidad de una ACL. Cada dispositivo puede publicar y subscribirse a cualquier tópico cuyo prefijo coincida con la definición del dispositivo.

Intercambio de datos: El sistema utiliza la tecnología TLS*, **T**ransport **L**ayer **S**ecurity*, seguridad en la capa de transporte, que consiste en encriptar los datos de la conexión segura por medio de un certificado emitido para tal fin. El intercambio de datos las interfaces (MQTT, HTML y REST) se encuentran asegurados por este método.

3 Caso de validación

A fin de comprobar el funcionamiento de la solución se han desarrollado dos dispositivos que son integrados a través del HoT. En esta implementación se abordan temas fundamentales de IoT como es la persistencia de los datos a largo plazo, la visualización de los datos almacenados y la arquitectura de implementación.

La implementación tiene tres elementos a destacar:

1. **Recolección de datos ambientales:** su función es la medición de variables ambientales de una planta (temperatura ambiente, humedad relativa del ambiente, intensidad lumínica y humedad de suelo).
2. **Cámara NDVI**: desarrollado sobre la plataforma Raspberry Pi Zero W y utiliza análisis de imagen para calcular el índice de vegetación diferencial normalizado o NDVI [5].
3. **HoT Host:** computadora que tendrá instalado todos los elementos de software correspondientes al HoT. En la Tabla 2, se presenta el detalle del hardware utilizado incluyendo los sensores conectados.

La solución se implementa utilizando la arquitectura de microservicios por la facilidad de actualización de cada uno de sus componentes; y con leves modificaciones puede ser implementado en diversas plataformas y sistemas operativos. *Docker* es una plataforma abierta para micro servicios elegida para esta implementación. Cabe destacar que la solución puede ser implementada prescindiendo de *Docker,* pero puede resultar más trabajoso dado que se debe configurar cada elemento manualmente.

Tabla 2. Detalle del hardware de la solución

Función	Hardware
RECOLECCION DE DATOS AMBIENTALES	ESP 32: Microcontrolador
	DHT22: Sensor de temperatura y humedad ambiente
	DS18B20: Sensor de temperatura de suelo
	BH1750: Sensor de intensidad lumínica
	Sensor de humedad de suelo capacitivo
CAMARA NDVI	Raspberry PI Zero W
	Pi NoIR: Cámara infrarroja
	Filtro azul para procesar el NIR (Near Infra red)
HoT Host	Rapsberry PI 3

La Figura 4, presenta un diagrama en bloques de los distintos componentes, incluidos los elementos de hardware.

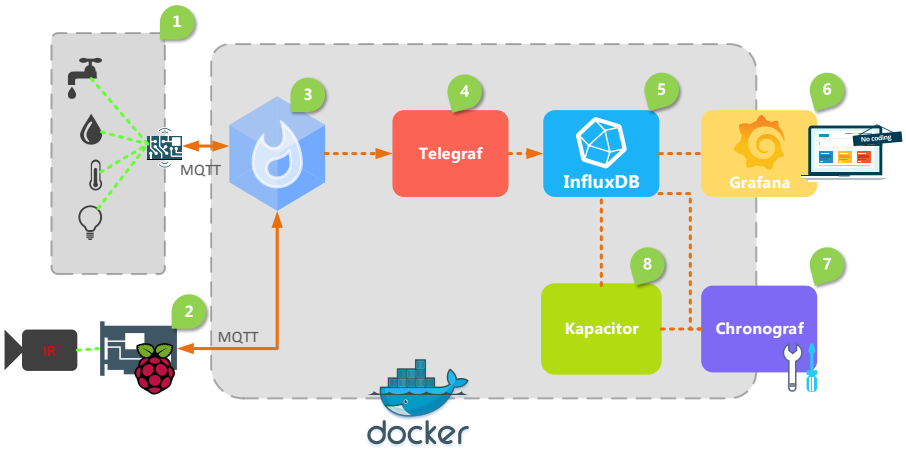


Figura 4. Arquitectura de la solución propuesta

La solución posee varios contenedores, que son instancias de una imagen. Una imagen es un conjunto de archivos inmutable que proporcionan un software pre-instalado[6]. La interrelación de los contenedores, representada en la figura 4, realizan el siguiente proceso:

1. Recolección de datos ambientales: sus sensores capturan las variables ambientales y el micro controlador las procesa y envían los valores vía MQTT al HoT.
2. Cámara *NDVI*: Captura imágenes en intervalos regulares, calcula el *NDVI* y envía el valor promedio vía *MQTT* al *HoT*.
3. *Hub Of Thing*: es un conjunto de componentes, como se describe en la **Figura 3.**

   a. Capa de adaptación: Adaptación de otros protocolos a formato de mensajes *MQTT*
   b. MQTT Bróker: Bróker *MQTT* conectado a la seguridad de HoT
   c. Interface HTM: Interface que permite la ver y crear dispositivos, monitorear sus valores y enviar comandos.
   d. REST API: interface que permite administrar la definición de dispositivos
   e. *Identity Server*: Servicio de validación de usuarios
   f. MongoDB: Base de persistencia de datos NoSQL
4. *Telegraf*: agente de recolección y persistencia de datos, se conecta al *MQTT* Bróker y persiste los valores numéricos en *InfluxDB*
5. *InfluxDB*: base de datos de series de tiempo para persistencia de datos a largo plazo.
6. *Grafana*: plataforma web para la visualización de datos, se conecta a InfluxDB y permite realizar gráficos personalizados con la información.
7. *Cronograf*: interface de administración de *InfluxDB* y *Kapacitor*, facilita la configuración de ambos elementos mediante una interface web.
8. *Kapacitor*: agente de procesamiento de datos en línea. Permite crear alerta de los valores transmitidos a *InlfuxDB*

3.1 Persistencia y visualización de datos

En los sistemas IoT los sensores, conectados a los dispositivos, generan valores en un momento determinado, es decir que la información generada tiene el formato [sensor, tiempo, valor]. Las bases de datos de serie de tiempo o *TSDB*, están diseñadas especialmente para manejar eventos y métricas combinadas con el tiempo.

Originalmente desarrolladas para analizar información financiera como la evolución de precios de las acciones, las bases de datos de serie de tiempo están encontrando un nuevo uso en el IoT. A diferencia de otro tipo de bases de datos, las *TSDB,* incluyen almacenamiento y compresión de datos con sello de tiempo, administración del ciclo de vida de los datos, funciones de agrupación, capacidad de manejar grandes escaneos dependientes de series de tiempo de muchos registros y consultas de series de tiempo.

3.2 Retención y reducción de muestras

Los sensores conectados a IoT pueden generar una gran cantidad de datos y con una implementación a mediano o largo plazo el espacio de almacenamiento es un punto de importancia a tener en cuenta. A fin de no causar una falla de sistema por falta de almacenamiento es necesario ajustar las políticas de retención y reducción de datos.

La política de retención permite poner un límite a la antigüedad de los datos. Una vez superado ese umbral de tiempo los datos son borrados.

Eliminar los datos "antiguos" no siempre es posible, en esos casos se puede optar por una técnica de reducción de muestras o *Downsampling*, Esta técnica consiste en agrupar todas las muestras en intervalos regulares (por ejemplo 30 minutos, una hora, o una instante de tiempo que se defina) y aplicar alguna función de agregación como puede ser suma, promedio, razón, etc., lo cual reduce ostensiblemente la cantidad de información almacenada.

El método de reducción de muestras puede ser aplicado en *InfluxDb* utilizando la política de retención de datos en la métrica donde impactan los valores de los sensores para que expire al transcurrir un cierto periodo de tiempo, por ejemplo, un día. Luego se programa una consulta continua (*Continuous Query*) que se ejecuta cada 24 horas y calcula la razón de las muestras de cada sensor agrupadas en intervalos de una hora y persistida.

4 Conclusiones y futuras líneas de investigación y desarrollo

El ***Hub Of Things*** es una herramienta que propone una base o punto de inicio a otras soluciones *IoT* a problemas particulares o como integrador para soluciones protocolos no *TCP/IP*.

La experiencia de usuario se extiende más allá del software y el tratamiento de los datos. El desarrollo de dispositivos de borde no es trivial ni sencillo. Si bien escapa al alcance de este trabajo, el dispositivo de borde es la "cosa" en la internet de las cosas.

La carencia de valor agregado o un funcionamiento deficiente no se puede subsanar con funcionalidad adicional.

La utilización de estadística para la reducción de datos es esencial para que la información importante no se pierda en el proceso.

El uso de herramientas de ciencia de datos también es un elemento que se debe considerar a fin para expandir el análisis de datos de la solución.

Docker es una gran herramienta a tener en cuenta, ideal para orquestar una solución que puede ser implementada tanto en la nube como localmente y que facilita la actualización.

Las futuras líneas de investigación y desarrollo se verán orientadas a la recopilación de datos y la integración de sistemas. Cada una de ellas, puede ser un trabajo independiente, pero se mencionan por que utilizarán el ***Hub Of Things*** como elemento central. Estas pueden ser:

- Medición de consumo eléctrico hogareño.
- Monitoreo de variables ambientales urbanas como la concentración de CO_2 o el nivel de ruido ambiente.
- Integración de sistemas inalámbricos *IoT* de largo alcance con *LoRa*.
- Integración de sistema cableados *Modbus*.

5 Referencias

[1] Cisco, «Fog Computing and the Internet of Things,» 2015. [En línea]. Available: https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf. [Último acceso: FEB 2016].

[2] Google, «Material Design - Guidelines,» 2014. [En línea]. Available: https://material.io/guidelines/. [Último acceso: NOV 2016].

[3] D. Strazzullo, D. D. Dailey and J. Frost, Building Web Applications with SVG, Sebastopol, California: O'Reilly Media, Inc., 2012.

[4] OASIS, «MQTT Version 3.1.1,» OCT 29 2014. [En línea]. Available: http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html. [Último acceso: ENE 2016].

[5] J. Weier y D. Herring, «NASA-Earthobservatory-Measuring Vegetation (NDVI & EVI),» 30 AUG 2000. [En línea]. Available: https://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring_vegetation_1.php. [Último acceso: JUN 2017].

[6] Docker Inc., «Docker glossary,» Docker Inc., [En línea]. Available: https://docs.docker.com/glossary/. [Último acceso: JUN 2018].

[7] InfluxData, Inc, «Open Source Time Series Platform,» InfluxData, Inc, 2018. [En línea]. Available: https://www.influxdata.com/time-series-platform/. [Último acceso: MAY 2018].

[8] Grafana Labs, «Grafana: testimonials,» Grafana Labs, 2018. [En línea]. Available: https://grafana.com/grafana/testimonials. [Último acceso: 25 MAY 2018].

[9] Cisco, «Cisco Visual Networking Index Predicts Near-Tripling of IP Traffic by 2020,» 07 JUN 2016. [En línea]. Available: https://newsroom.cisco.com/press-release-content?type=press-release&articleId=1771211. [Último acceso: FEB 2018].

[10] Ericsson, «Ericsson Mobility Report November 2015,» NOV 2015. [En línea]. Available: https://www.ericsson.com/assets/local/news/2016/03/ericsson-mobility-report-nov-2015.pdf. [Último acceso: FEB 2018].

[11] Gartner, «Gartner Says 6.4 Billion Connected "Things" Will Be in Use in 2016,» 10 NOV 2015. [En línea]. Available: https://www.gartner.com/newsroom/id/3165317. [Último acceso: FEB 2018].

[12] S. Jankows, J. Covello, H. Bellini, J. Ritchie y D. Costa, «The Internet of Things: Making sense of the next mega-trend,» 3 SEP 2014. [En línea]. Available: http://www.goldmansachs.com/our-thinking/outlook/internet-of-things/iot-report.pdf. [Último acceso: FEB 2018].

Redes 5G y Virtualización de funciones de red en Argentina

Carlos Peliza, José Luis Roca, Fernando Dufour, Ariel Serra, Gustavo Micieli

Universidad Nacional de La Matanza
Florencio Varela 1903 (B1754JEC) - San Justo, Buenos Aires, Argentina
pelizac@gmail.com

Abstract. Este trabajo, primero pretende introducir los conceptos fundamentales de una arquitectura con funciones de red virtualizadas (NFV), revisando para ello la bibliografía disponible para enunciar las posibilidades de desarrollo de la arquitectura. Como segundo punto, clasificar el estado de las redes de quinta generación en Argentina. Mediante el análisis de lo actuado en Argentina, buscaremos describir el estado de las redes de 5G virtualizadas en el país y cuál ha sido el tipo de desarrollo NFV que se ha decidido implementar.

Software Defined Networking (SDN) se centra en la separación entre el plano de control, encargado del mantenimiento y control de las conexiones en la red, y el plano de datos en las redes, propio del intercambio de información de usuario mientras que la virtualización de las funciones de red (NFV) desacopla las funciones de red en ambos planos, del hardware que se utilice.

Este desacoplamiento de planos permite el despliegue de componentes de software de plano de control y la implementación de software de plano de usuario (llamados VNF) en plataformas informáticas que son mucho más potentes que los equipos de red tradicionales.

La separación de planos permite desarrollos generados por distintos proveedores de software y de hardware. La arquitectura NFV postula el desarrollo y despliegue de soluciones interoperables y el trabajo cooperativo entre fabricantes y desarrolladores.

Las redes móviles 5G postulan un avance en redes móviles que es tomado como driver de desarrollo para NFV por lo que resulta procedente preguntarse, ¿cuáles son las funciones virtualizadas para las nuevas redes? ¿Cuáles son las características del desarrollo 5G? ¿Cuál es el estado en Argentina?

Keywords: Virtualización, SDN, VNF, NFV, Redes Móviles, 5G.

1 Introducción

En el año 2012, durante el Congreso “SDN and OpenFlow World Congress” del mes de Octubre en Darmstadt, Alemania fue presentado el whitepaper “Network Functions Virtualization” con el objetivo de describir los beneficios, habilitadores y desafíos para la virtualización de funciones de red y como justificación para fomentar una colaboración internacional que acelerara el desarrollo y despliegue de soluciones interoperables para servidores de alto volumen, basadas en estándares. En la propues-

ta del whitepaper estaban definidos los campos de acción y casos de uso (NFV_White_Paper, p 6) entre los cuales se enumeraban:

- Elementos de conmutación: BNG, CG-NAT, enrutadores
- Nodos de red móvil: HLR / HSS, MME, SGSN, GGSN / PDN-GW, RNC, nodo B, eNodo B.
- Funciones contenidas en enrutadores y decodificadores domésticos.
- Análisis de tráfico: DPI, medición de QoE.
- Señalización NGN: SBCs, IMS.
- Funciones convergentes y de toda la red: servidores AAA, control de políticas.
- Optimización a nivel de aplicación: CDN, servidores de caché, equilibradores de carga.
- Funciones de seguridad: cortafuegos, escáneres de virus, sistemas de detección de intrusos, protección contra correo no deseado.

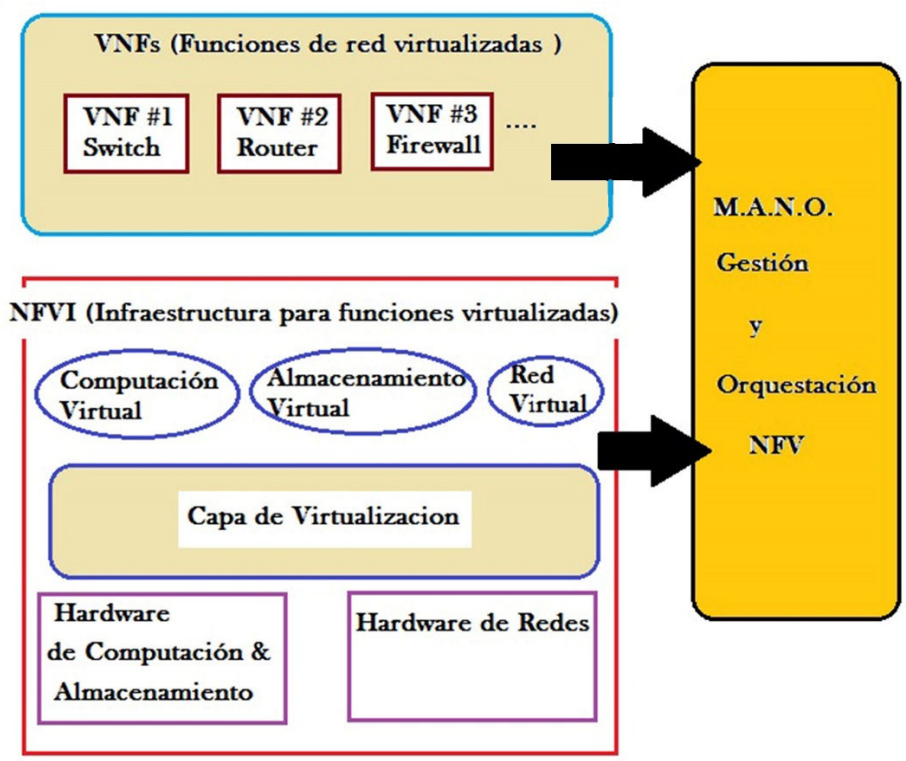


Ilustración 1 - Framework NFV

Dentro de lo que se denomina drivers en la industria de Telecomunicaciones, o sea factores que se analizan para adoptar una tecnología novedosa en el mencionado congreso se enumeraban:

- Necesidades de diseño para nuevos equipamientos

- Costes y restricciones físicas de fabricación
- Alto nivel de conocimiento necesario para operar las soluciones propietarias de HW/SW
- Complejidad HW en las soluciones de fabricante

El ciclo de producto comienza antes de haber podido comenzar el retorno de inversión

2 Objetivos

El presente trabajo tiene como objetivo analizar la arquitectura NFV y distinguir en ella los desarrollos que son posibles de seguir y a partir del estudio del estado de implantación una red 5G en Argentina, verificar cuál ha sido el desarrollo de arquitectura NFV utilizado.

El estilo de este trabajo de investigación es comparativo y se basa en el análisis de fuentes bibliográficas y documentación existente, junto a la realización de pruebas de concepto y funcionamiento del servicio 5G.

3 Entendiendo NFV

La virtualización de las funciones de red (NFV) es un enfoque de red propuesto por ETSI (European Telecommunications Standards Institute) que permite la sustitución de dispositivos de hardware dedicado, tales como routers, firewalls y equilibradores de carga entre otros equipamientos, por dispositivos basados en software que se ejecutan como máquinas virtuales en servidores estándares de la industria.

El viaje a una red NFV completamente operacional requiere la coordinación de tres desarrollos interconectados, pero totalmente separados: Virtualización, Orquestación y Automatización, ninguno de estos caminos puede considerarse de manera aislada (Ashton Metzler, 2015, p.47)

NFV desacopla las funciones de la red de dispositivos de hardware dedicados y las traslada a uno o varios servidores virtuales, que pueden cumplir múltiples funciones en un único servidor físico. Este enfoque reduce los costos y minimiza el mantenimiento, debido a que los dispositivos virtuales reemplazan dispositivos de red basados en hardware dedicado. La NFV no debe confundirse con una red virtualizada, porque la NFV, es un marco o arquitectura que busca descargar sólo las funciones de red, y no toda la red. También es importante señalar que la NFV es diferente de una función virtual de red (VNF), un término comúnmente utilizado para describir una función de red que se ejecuta en el software de una máquina virtual (Terminology for Main Concepts in NFV, gs_NFV003v010201p).

La amplia gama de opciones de desarrollo que presenta NFV puede expresarse mediante el siguiente cuadro de elaboración propia:

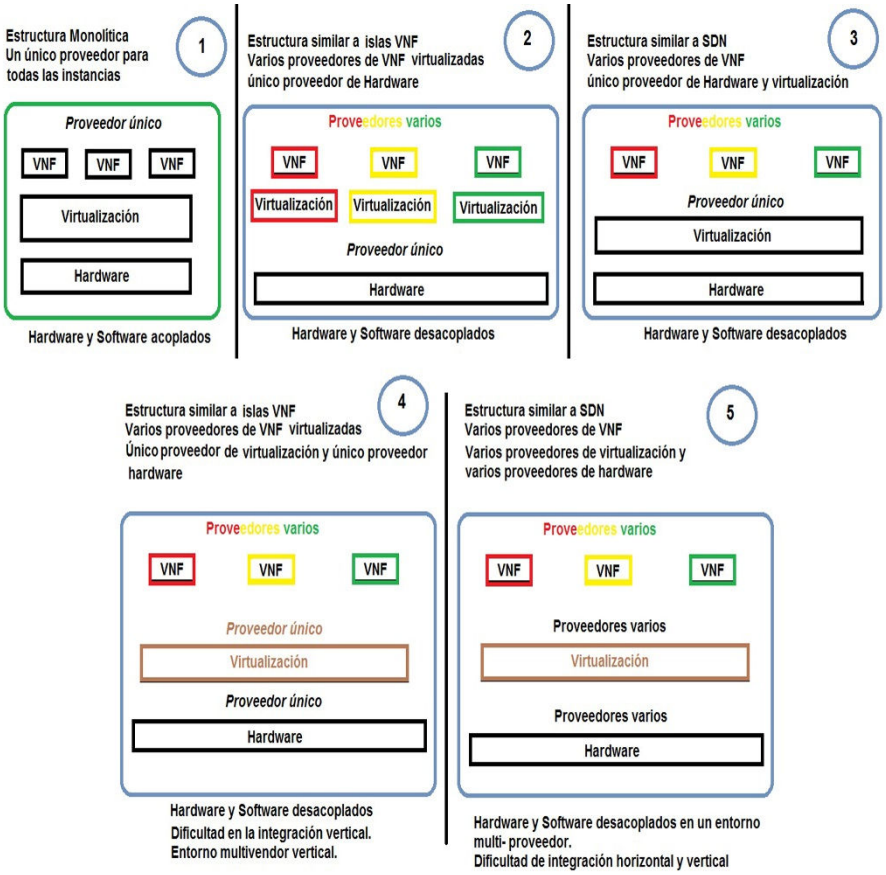


Ilustración 2 - Opciones de desarrollo NFV

4 Redes de Quinta generación 5G

En Diciembre de 2017, el 3rd Generation Partnership Program (3GPP) aprobó, en encuentro plenario en Lisboa (Portugal), las especificaciones de Nueva Radio 5G No Autónoma (NSA 5G NR, o Non-Standalone 5G New Radio). Se trata del primer estándar de Quinta Generación (5G) de redes móviles aprobado oficialmente por el 3GPP (Fuente 5G Americas). (http://www.5gamericas.org/en/)

Se trata de las especificaciones técnicas para Non-Standalone 5G New Radio (Nueva Radio 5G No Autónoma), que se apoyará en las redes existentes 4G LTE. Este primer hito contribuirá a acelerar la llegada de una quinta generación (5G) de tecnologías móviles estandarizadas. El cronograma de avance en la estandarización propuesto por el 3GPP se puede ver en la siguiente figura.

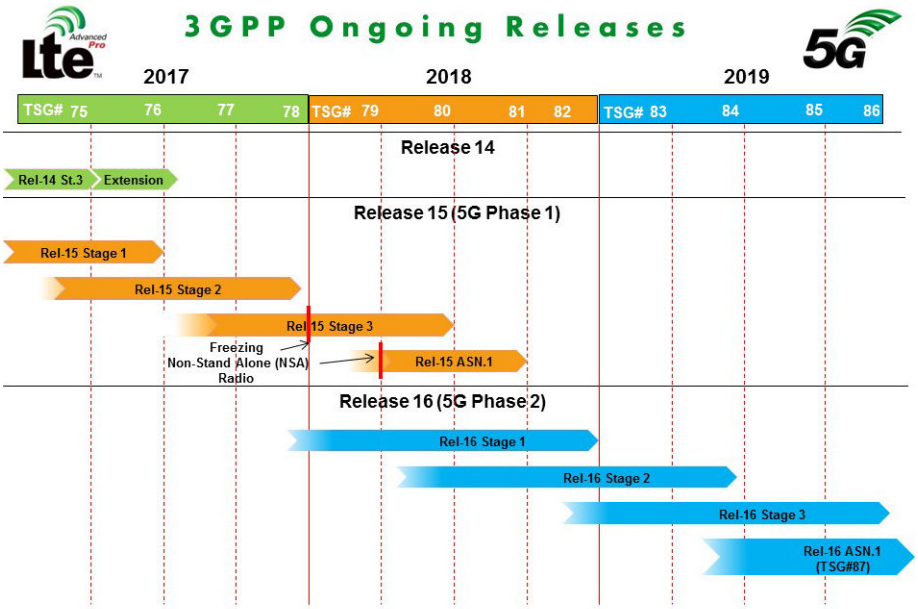


Ilustración 3 – Cronograma de avance de estandarización.

El análisis de las releases del 3GPP, pasadas y futuras, permite establecer una línea lógica en el desarrollo de las nuevas tecnologías. Así, queda determinado que se priorizará el desarrollo integrado de las redes 4G y 5G, compartiendo en un primer paso la infraestructura de Core y de Backhaul provenientes de LTE o 4G, para en un paso posterior estandarizar los desarrollos de 5G que serán virtualizados.

Se ha establecido como nombre de cada desarrollo de la estandarización de 5G,el primer paso es llamado 5G Non Stand Alone o 5G NSA y su desarrollo posterior 5G Stand Alone, donde se realizaran las especificaciones propias de una red 5G sin compartir recursos de su predecesora, la red LTE 4G.

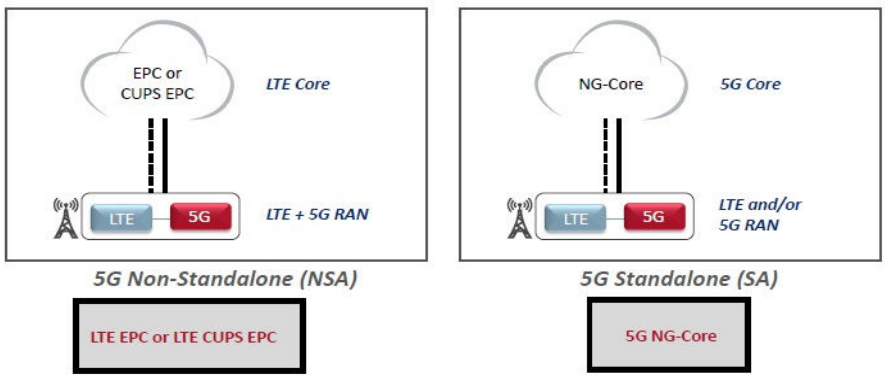


Ilustración 4 Opciones de desarrollo 5G

En la actualidad, se encuentran disponibles en vías de estandarización dispositivos de acceso por radio, tradicionalmente llamados Radiobases, que pertenecen a la Radio Access Network (RAN) para redes 5G NSA (Non Stand Alone).

4.1 Características esperadas de la tecnología 5G

El mercado de las telecomunicaciones pretende las siguientes mejoras de la nueva tecnología móvil

- 10 a 100 veces más velocidad del usuario.
- 1000 veces de volumen de datos móviles.
- latencia 5 veces menor
- mayor duración de la batería
- Incremento por 100 de dispositivos conectados

Dichas pretensiones están basadas en la nueva gama de servicios que pretenden ofrecer a sus usuarios.

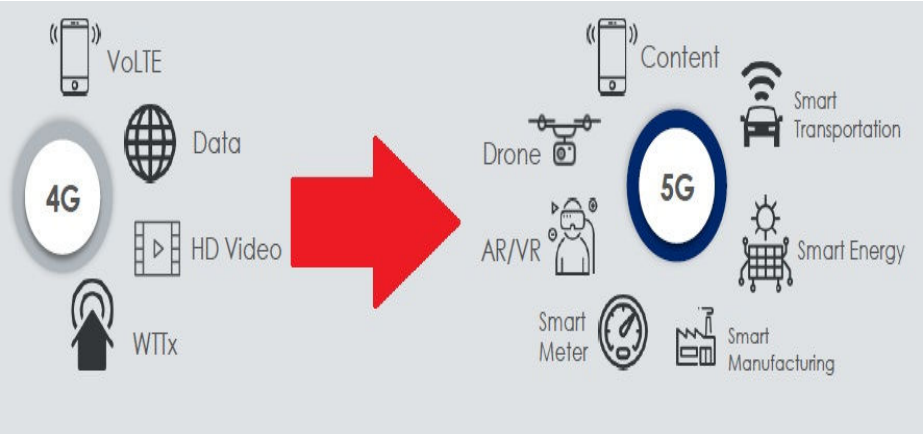


Ilustración 5 Nuevos servicios 5G

5 Las pruebas realizadas

Dentro del marco controlado de la Radio Access Network (RAN) se pretendió mediante dispositivos prototipo, generar las condiciones que demandarán los usuarios de la RAN 5G, como ser tasas de descarga de información (Down Link) de 20 Gbps, seguimiento y formación de haces (Beam Tracking / Beamforming) para mejorar la capacidad de cobertura, descarga de streams de video 4K que permitirá ver video 4K a través de la red móvil.

Debió ser utilizado como cliente, un prototipo que emula a los múltiples dispositivos que utilizaran la 5G RAN, ya que al momento de la prueba, no existen en el mercado de consumo masivo dispositivos móviles que acepten las tecnologías 5G.Existe una reducida cantidad de celulares o equipos de Internet de banda angosta (NB-IoT) que las aceptan y son usados en las pruebas en casa matriz de los fabricantes de dispositivos.

El prototipo utilizado se puede ver en la imagen siguiente:



Ilustración 6 Prototipo de simulación

Mientras que para la simulación del lado proveedor del servicio se utilizó la siguientes radiobase con su array de antenas.

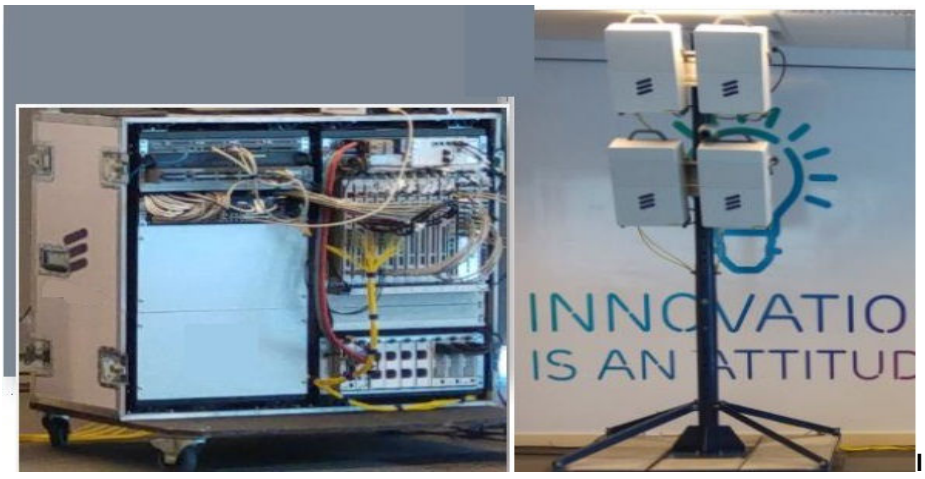


Ilustración 7 Prototipo de Radiobase y antenas

Los resultados obtenidos de las pruebas de simulación y concepto pueden ser considerados satisfactorios.

6 Unión de proveedores NFV para este caso de estudio

Dentro de los más de 800 miembros de 66 países en los cinco continentes se pueden encontrar desarrolladores de software, hardware y servicios de red[1] con membresía en ETSI. Una de las premisas sugeridas por ETSI para el desarrollo de NFV fue la búsqueda de plena interacción entre fabricantes y desarrolladores, sin que ello implicara membrecía al grupo de trabajo[2], de igual manera resulta pertinente verificar la pertenencia de cada proveedor propuesto para trabajos de NFV, al universo ETSI, pero debe mencionarse que ello no implica un diferencial favorecedor. Sin embargo, en este caso Telefónica junto a Ericsson, forman parte del board del ETSI NFV, de lo que resulta apropiado afirmar que la 5G RAN va a estar en posición de ser testeada en simultaneo a la estandarización de la integración de 5G con las redes LTE conformando la implantación 5G Non Stand Alone.

7 Conclusiones y posibles líneas de Investigación

El análisis del marco teórico de la tecnología NFV nos permite conocer que uno de los driver que permitirá su desarrollo está ligado a brindar nuevos servicios para la telefonía móvil, que permitan lograr una llegada rápida al mercado.

Los análisis de mercado de las empresas de Telecomunicaciones indican que las tendencias se dirigen al uso del servicio móvil en tiempo real o de muy baja latencia y esa cualidad solo es posible con tiempos de latencia como se ha estandarizado en 5G.

En Argentina a finales de 2017, en simultaneo con el cierre del proceso de estandarización, se han realizado pruebas satisfactorias de una implantación 5G RAN para Non Stand Alone con dispositivos demostrativos.

Al momento de la prueba, existen disponibles redes NFV con solución monolítica de LTE que por problemas de disponibilidad de tiempo no fueron exploradas para conexión con la RAN 5G, por o que no resulta factico considerar una conectividad posible. Será una experiencia que debe transitarse.

Por otro lado, se están desarrollando pruebas de VEPC (Core Evolucionado de Red de Paquetes Virtualizado) para hacer de un corazón de red de paquetes evolucionados, propio de las redes LTE, una VNF, un servicio de red virtualizado que respete el concepto de la arquitectura NFV.

Todo lo expuesto nos permite afirmar que en la Argentina se ha comprobado el funcionamiento del servicio de 5G que cumplirá con los estándares a cumplimentar en el mundo.

[1] http://www.etsi.org/membership/current-members.

[2] http://www.etsi.org/images/files/ETSIGenericPresentation.pdf

El estado de las redes 5G en el mundo, se encuentra dando los primeros pasos para la completa estandarización, en la Argentina, se cuenta con el potencial de elementos para realizar la segunda parte del proceso de estandarización de a implantación 5G Non Stand Alone, a saber, Núcleo de Red Virtualizado (por un único proveedor) y núcleo de red virtualizado por arquitectura NFV no monovendor y bajo el concepto de VNF

Una posible línea para futuras investigaciones es realizar un análisis de rendimiento cuando se presente el caso de varios proveedores del servicio de orquestación, por ejemplo, la existencia de dos o más VNFs cuyo plano de control u orquestación sea operado con dos proveedores diferentes, por ejemplo, OpenMano y Cyan u otros orquestadores a desarrollarse.

8 Glosario

AAA	Authentication, Authorization & Accounting
BNG	Broadband Network Gateway
BRAS	Broadband remote access server
CCMP	Counter Mode CBC-MAC Protocol
CDN	Content Delivery Network
CPEs	Customrer Premise Equipments
DHCP	Dynamic Host Configuration Protocol
DPI	Deep Packet Inspector
GGSN	Gateway GPRS Support Node
HLR	Home Location resources
HSS	Home Subscriber Server
IEEE	Inst, of Electrical and Electronics Engineers
IMS	IP Multimedia System
MAC	Media Access Control
ME	Mobile Equipment
MME	Mobility Management Entity
MS-BNG	Multi Service Broadband Network Gateway

NA(P)T	Network Address (Port) Translation
NAI	Network Access Identifier
NGN	Next Generation Network
PCC	Policy and Charging Control
PCEF	Policy and Charging Enforcement Function
PCRF	Policy and Charging Rules Function
PDN	Packet Data Networks
PDN GW	Packet Data Networks Gateway
PLMN	Public Land Mobile Network
PSK	pre-shared key
QCI	QoS Class Identifier
QoE	Quality of Experience
QoS	Quality of Server
RA	Router Advertisement
RADIUS	Remote Authentication Dial In User Service
RBN	Regional Broadband Network
RNC	Radio Network Controller
SBCs	Service Border Controllers
SDN	Software Defined Network
SDWN	Software-defined wireless network
SeGW	Security Gateway
TDF	Traffic Detection Function
TIER x	ISP de nivel x
UDP	User Datagram Protocol
VLAN	Virtual Local Area Network
VPN	Virtual Private Network
WG	Working Group

9 Referencias

1. Akram Hakiri, Aniruddha Gokhale, Pascal Berthou, Douglas C.Schmidt, Thierry Gayraud. (2014). Software Defined Networking: Challenges and research opportunities for Future Internet. Computer Networks Volume 75, Part A (pp 453-471).
2. AT&T: Margaret Chiosi.BT: Don Clarke, Peter Willis, Andy Reid. China Mobile: Dr. Chunfeng Cui, Dr. Hui Deng.et all. (s.f.). European Telecommunications Standards Institute. Recuperado el 29 de 04 de 2018, de ETSI: http://portal.etsi.org/NFV/NFV_White_Paper.pdf
3. Ashton Metzler, Jim Metzler. The 2015 guide to SDN and NFV. Webtorials. 2015
4. Chris Wolf, Erick M. Halter. Virtualization: From the Desktop to the Enterprise (Books for Professionals by Professionals). Apress; 1st ed. edition (May 26, 2005), ISBN-10: 1590594959, ISBN-13: 978-1590594957.
5. Edouard Bugnion, Jason Nich, Dan Tsafir. Hardware and Software Support for Virtualization (Synthesis Lectures on Computer Architecture). Morgan & Clay-pool Publishers (February 21, 2017), ISBN-10: 1627056939, ISBN-13: 978-1627056939.
6. ETSI GS. Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV. NFV 003 V1.2.1.2014
7. Ken Gray, Thomas D. Nadeau - Network Function Virtualization – Morgan Kaufmann - 2016 - ISBN 978-0128021194
8. Manuel Paul, Sibylle Schaller, Malcolm Betts, Dave Hood, Meral Shirazipour, Diego Lopez, John Kaippallimalil, Applying_SDN_Architecture_to_5G_Slicing_TR-526.pdf recuperado de https://www.opennetworking.org/software-defined-standards/archives/
9. N. Mckeown, T. Anderson, L. Peterson, J. Rexford, S. Shenker, and S. Louis, “OpenFlow: Enabling Innovation in Campus Networks.(2008).ACM SIGCOMM Computer Communication. Volume 38 Issue 2, April 2008 (pp 69-74).
10. Rajendra Chayapathi, Syed F. Hassan, Paresh Shah - Network Functions Virtualization (NFV) with a Touch of SDN – Pearson Education - 2017 – ISBN 978-0134463056
11. Sevil Mehraghdam, Matthias Keller, Holger Karl - Specifying and placing chains of virtual network functions – IEEE – 2014 - ISBN 978-1-4799-2730-2
12. 5G New Radio and System Standardization in 3GPP descargado de https://www.itu.int
13. https://images.samsung.com/is/content/samsung/p5/global/business/networks/insights/event/the-silicon-valley-5g-summit-2017/Session-1_3GPP_Balazs-Bertenyi.pdf
14. (3GPP TS 36.413 versión 10.9.0 Release 10) LTE; Evolved Universal Terrestrial Radio Access Network (E-UTRAN);S1 Application Protocol (S1AP).
15. (3GPP TS 36.331 version 10.19.0 Release 10) LTE; Evolved Universal Terrestrial Radio Access (E-UTRA);Radio Resource Control (RRC) Protocol specification.

16. http://howltestuffworks.blogspot.com.ar/2014/10/carrier-aggregation-secondary-cell.html.
17. http://www.3gpp.org/keywords-acronyms/1612-ue-category.
18. Chris Johnson, 2012, Long Term Evolution in BULLETS - 2nd Edition, Inglaterra, Independiente.
19. http://howltestuffworks.blogspot.com.ar/2011/10/ue-capability-information.html.
20. http://www.sharetechnote.com/html/Handbook_LTE_UE_Capability.html.
21. Anders Hedlund y Irina Cotanis, 2014,Introduction to Carrier Aggregation Testing, Ascom.
22. http://es.slideshare.net/veermalik121/throughput-calculation-for-lte-tdd-and-fdd-system.
23. (3GPP TS 36.213 version 10.13.0 Release 10) LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures.
24. http://www.3gpp.org/technologies/keywords-acronyms/101-carrier-aggregation-explained,
25.

Variante Heurística para el Problema del Viajante. Caso de Aplicación: Circuito de Pesca Deportiva

Ana Priscila Martinez[1],
Lidia Marina López[2]

Universidad Nacional del Comahue, Facultad de Informática, Neuquén, Argentina.
[1] martinezanapriscila@gmail.com
[2] lidia.lopez@fi.uncoma.edu.ar

Abstract. El presente trabajo se enfoca en la construcción de un algoritmo para resolver un circuito de pesca deportiva, aplicando técnicas de optimización combinatoria con el objeto de generar la mejor solución al problema del recorrido para la pesca deportiva en la provincia de Neuquén. La planificación y gestión de caminos para recorridos con preferencias exige disponer de sistemas eficientes de optimización de rutas. Su complejidad es exponencial. Para la resolución de este tipo de problemas se deben emplear heurísticas que permitan soluciones factibles. Para modelar un circuito turístico asociado a la pesca deportiva se utiliza la exploración de un grafo con restricciones. El mismo se encuadra dentro del Problema del Viajante. Se propone el diseño de un algoritmo metaheurístico de búsqueda tabú, basado en una búsqueda local, para encontrar una solución al problema.

Palabras Claves: Grafo, Heurística, Problema del Viajante, Búsqueda Tabú, Algoritmo.

1 Introducción

El presente trabajo se enfoca en la construcción de un algoritmo para resolver un circuito turístico con determinadas características, aplicando técnicas de optimización combinatoria con el objeto de generar la mejor solución al problema del recorrido para la pesca de salmónidos en la provincia de Neuquén.

1.1 Objetivo

El objetivo principal de este trabajo es el estudio, construcción e implementación de un algoritmo que resuelva el armado de un circuito turístico específico, en este caso para la pesca deportiva, mediante técnicas de exploración de grafos. Es decir, dada una fecha de inicio y fin del recorrido, tipo de pesca y ciudad de origen, devuelva un camino óptimo con todos los accesos de pesca habilitados para esas especificaciones.

El diseño de recorridos óptimos o con restricciones sobre rutas geográficas involucra la aplicación de algoritmos de exploración de grafos y de sistemas de información geográfica. Dado que estos caminos no se modifican, el problema de diseño de rutas, con preferencias determinadas, se encuadra en las variaciones del clásico *Problema del Viajante.* Este tipo de problemas no es abordable con técnicas de resolución exactas, salvo para problemas muy pequeños, debiéndose emplear heurísticas para encontrar soluciones factibles.

Este trabajo presenta un algoritmo basado en grafos representados por matrices de adyacencia, y una heurística para la obtención del grafo resultado.

2 El Problema del Viajante (TSP)

Un viajante quiere visitar n ciudades, una y sólo una vez cada una, empezando por una cualquiera de ellas y regresando al mismo lugar del que partió. Supongamos que conoce la distancia entre cualquier par de ciudades. ¿De qué forma debe hacer el recorrido si pretende minimizar la distancia total? [6]

A este problema se lo conoce con el nombre de *problema del viajante* o TSP (Travelling Salesman Problem) y resulta ser uno de los más prominentes dentro del campo de la optimización combinatoria. Es un problema de complejidad *NP- difícil.* A pesar de su aparente sencillez aún no se ha conseguido encontrar un algoritmo que sea capaz de resolverlo en un tiempo razonable (polinómico). [2]

Matemáticamente el TSP corresponde al problema de hallar un *ciclo hamiltoniano* de mínima distancia en un grafo completo $G = (V, E)$ en donde $V = \{1, 2, ...,n\}$ es el conjunto de nodos y representan las ciudades, E es el conjunto de ramas que denotan la conexión entre ellas (rutas) y $d : E \rightarrow R^+$ una función que a cada $(i, j) \in E$ le asigna la distancia d_{ij} entre las ciudades i y j.

3 Métodos Algoritmos Heurísticos

"Un método heurístico es un procedimiento para resolver un problema de optimización bien definido mediante una aproximación intuitiva, en la que la estructura del problema se utiliza de forma inteligente para obtener una buena solución." [8]

En la *optimización combinatoria* se plantea un problema de optimización como un par *(F, h)* donde F es el conjunto de soluciones factibles de una instancia de un problema de optimización y $h: F \rightarrow R^+$ es la función costo. [7]

El objetivo es encontrar un $f_0 \in F$ tal que $h(f_0) \leq h(f)$ para todo $f \in F$. La base del método consiste en asignar a cada $t \in F$ un entorno de t, es decir, un subconjunto $N(t) \subseteq F$ que contiene a t y a cuyos elementos los consideramos vecinos de t.

El *algoritmo de búsqueda local* consiste en partir de una solución inicial t_0, encontrar un $s \in N(t)$ con coste menor, para de esa forma disminuir $h(t)$, y repetir este procedimiento hasta llegar a un t en cuyo entorno resulte imposible disminuir el coste. Si esto pasa, se está en presencia de un *mínimo local*. [1]

Una *heurística,* es una función matemática *h (f)* que sirve como una estimación del coste del camino más económico de un nodo dado hasta el nodo objetivo. [8]

No garantizan una solución óptima, pero si aproximadas al resultado óptimo. Un buen algoritmo heurístico debe ser e*ficiente*, b*ueno*, y *robusto*.

4 Planteamiento del TSP con Algoritmos Heurísticos

Para enunciar el TSP a través de métodos heurísticos hay distintos tipos de categorías de heurísticas. Sólo se estudiara:

Mejora Iterativa. La idea básica consiste en comenzar con una solución inicial e ir realizando modificaciones para mejorar su calidad. El objetivo de esta técnica consiste en explorar iterativamente el estado de soluciones para encontrar las soluciones óptimas. Se detallan algunas técnicas a continuación [5]:

Intercambio par a par: El intercambio par a par involucra en cada iteración la eliminación de dos aristas y el reemplazo de éstas, con dos aristas diferentes que reconecten los fragmentos creados por la eliminación de las aristas produciendo un camino nuevo más corto. Esto es un caso especial del método *k-opt*.

Heurística k-opt o heurística Lin-Kernighan: Toma un recorrido dado y elimina *k* aristas mutuamente disjuntas. Reconecta los fragmentos conformando el recorrido, sin dejar ningún subcamino. Esto, en efecto, simplifica las consideraciones del TSP convirtiéndolo en un problema más simple. Cada extremo de un fragmento tiene $2k-2$ posibilidades de ser conectado: del total de $2k$-extremos de fragmentos disponibles, los dos extremos del fragmento que se está considerando son descartados. La técnica *k-opt* es un caso especial del *V-opt* o técnica *Variable-opt*.

Heurística V-opt: Mientras el método *k-opt* elimina un número fijo *k* de aristas del recorrido original, el método *variable-opt* no fija el tamaño del conjunto de aristas eliminadas. En cambio, este método aumenta el conjunto a medida que el proceso de búsqueda continúa. Este método es considerado la heurística más poderosa para el problema.

5 Caso de Estudio: Circuito de Pesca Deportiva

La construcción de un circuito turístico de pesca en la provincia de Neuquén, es el caso de estudio seleccionado para el desarrollo de una solución algorítmica aplicando las técnicas precedentes para el TSP.

5.1. Planteamiento del problema

El TSP corresponde a un recorrido de un grafo donde el origen viene determinado.

Para resolver el problema, se decidió implementar un algoritmo heurístico, ya que está demostrado que estos algoritmos son muy eficientes encontrando soluciones cercanas a la óptima.

Modelo. El mapa M de la región a estudiar, donde se encuentran georreferenciados los accesos de pesca, se modelará como un grafo *G = (V, E)*. Dicho grafo cumplirá que:

El conjunto de *aristas* no dirigidas $E \subseteq (V \times V)$ representará los segmentos de rutas o caminos entre dos accesos de pesca. El conjunto de *vértices V* representará entonces los accesos de pesca. Estos tendrán asociadas sus respectivas coordenadas geográficas respecto de algún punto en el espacio.

Requerimientos. El objetivo principal busca partir desde una región determinada, definir solamente los puntos en los que hay *accesos de pesca de salmónidos* (según la temporada en la que ese acceso está habilitado para la pesca, y el tipo de pesca), y volver al punto de partida.

Esta etapa de modelado requiere en primer lugar un proceso de obtención de la información cartográfica digital asociada a cada una de las zonas de pesca a trabajar y el pre-procesamiento realizado sobre la misma. A tal efecto, los accesos de pesca y sus caminos serán modelados por un grafo, y sus aristas (caminos) tendrán un costo asociado (la distancia/kilometraje).

5.2 Requerimientos

Se confeccionó la Matriz de Adyacencia, base generadora del grafo que mostrará gráficamente toda la información de los accesos de pesca. Esta matriz tiene la información en kilometraje, de todas las distancias que hay entre los distintos puntos del mapa, siempre y cuando exista un camino físico que los una. La información fue proporcionada por la Subsecretaría de Turismo de Neuquén. También se cuenta con la información de temporada y tipo de pesca para cada acceso de pesca.

Una vez confeccionada la matriz de adyacencia, se realiza el grafo, el cual es un grafo no dirigido y etiquetado, como peso de cada arco está el kilometraje entre cada nodo, y como nodo, las regiones detalladas en las tablas antes mencionadas.

5.3 Búsqueda Tabú

Este algoritmo es atribuido a Fred Glover quien menciono en 1986 [4]:

"Es mejor una mala decisión basada en información, que una buena decisión al azar, ya que, en un sistema que emplea memoria, una mala elección basada en una estrategia proporcionará claves útiles para continuar la búsqueda. Una buena elección fruto del azar no proporcionará ninguna información para posteriores acciones".

La búsqueda tabú [4] es un algoritmo metaheurístico, éste se aplica, generalmente, a problemas que no tienen un algoritmo o heurística específica que dé una solución satisfactoria.

La estructura de memoria más importante usada para determinar las soluciones permitidas es la *lista tabú*. Esta lista es la que permite salir de óptimos locales para alcanzar un óptimo global. En su forma más simple, una lista tabú es una memoria de corto plazo que contiene las soluciones que fueron visitadas en el pasado reciente

(menos de *n* iteraciones atrás, donde *n* es el número de soluciones previas que van a ser almacenadas, *n* es llamado la *tenencia del tabú*). Los atributos seleccionados de las soluciones recientemente visitadas son denominados *tabú-activos*. Las posibles soluciones que contengan elementos *tabú-activos* son *tabú*.

Cuando sólo un atributo es marcado como tabú, por lo general resulta en que más de una solución, es marcada como tabú. Algunas de estas soluciones, que ahora deben ser evitadas, podrían ser de excelente calidad y no serían visitadas. Para mitigar este problema, se introducen los *criterios de aspiración*: estos pueden modificar el estado de tabú de una solución, por lo tanto incluyendo la antes excluida solución en el conjunto de soluciones permitidas. Un criterio de aspiración muy utilizado es admitir soluciones que no son mejores que la *mejor solución* conocida al momento.

6 Solución Algorítmica Alcanzada

Se opta por desarrollar e implementar un *algoritmo metaheurístico de búsqueda tabú* basado en una búsqueda local, para encontrar una solución al problema.

En la búsqueda de la solución se codifica un algoritmo con el método de búsqueda tabú: *mejor vecino* y luego se fueron realizando *mejoras iterativas*.

El algoritmo propuesto comienza a buscar la solución en base a una *solución posible*. Esto es, a partir de un arreglo inicial con una posible solución, cuyo nodo inicial y final coinciden.

Los datos iniciales son: cantidad de nodos, nombre de los nodos y descripción, distancias entre los nodos, tipo de pesca, fecha de comienzo y fin de temporada según el tipo de pesca y la región.

Las estructuras de datos iniciales son: Matriz de distancias: *distancia*, Matriz de fecha de inicio y fin de temporada: *fechaTemp*, Matriz de tipos de pesca: *tipoPesca*, Arreglo con la solución inicial: *posibleSol*.

Para llenar con el contenido inicial el arreglo *posibleSol* se realizó un algoritmo que busca en la matriz de temporada *fechaTemp,* y tipos de pesca, *tipoPesca,* los accesos de pesca (nodos) que cumplan con los criterios ingresados por el usuario, los cuales son: Ciudad en la cual quiere empezar su recorrido, Fecha de inicio y fin del recorrido planeado, Tipo de pesca que va a realizar, las cuales pueden ser: Spinning, Fly Cast, Trolling o Spinning Exclusivamente con mosca. El arreglo *posibleSol* es el que se debe ordenar.

6.1 Codificación y Resolución del TSP mediante el Algoritmo de Búsqueda Tabú.

En esta sección se muestra cómo se resuelve el *Problema del Viajante (TSP)* usando una variación heurística del Algoritmo de *Búsqueda Tabú*.

El algoritmo de *Búsqueda Tabú* [15] es una mejora sobre la búsqueda local básica, que intenta superar los problemas de búsqueda local al no quedar atrapado en un mínimo local. De esta manera permite contemplar más resultados y compararlos; se logra permitiendo la *aceptación de movimientos que no mejoran el resultado*. Esta

implementación también permite escapar de soluciones sub-óptimas, mediante el uso de una *Lista Tabú*.

La lista Tabú es una lista de movimientos posibles que podrían incorporarse a una solución. Estos movimientos son operaciones de *intercambio*. En el caso en que se encuentra una nueva mejor solución, ese movimiento se hace *tabú* para un cierto *número de iteraciones.* Entonces, se añade a la lista de tabúes con un cierto valor denominado *tenencia Tabú* (Longitud Tabú), la cual se decrementa con cada iteración. Cuando esta es cero, el movimiento puede ser realizado y aceptado.

Una lista de tabú se representa como una matriz, donde cada celda representa la *tenencia tabú* de una operación de intercambio, si se intercambian dos ciudades. Esto evita que se retome un movimiento antes de explorar un poco.

Para permitir un movimiento tabú, es necesario aplicar criterios que permitan seleccionar un movimiento tabú en base a ciertas restricciones. Por ejemplo, un movimiento permite una nueva solución global mejor, por lo tanto, es aceptado, y su tenencia tabú se renueva.

En resumen, la *búsqueda tabú* realiza los siguientes pasos:

1. Crear una solución inicial, *posible solución*, luego llamarla *mejor solución*.
2. Encontrar el mejor vecino, basado en la *posible solución* aplicando ciertas operaciones de intercambio.
3. Si se alcanza el mejor vecino realizando un movimiento no tabú, se acepta como la nueva *mejor solución*, sino, encontrar otro mejor vecino.
4. Si se alcanza el número máximo de iteraciones (o alguna otra condición de parada), ir al paso cinco, caso contrario, seguir iterando desde el paso dos.
5. La mejor solución global, es la mejor solución al final de las iteraciones.

Este método heurístico está diseñado para escapar de la optimalidad local; se basa en el manejo y uso de una colección de principios que sirven para resolver el problema de manera inteligente, esto es, haciendo uso de memoria flexible para involucrar dos procesos: -la adquisición y, -el mejoramiento de la información; así, al tener cierta historia de los caminos ya recorridos y de los óptimos encontrados, se puede evitar permanecer en las mismas regiones, y recorrer regiones nuevas para encontrar otras soluciones mejores. [3]

Resolución y codificación. El código empieza en la clase de nivel superior *o main.* Esta clase contiene el método principal que inicializa los datos del problema y realiza la búsqueda tabú.

El algoritmo trata de encontrar la mejor solución vecina en cada iteración que no implique un movimiento tabú. La solución siempre es aceptada.

Cada iteración comienza desde la solución que dio la iteración anterior. Durante las iteraciones se realiza un seguimiento de cada solución generada y se almacena. La mejor solución, es la que se logra cuando el número predefinido de iteraciones ha terminado. A la hora de implementar el algoritmo, surgió el tema de la cantidad de caminos inexistentes entre dos pares de ciudades o accesos de pesca. Para ello se aplica Dijkstra (que devuelve el camino más corto entre dos nodos distintos) en cada iteración, antes de cada llamado al *getMejorVecino*.

A continuación la codificación en java el algoritmo del mejor vecino.

```
public static int[] getMejorVecino(ListaTabu tabuList,
AmbienteTSP ambienteTSP,int[] solInicial) {
     int[] mejorSol = new int[solInicial.length];
     System.arraycopy(solInicial, 0, mejorSol, 0,
mejorSol.length);
     int mejorCosto = ambienteTSP.getValorFuncion(solInicial);
     int ciudad1, ciudad2 = 0;
     boolean primerVecino = true;
     for (int i = 1; i <mejorSol.length - 1; i++) {
        for (int j=2; j<mejorSol.length-1; j++){
           int[] newMejorSol = new int[mejorSol.length];
           System.arraycopy(mejorSol, 0, newMejorSol, 0,
newMejorSol.length);
            newMejorSol = intercambio(i,j, solInicial);
            int newMejorCosto =
ambienteTSP.getValorFuncion(newMejorSol);
            if ((newMejorCosto > mejorCosto ||
primerVecino) &&
tabuList.listaTabu[i][j] == 0) {
                  primerVecino = false;
                ciudad1 = i;
                  ciudad2= j;
                System.arraycopy(newMejorSol, 0, mejorSol, 0,
newMejorSol.length);
                mejorCosto = newMejorCosto;
             }
          }
     }
     if (ciudad1 != 0 ) {
          tabuList.decrementaTabu();
          tabuList.moverTabu(ciudad1, ciudad2);
     }
      return mejorSol;
 }
```

Permite un movimiento que no mejora

No es un movimiento tabú. Solo cuando la tenencia tabú es cero, el movimiento puede ser aceptado

La *tenencia tabú* se decrementa con cada iteración

Cuando un movimiento se hace tabú se añade a la lista de tabúes.

La solución algorítmica encontrada combina adaptaciones de los algoritmos heurísticos iterativos: intercambio par a par, Búsqueda Tabú y de Dijkstra. También utiliza arreglos auxiliares con información útil para construir el recorrido y una estructura de pila.

Datos de Prueba.

Fecha desde: 02/07/2017 - Fecha Hasta: 01/08/2017

Ciudad Origen: Neuquen (Nodo 28) - Tipo de Pesca: Spinning (Int 0)

En la Figura 1 se muestra el grafo final y los resultados de las distintas estructuras utilizadas. En rojo están los *nodosPaso,* estos son los nodos por los cuales sólo se utilizan para unir dos destinos, y en verde los *nodosPesca.* En cuanto a las aristas o caminos están representados de la misma forma que los nodos. Se puede observar a simple vista que el camino que une el nodo 28 con el 29 es recorrido dos veces. Esto se puede deber a dos cosas: 1) no existen dos caminos distintos que la unan, o 2) al ser caminos de paso, ya que el nodo 29 al ser visitado por primera vez

es un *nodoPesca,* pero al visitarse por segunda vez es *nodoPaso,* éste no controla que los nodos ya hayan sido visitados. Como se aclaró en el presente trabajo, los caminos que sirven para unir dos *nodosPesca* que no son abordables mediante otros *nodosPesca*, no se controla si ya fueron visitados o no.

Camino Inicial: {28 29 27 13 9 6 3 1 0 28}
Camino antes de Dijkstra: {28 29 27 13 9 6 3 1 0 28}
Solución final: {28 29 27 13 9 11 10 7 6 5 3 1 0 29 28}
Costo: 1253 km

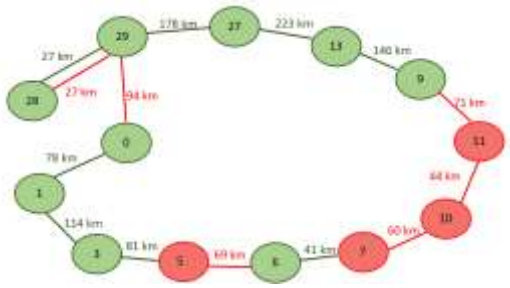


Figura 1. Grafo final del ejemplo uno. Distinguiendo *nodosPaso* de *nodosPesca*.

7 Eficiencia y Complejidad

Se modela un TSP perteneciente a problemas *NP-difíciles* debido a que el algoritmo solución conforma una transformación en tiempo polinomial del problema TSP genérico.

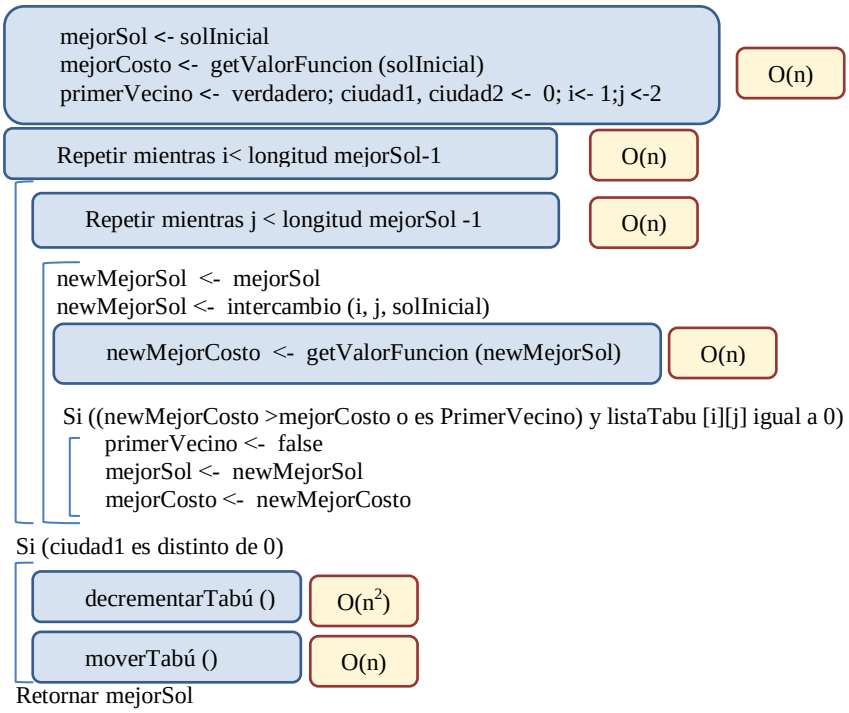


Figura 2. Diagrama de flujo y órdenes de ejecución del Algoritmo *GetMejorVecino* Final

En la Figura 2 se muestra el flujo de trabajo del algoritmo de mejor vecino junto con el orden de ejecución, el cual es O (n^3). En la Figura 3 se detalla un flujo general del algoritmo propuesto. Sobre el mismo se calcula el tiempo de ejecución total del algoritmo. Dicho algoritmo presenta un orden de ejecución polinomial, más precisamente $O(n^4)$.

El orden de un algoritmo se usa como medida de su eficiencia. Un algoritmo es considerado *eficiente* sólo en el caso polinomial.

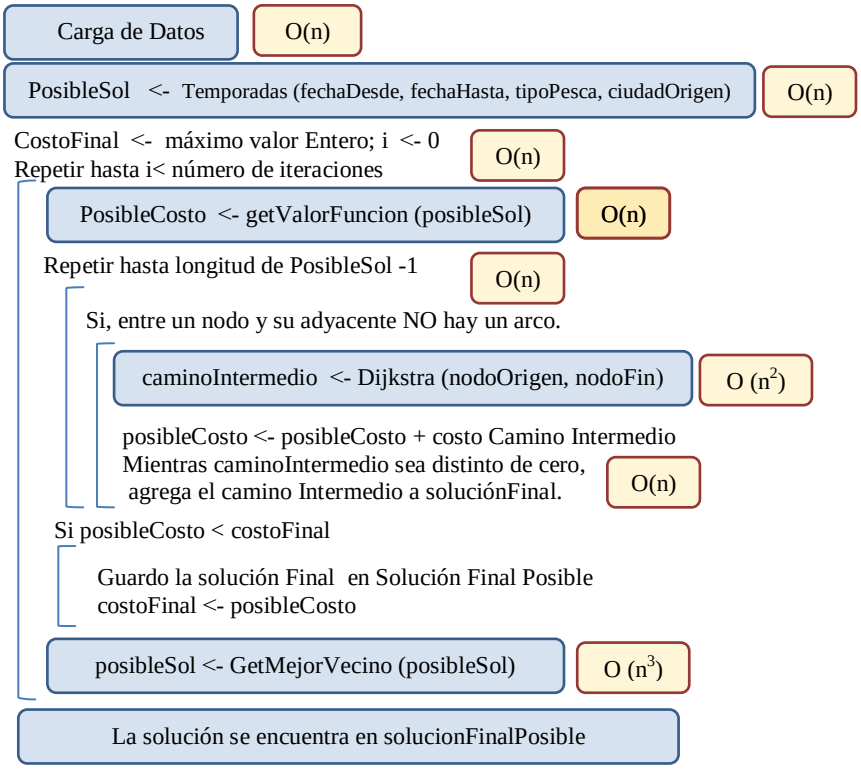


Figura 3. Diagrama de flujo y ordenes de ejecución del esquema principal de llamadas, *main*.

8 Conclusión

Con este algoritmo resultante se ha comprobado que las heurísticas pueden encontrar, en un tiempo razonable, soluciones aproximadas a la óptima, para la aplicación del TSP a partir de un mapa, los accesos de pesca, una fecha y un tipo de pesca. El método de *búsqueda tabú* se mostró efectivo para este problema

Se modeló el problema en el ámbito de los grafos. Donde los accesos de pesca son los nodos, y las rutas que unen a cada uno de ellos las aristas.

En primer lugar se desarrolló un algoritmo que genere como resultado un arreglo conteniendo aquellos nodos que cumplan con la condición que el usuario ingrese, tipo de pesca, fechas y ciudad origen. Este arreglo es una solución inicial del problema.

En segundo lugar, dada la magnitud y el alcance del problema y dado que su complejidad es *NP-Difícil*, se construyó un algoritmo *metaheurístico de búsqueda tabú* con mejoras iterativas.

Este algoritmo es una mejora sobre la búsqueda local básica, que intentó superar los problemas de la búsqueda local al no quedar atrapados en un mínimo local. Esto permitió contemplar más resultados, intercambiarlos y compararlos. Es por ello que acepta movimientos que no mejoran.

Se desarrolló entonces una heurística que demostró dar muy buenos resultados en muy corto tiempo. Pero el inconveniente inmediato presentado se relacionó con la inexistencia de un camino o arista que conecte dos nodos consecutivos del camino resultante. Esto hizo que el camino sea impracticable en la vida real. Se resolvió este inconveniente mediante la utilización de un algoritmo de *Dijkstra*, el cual devuelve el camino más corto entre dos nodos distintos.

Los resultados finales se ven claramente afectados por la solución inicial ingresada. Si la solución inicial es mala, su resultado final no será el óptimo, debido a que no se llega a realizar todos los intercambios necesarios para llegar a una solución cercana a la óptima. En lo que respecta a los requerimientos, mediante las heurísticas desarrolladas, se ha logrado cumplir con el más importante, la realización de un circuito de pesca óptimo en base a las restricciones planteadas en un comienzo. Los resultados que se obtuvieron son simples, claros y se asemejan al óptimo. Incluyendo ciudades que solo se usarán para conectar con otras que sí serán accesos de pesca, y no solo de paso. Concluyendo, el algoritmo propuesto se encuadra dentro de los Heurísticos Iterativos, y su complejidad es de $O(n^4)$, polinomial.

Referencias

1. Christofides, N. (1976), "*Worst-case analysis of a new heuristic for the travelling*
2. E.Pesch, F.Glover, Local Search and metaheuristics, in: G. Gutin, A. Punnen (Eds). The Traveling Salesman Problem and Its Variations, Kluwer Academic Publishers, Dordrecht, 2002, pp 309-368.
3. El método de Búsqueda Tabú: http://tesis.uson.mx/digital/tesis/docs/18920/Capitulo2.pdf –Consultado el 10/03/1017
4. Glover, F.; Melián, B. (2003) "Busqueda Tabú", *Revista Iberoamericanda de Inteligencia Artificial* **19:** pp 29-48
5. Johnson, D.S. and McGeoch, L.A.. "*The traveling salesman problem: A case study in local optimization*", Local search in combinatorial optimization, 1997.
6. María Lorena Stockdale .Tesis de Licenciatura - *El problema del viajante: un algoritmo heurístico y una aplicación*.
7. Miguel Sánchez García- *Las Matemáticas del Siglo XX. Optimización Combinatoria*.
8. Rego, César; Gamboa, Dorabela; Glover, Fred; Osterman, Colin. *Traveling salesman problem heuristics: Leading methods, implementations and latest advances* - European Journal of Operational Research. Volume 211, Issue 3, 16 June 2011, Pp 427–441

Interacción de elementos de networking con contenedores de software (docker)

Carlos Binker[1], Hugo Tantignone[1], Guillermo Buranits[1], Rubén Darío Moreira[1], Eliseo Zurdo[1]

[1]Universidad Nacional de La Matanza, Florencio Varela 1903 (B1754JEC) -- San Justo, Buenos Aires, Argentina
{cbinker, htantignone, gburanits, rmoreira}@unlam.edu.ar; ezurdo@alumno.unlam.edu.ar

Abstract. En este trabajo se pretende mostrar la interacción de elementos de networking tales como switches de capa 2 y de capa 3 interactuando con hosts con diversos sistemas operativos, los cuales son emulados a partir de contenedores de software a través del empleo de docker. Se emularán terminales de consola linux haciendo uso del intérprete de comandos *bash*. También se empleará un servidor tomcat y un terminal de cliente gráfico Firefox construidos con docker. Se empleará la plataforma GNS3 (Graphical Simulator Network 3) Dicha plataforma albergará toda la configuración de los dispositivos enumerados anteriormente en forma virtualizada (Appliances) [1]. Después de hacer un análisis más pormenorizado del GNS3 y su máquina virtual asociada se analizará un caso de estudio en donde se simula una red corporativa que conecta varias sucursales entre sí mediante enlaces punto a punto.

Keywords: GNS3, VMware, Dynamips, Appliances, docker

1 Introducción

El enorme crecimiento registrado en el ámbito del networking, teniendo en cuenta el avance vertiginoso de Internet, ha traído aparejado un importante incremento de las estructuras de hardware. Por esta razón y dada la necesidad de reducir físicamente estas enormes estructuras es que ha nacido un nuevo paradigma denominado *Virtualización* [2]. Clásicamente si por ejemplo se requería montar una granja de servidores, era necesario contar con muchas cajas físicas en donde cada caja constituía un servidor brindando una determinada aplicación. Con la virtualización, una sola caja física podría albergar múltiples servidores *lógicos* con la consiguiente disminución de las estructuras de hardware. Pero claro, esto no resulta gratuito, se requieren procesadores con alta potencia de cómputo (varios núcleos para procesamiento paralelo) y mucha capacidad de memoria RAM, como así también gran capacidad de almacenamiento de disco rígido. De esta manera nace el concepto de *Virtual Machine*, en castellano máquina virtual (en adelante VM), es decir una estructura lógica que emula los componentes básicos del hardware (disco duro, CPU, RAM, placas de red, lector de DVD, audio,

video, puertos USB, etc.) y que ejecuta un determinado sistema operativo. La VM es en esencia software que utiliza los recursos de hardware de una máquina principal denominada *host*. Cuando se instala una VM en una máquina host, para el usuario ésta se ve como si fuera una máquina física más con las mismas características de una máquina host, en donde los recursos a asignar a la VM son totalmente configurables. La VM instalada en una máquina host se denomina comúnmente *guest*, y el sistema operativo (en adelante SO) asociado también se denomina SO *guest*.

1.1 Descripción de la tecnología docker

Los containers [3] son una abstracción en la capa de aplicación que agrupa código y dependencias juntos. Múltiples contenedores pueden ejecutarse en una misma máquina y compartir el núcleo del sistema operativo (kernel linux preferentemente aunque también puede ser Windows) con otros contenedores, además de archivos binarios y bibliotecas. Cada uno de estos contenedores se ejecuta como un proceso aislado en el espacio de usuario. Los contenedores ocupan menos espacio que las VM (las imágenes de contenedores suelen tener decenas de MB de tamaño) y comienzan su procesamiento casi de manera instantánea. El hypervisor es la interface gráfica que permite que múltiples VM se ejecuten en un solo host físico. Cada VM incluye una copia completa de un SO, una o más aplicaciones, binarios y bibliotecas necesarias, ocupando decenas de GB. Las VM suelen ser lentas para arrancar. Esto es una diferencia sustancial con los contenedores. Los contenedores y las VM utilizados conjuntamente proporcionan una gran flexibilidad en la implementación y administración de aplicaciones (ver Fig. 1).

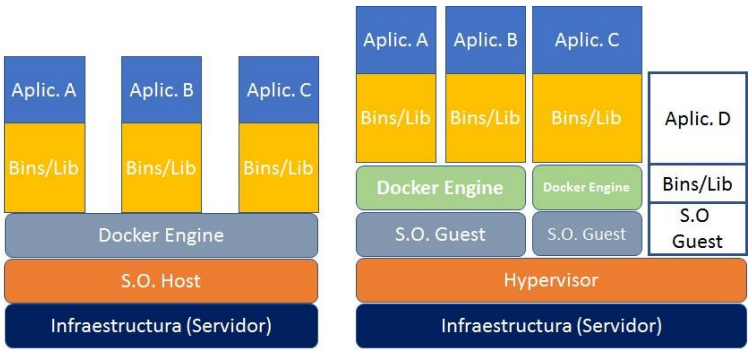


Fig. 1. Análisis comparativo entre VM y docker y combinación de VM con docker

1.2 Empleo del espacio de usuario para el lanzamiento de contenedores

Según la Fig. 2 ejecutar docker significa ejecutar dos programas en el espacio de usuario. El primero es el *docker daemon*, que se trata de un proceso que debe estar siempre ejecutándose. El segundo es el docker CLI. Este es el programa *docker engine* que interactúa con los usuarios ejecutando las aplicaciones a partir de los recursos del kernel. Si se desea iniciar, detener o instalar software, se emitirá un comando utilizando

el programa docker. La Fig. 2 también muestra tres contenedores en ejecución. Cada uno se ejecuta como un proceso hijo del demonio docker, encapsulado dentro de un contenedor, y el proceso delegado se ejecuta en su propio subespacio de memoria del espacio de usuario. Los programas que se ejecutan dentro de un contenedor pueden tener acceso solamente a su propio subespacio de memoria y acceso a los recursos según el alcance del contenedor.

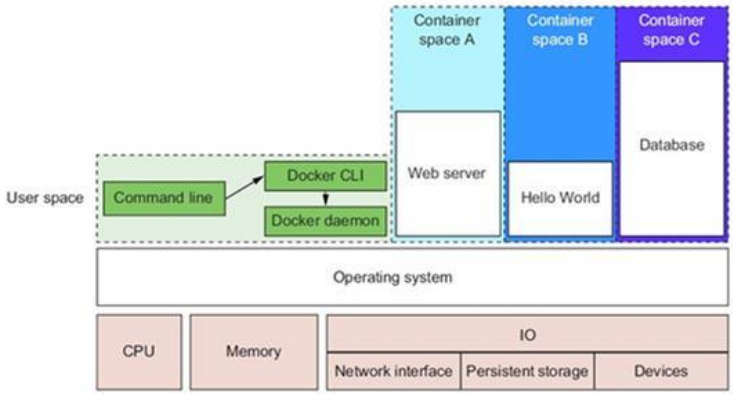


Fig. 2. Empleo del espacio de usuario para el lanzamiento de programas

1.3 Problemática de las aplicaciones con dependencias compartidas

Sin docker, una PC parece un caos en cuanto a las aplicaciones se refiere. Las aplicaciones tienen todo tipo de dependencias. Algunas dependen de bibliotecas de sistemas para aspectos tales como sonido, redes, gráficos, etc. Otras, en cambio dependen de bibliotecas estándar para el idioma en el que están escritas. Algunas dependen de otras aplicaciones, como la forma en que un programa Java depende de su versión específica. La VM o una aplicación web puede depender de una base de datos, etc. Es común que un programa requiera acceso exclusivo a algún recurso escaso tal como una red, conexión o un archivo. Sin docker, las aplicaciones se distribuyen por todo el sistema de archivos y terminan creando una confusa red de interacciones. La Fig. 3 ilustra cómo diferentes aplicaciones dependen de distintas bibliotecas comunes presentes en el sistema. Por ende, docker mantiene las cosas organizadas aislando todas estas dependencias en contenedores de software que incluyen todos los elementos necesarios, básicamente binarios y librerías, para evitar el caos de la Fig. 3. Ver Fig.4

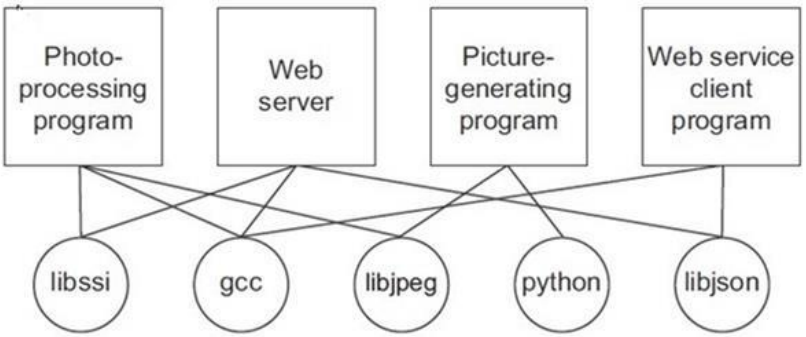


Fig. 3. Ejemplo de aplicaciones con dependencias compartidas

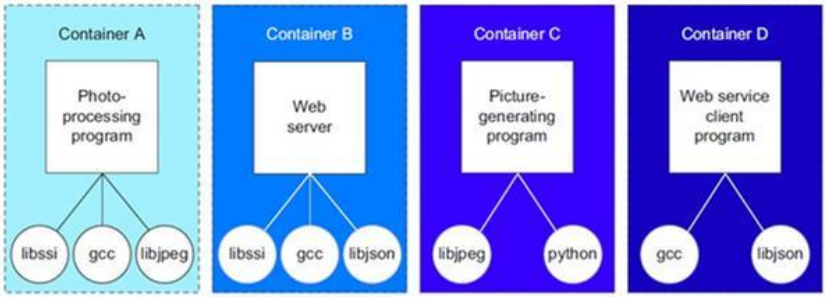


Fig. 4. Programas corriendo dentro de un contenedor con copia de sus dependencias

2 Descripción sucinta de la Plataforma de simulación GNS3

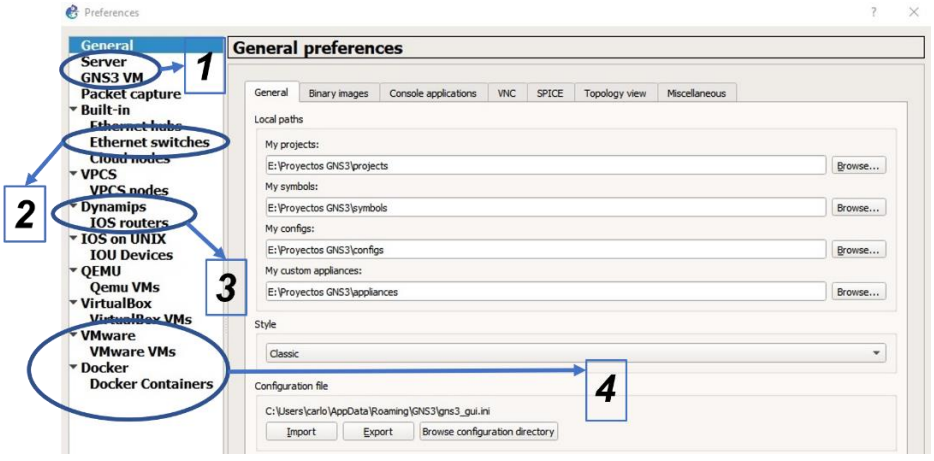


Fig. 5. Aspectos relevantes de la configuración de la plataforma GNS3

En la Fig. 5 en la configuración ***1*** (Server – GNS3 VM) se fijan la habilitación del servidor local (host donde reside el GNS3) y la habilitación de la VM que albergará

todas las configuraciones de routers, switches, VM y contenedores docker. Además se elige el tipo de Hypervisor, en este paper se empleará Vmware Workstation. En la configuración ***2*** (Ethernet Switches) se configura el switch de capa 3 (L3). Se emplea para ello un router Cisco C3745 con un módulo NM-16ESW que es un switch de 16 puertos que implementa VLAN. Este switch se emplea como switch de *core* de la topología en estudio. Como switches de acceso se emplean los Ethernet Switches incluidos en la misma plataforma. En realidad el switch L3 es un router que al agregarle el módulo mencionado lo convertimos en un switch de capa 3, para ello es necesario marcarlo como *Etherswitch router*. En la configuración ***3*** el router 3745 se simula mediante Dynamips, que permite emular el SO iOS (Internet Working Operating System) de cisco. En el C3745 se ha agregado también un módulo WIC 2T que contiene dos interfaces de WAN para establecer los enlaces punto a punto correspondientes. En consecuencia se han generado dos templates, uno como router convencional, y el otro como un switch de capa 3. Finalmente en la configuración ***4*** se configuran los containers en base a las imágenes bajadas del docker Hub [4].

3 Diseño de un caso de estudio empleando GNS3 con docker

Se hará la síntesis de la siguiente topología de red:

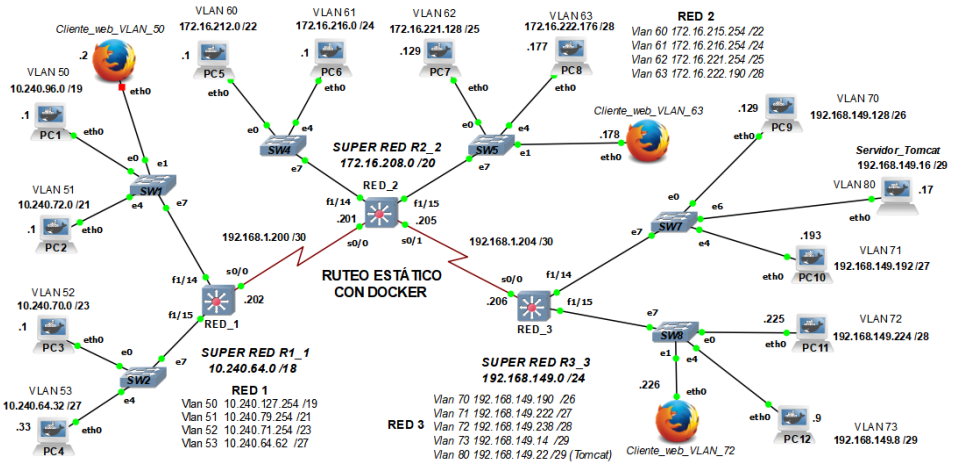


Fig. 6. Topología de red a implementar mediante GNS3

La topología consiste en tres redes principales indicadas como RED_1, RED_2 y RED_3 respectivamente. Cada una de estas redes posee cuatro vlans (a excepción de la RED_3 que incorpora la vlan 80 para el servidor Tomcat, que está implementado por docker) cuyas direcciones de red y máscaras se observan en la Fig. 6. En los switches de acceso (SW1, SW2, SW4, SW5, SW7 y SW8) se han asociado las *vlan* a los puertos del switch y se han configurado los *trunks* mediante el estándar 802.1q. Igual situación se ha hecho con los switches L3 en donde además se ha configurado el ruteo de las vlans. Además las vlans en cada red principal se han sumarizado a fin de minimizar el tamaño de las tablas de enrutamiento, resultando la superred ***10.240.64.0/18*** para la RED_1, la superred ***172.16.208.0/20*** para la red RED_2 y la superred ***192.168.149.0/24***

para la red RED_3. El ruteo global de toda la topología se lleva a cabo mediante rutas estáticas, ya que el hecho de sumarizar las vlans implica tan sólo tres sentencias de configuración en cada switch de capa 3 con la consiguiente performance que el ruteo estático otorga. A continuación se muestran las tablas correspondientes a las redes y a las superredes de la topología en cuestión.

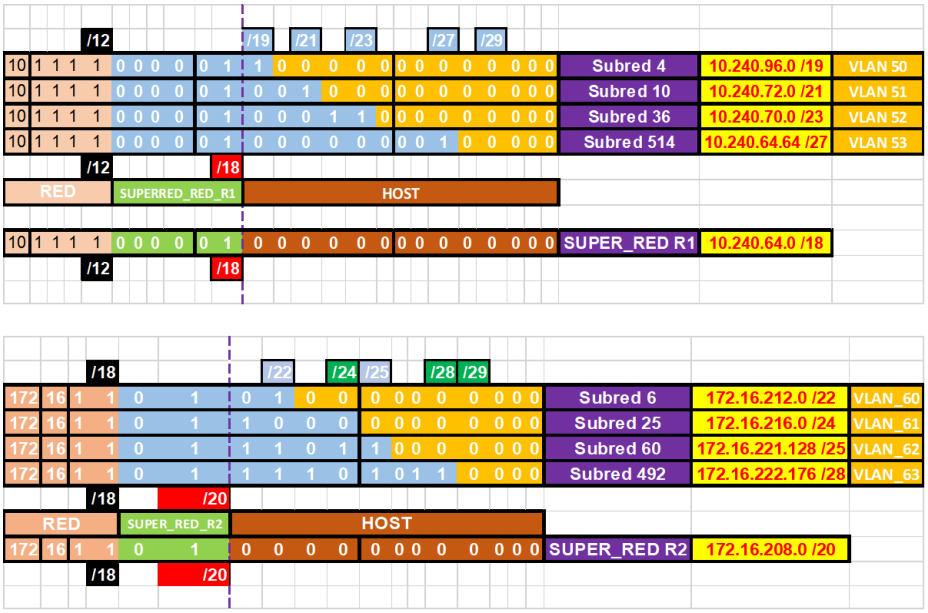


Tabla 1. Redes y superredes correspondientes a los switches RED_1 y RED_2

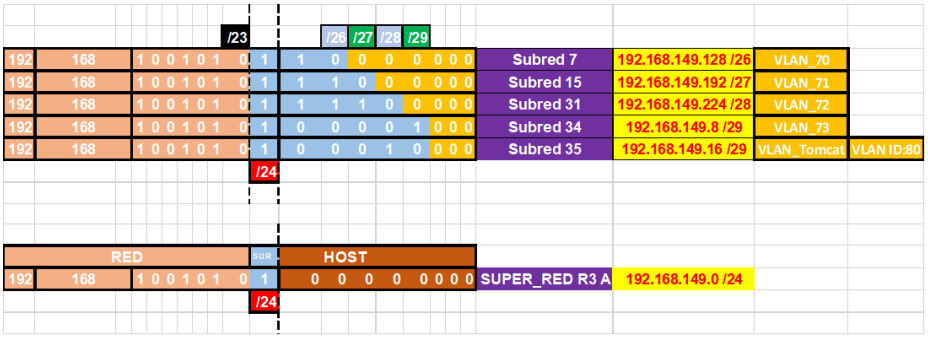


Table 2. Red y superred correspondiente al switches RED_3

4 Desarrollo de la experiencia

En este apartado se aportará la experiencia de vislumbrar la interacción del mundo del networking con las computadoras y los servidores simulados mediante *docker* . Para

ello primero haremos una síntesis preliminar de las herramientas utilizadas y finalmente presentaremos las capturas más relevantes de los resultados obtenidos.

4.1 Herramientas utilizadas

Para la simulación de la topología de red mencionada con anterioridad emplearemos los siguientes elementos de hardware y software:

- HP Notebook Pavilion AMD A12-9720P RADEON R7, 12 COMPUTE CORES 4C + 8G 2,70 GHZ con 16 GB de RAM. Windows 10 home de 64 bits
- Plataforma de software GNS3 versión 2.1.8
- VMware Workstation 14 Pro version 14.1.2 build-8497320
- 3 routers C3745 configurados como Etherswitch con una interfaz WIC-2T
- 12 containers emulando una consola bash Ubuntu server 14.04 (ubuntu:14.04)
- Servidor tomcat (imagen latest obtenida en Docker Hub https://hub.docker.com/)
- 3 containers emulando un Terminal web (browser Firefox). La imagen es gns3/webterm:latest https://hub.docker.com/r/gns3/webterm/, en Docker Hub)
- 6 switches de acceso (SW1, SW2, SW4, SW5, SW7 y SW8) embebidos en GNS3

4.2 Configuraciones de los dispositivos involucrados en la topología de red

```
gns3@gns3vm:~$ docker images
REPOSITORY          TAG                 IMAGE ID            CREATED
SIZE
tomcat              latest              df50c9d355cf        4 days ago
463 MB
ubuntu              14.04               578c3e61a98c        3 weeks ago
223 MB
gns3/webterm        latest              d504c725cf86        4 weeks ago
389 MB
gns3@gns3vm:~$ _
```

Fig. 7. Imágenes docker con sus hashes correspondientes empleadas en el caso de estudio

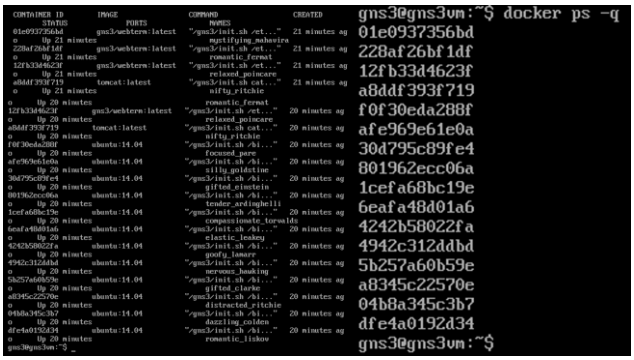


Fig. 8. Contenedores lanzados en docker. La parte izquierda se ve con ps -a y la derecha con ps -q

A fin de demostrar el correcto funcionamiento de la red, y por cuestión de espacio, ponemos la configuración del switch L3 que implementa la RED_3, tal como se ve en las Figuras 9, 10, 11 y 12 respectivamente.

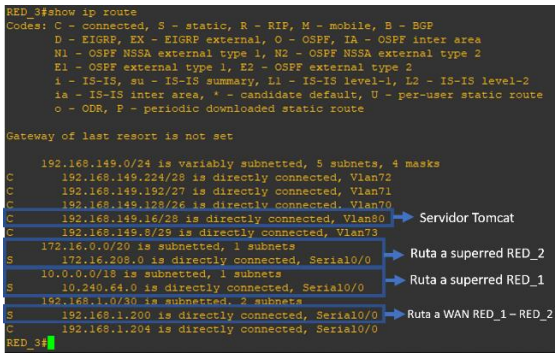


Fig. 9. Tabla de ruteo del switch L3 RED_3

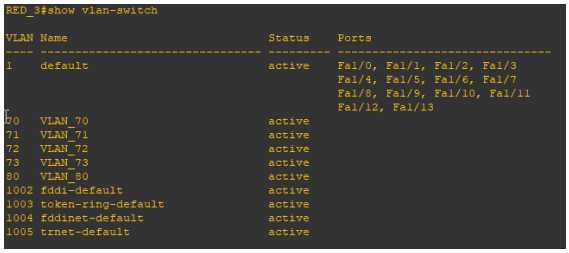


Fig. 10. Configuración de vlans en el switch L3 RED_3

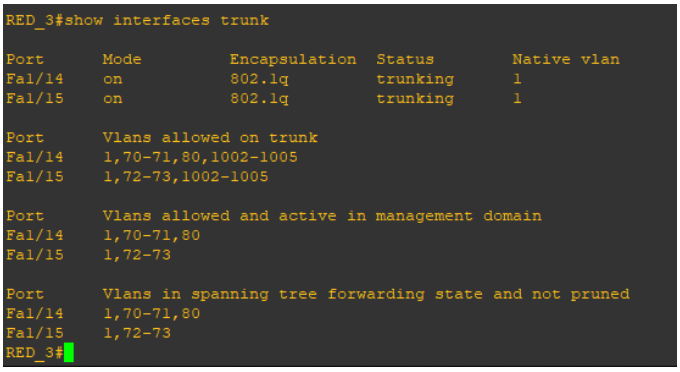


Fig. 11. Observación de los puertos de trunk en el switch L3 que implementa RED_3

```
interface FastEthernet1/14
 switchport trunk allowed vlan 1,70,71,80,1002-1005
 switchport mode trunk
!
interface FastEthernet1/15
 switchport trunk allowed vlan 1,72,73,1002-1005
 switchport mode trunk
!
interface Vlan1
 no ip address
!
interface Vlan70
 ip address 192.168.149.190 255.255.255.192
!
interface Vlan71
 ip address 192.168.149.222 255.255.255.224
!
interface Vlan72
 ip address 192.168.149.238 255.255.255.240
!
interface Vlan73
 ip address 192.168.149.14 255.255.255.248
!
interface Vlan80
 ip address 192.168.149.22 255.255.255.240
!
!
ip forward-protocol nd
ip route 10.240.64.0 255.255.192.0 Serial0/0
ip route 172.16.208.0 255.255.240.0 Serial0/0
ip route 192.168.1.200 255.255.255.252 Serial0/0
```

Fig. 12. Direcciones de red de vlans y rutas estáticas a las superredes y WAN RED_1 – RED_2

4.3 Resultados alcanzados

A continuación, mostraremos tres capturas realizadas desde cada container en cada RED enviando ping hacia los otros containers en otras redes. Finalmente mostraremos la captura desde un terminal web hacia el servidor tomcat. Con esto se corrobora el funcionamiento del laboratorio.

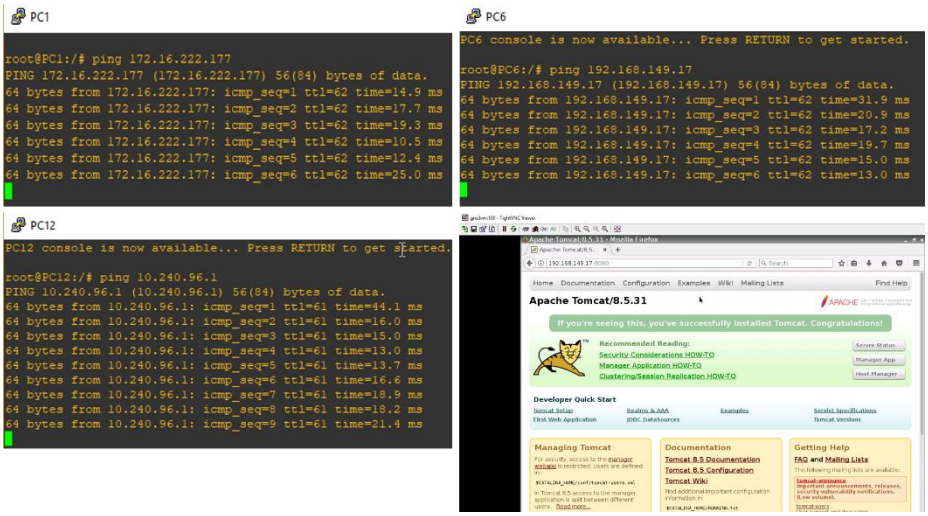


Fig. 13. Prueba de ping entre diferentes contenedores y desde un terminal web al *TOMCAT*

5 Conclusiones y trabajo futuro

1. Ahora en GNS3 es posible conectar a la infraestructura de networking VMs más contenedores de software. Los containers a diferencia de las VMs se ejecutan muy rápido y ocupan algunos MB frente a los GB de las VMs.
2. En una futura investigación se estudiará docker-machine. Esto permite instalar en una VM un conjunto de VMs, donde en cada una de ellas se instala docker y pueden ejecutarse aplicaciones distribuidas entre estas VMs además de administrarse a todas estas VMs de manera centralizada.
3. En relación con docker-machine y a GNS3 se trabajará en un futuro con redes en IPv6, dando de esta manera conectividad dual stack.
4. Los requerimientos de hardware al trabajar con contenedores de software resultan más flexibles que al trabajar con máquinas virtuales, en donde es necesario gran capacidad de RAM, de disco y de procesamiento.

6 Referencias

1. Appliances. Emulación de elementos de networking interactuando con máquinas virtuales. Carlos Binker. CACIC 2016.
2. Virtualization Essentials 2nd Edition. Matthew Portnoy. Publicación: 29 de agosto de 2016. Editorial Sybex. ISBN 10: 1119267722.
3. Virtual Machines: Versatile Platforms for Systems and Processes (The Morgan Kaufmann Series in Computer Architecture and Design) 1st Edition. Publicación:17/062005. Editorial: Morgan Kaufmann. ISBN 10: 1558609105.
4. Docker Up and Running. Karl Matthias & Sean P. Kane. Publicación: 03/07/2015. Editorial: O'Reilly Media. ISBN 10: 1491917571. Docker in Action. Jeff Nickoloff. 1st Edition. Publicación: 27/03/16. Editorial: Manning Publications. ISBN 10: 1633430235

Mejoras en maquinaria industrial con IoT. Hacia la Industria 4.0

Luis Damian Candia[1], Anahí S. Rodríguez[1], Nestor Castro[1], Patricia Bazán[1], Viviana M. Ambrosi[1, 2], Javier Díaz[1]

1- LINTI – Facultad de Informática – UNLP
2- Comisión de Investigaciones Científicas de la Provincia de Buenos Aires - CIC
dcandia@linti.unlp.edu.ar, arodriguez@linti.unlp.edu.ar, ncastro@isis.unlp.edu.ar, pbaz@info.unlp.edu.ar , javier.diaz@info.unlp.edu.ar, vambrosi@info.unlp.edu.ar

Resumen. Se calcula que en el año 2025 existirán más de 26 billones de dispositivos conectados incluyendo casas, teléfonos, autos y fábricas. Esta era de hiperconectividad se conoce como "La era de Internet de las Cosas (IoT – Internet of Things)", la cual impone grandes desafíos y conduce a las industrias a adaptar sus procesos de producción a la nueva realidad. La aplicación de este nuevo concepto en la industria es conocida como Industria 4.0 o la nueva revolución industrial. La misma consiste en la digitalización de los procesos industriales por medio de la interacción de la inteligencia artificial con las máquinas y la optimización de recursos enfocada en la creación de efectivas metodologías comerciales. Esto implica cambios orientados a infraestructuras inteligentes y a la digitalización de metodologías. Este trabajo propone la actualización de una maquinaria industrial mediante la incorporación de sensores que alimentan sistemas informáticos de control, mediante una arquitectura por capas.

Palabras Clave: Industria 4.0, IoT, I2oT, Servicios web.

1 Introducción

Nos encontramos en los inicios de una nueva revolución industrial, que cambiará para siempre los procesos de negocio de las fábricas, esta nueva revolución es conocida como "*Industria 4.0"*. El término se refiere a un nuevo modelo de organización y de control de la cadena de valor a través del ciclo de vida del producto apoyado por las tecnologías de la información. En definitiva se trata de la aplicación a la industria del modelo "Internet de las cosas" (IoT) [1].

En este contexto, en octubre de 2017 la Secretaría de la Transformación Productiva del Ministerio de Producción de Argentina, realizó una encuesta online a 78 ejecutivos de 66 grandes empresas industriales argentinas de diferentes sectores económicos. El objetivo era conocer el progreso en la implementación de nuevas tecnologías en sus procesos productivos y los desafíos más importantes que encuentran a la hora de ejecutar esta transformación y el resultado fue que solamente el 30 % de la industria de nuestro país había implementado alguna solución referida con la Industria 4.0 [2].

Este resultado demuestra la necesidad de pensar soluciones que sean económicas y de rápida implementación para que nuestra industria nacional no pierda competitividad. Un enfoque posible es utilizando herramientas de hardware y software libre para realizar el despliegue de soluciones de IoT e integrarlas con las máquinas y los sistemas ya existentes en las industrias, de modo de alcanzar la cobertura total de la actividad de la organización. Es necesario adaptar los elementos con los que ya cuenta la industria a las nuevas tecnologías; esto permite una reutilización que no solamente favorece económicamente a las empresas, sino que también reduce el impacto de descarte de las mismas

El objetivo del presente trabajo es mostrar un caso de estudio en el cual se implementó una red de sensores de IoT con un backend de administración basado en web en una fábrica de bolsas de plástico, adaptando con nuevas tecnologías toda la maquinaria de la planta para poder medir en tiempo real la producción de la misma.

El artículo se organiza de la siguiente manera: en la Sección 2 se muestra el estado del arte y antecedentes acerca del concepto de Industria 4.0. En la Sección 3 se presenta la propuesta arquitectónica para abordar soluciones basadas en Industria 4.0. En la Sección 4 se aplica a un caso de estudio en la industria del plástico. Finalmente se arriban a algunas conclusiones y trabajos futuros.

2 Estado del Arte de la Industria 4.0

Los procesos industriales han atravesado tres grandes transformaciones tecnológicas a lo largo de la historia, las primeras revoluciones industriales como fueron la creación de la máquina de vapor, la electricidad y la automatización causaron un profundo impacto en la sociedad luego de su implementación. La cuarta revolución industrial llega de la mano de la tecnología.

A diferencia de las tres revoluciones industriales anteriores, la cuarta revolución industrial se generó primero en la sociedad como un cambio cultural y en la actualidad la industria intenta alcanzar el mismo nivel de tecnología con el cual viven las personas en sus hogares. El concepto de “Industria 4.0” fue nombrado por primera vez en Alemania en el año 2010. Esta nueva ola de tecnologías, que hoy avanza en diversas regiones del mundo a distintas velocidades, implica una transformación a partir de nuevas tecnologías industriales con foco digital, con potencial de crear fábricas con procesos productivos totalmente integrados y automatizados, permitiendo que diferentes sistemas interactúen analizando información en tiempo real para optimizar la producción, predecir fallas e integrar las cadenas de suministros para volverlas más eficientes. El impacto en los procesos productivos será significativo, volviéndolos más flexibles, eficientes y veloces, con importantes implicancias para los nuevos empleos que demandarán las empresas [3].

Se pueden identificar tecnologías clave de esta transformación, descritas a continuación:

- Robótica: robots industriales autónomos y cooperativos.
- Internet industrial: red de máquinas y productos con comunicación multidireccional entre objetos de la red.
- Simulación: simulación y optimización de redes de valor basada en información en tiempo real procesada por sistemas integrados.

- Cloud y ciberseguridad: administración de grandes volúmenes de datos en sistemas abiertos con alto nivel de interacción entre máquinas inteligentes, productos y sistemas que exige altos requisitos de seguridad.
- Manufactura aditiva: impresión 3D particularmente para prototipos y repuestos y para reducir envíos e inventario.
- Realidad aumentada: realidad aumentada para mantenimiento, logística y procesos operativos, Información de apoyo, ej.: mediante lentes.
- Big Data y analítica: análisis integral de la información disponible (ej.: ERP, SCM, CRM e información de robots productivos para apoyo a toma de decisiones en tiempo real.

La adopción de tecnología a pesar de parecer un gasto inicial, podría aumentar considerablemente la productividad y competitividad de una organización, haciendo más eficientes los procesos, generando nuevos negocios para las empresas y favoreciendo la creación de nuevos puestos de trabajo. Cada día más, los gobiernos diseñan e implementan instrumentos e incentivos para potenciar la economía del conocimiento y acompañar a las empresas en la adopción de nuevas tecnologías.

Argentina no es ajena a esta tendencia, empezó a plantear una estrategia integral de políticas públicas para que la adopción de tecnología cobre mayor protagonismo [4].

Sin embargo, incluir nuevas tecnologías depende, en gran parte, de la posibilidad de que todas las empresas del entramado productivo incorporen estos procesos de innovación en sus modelos de negocio, sin estos esfuerzos complementarios no existirán mejoras sustentables.

Industria 4.0 es un concepto que forma parte de la agenda de las empresas de todo el mundo y las compañías argentinas no son una excepción. Esta transformación involucra cambios en múltiples frentes y abarca diversos tópicos, siendo los más importantes aquellos relacionados con fábricas inteligentes, control de la producción y diseño de fábricas digitales, gestión del desempeño en tiempo real, tableros electrónicos de desempeño y mantenimiento predictivo.

3 Industria 4.0 en una solución integral por capas

El escenario que plantea la Industria 4.0 amerita el planteo de soluciones que consideren distintas capas de abstracción y un modelo arquitectónico orientado a servicios [5].

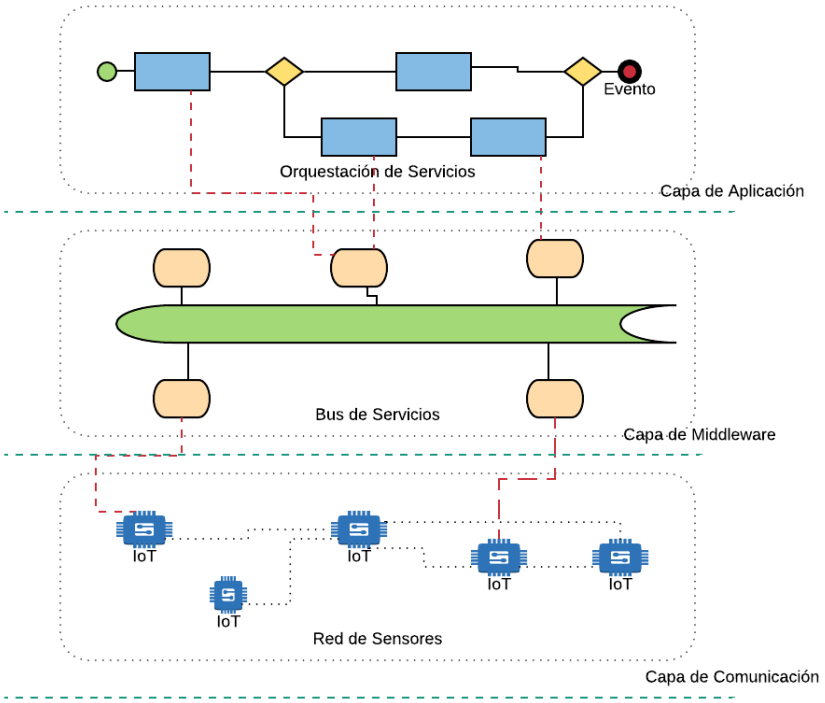


Figura 1 – Modelo Arquitectónico por capas

La Figura 1 plantea dicho modelo arquitectónico:

1. Capa de Comunicación: está conformada por una red de sensores que producen la información que alimenta a los sistemas de la organización. Esta capa posee sus propias reglas y protocolos de comunicación y establece un mecanismo uniforme para propagar sus registros hacia una capa de servicios.
2. Capa Middleware: está conformada por servicios o componentes, que son las piezas de software que se comunican con la red de sensores y procesan los datos obtenidos para disponerlos al usuario.
3. Capa de Aplicación: realiza la orquestación de los servicios y resuelve los mecanismos y circuitos propios del dominio conformando el flujo de trabajo (workflow).

Este trabajo propone una solución aplicada a un caso de estudio para la capa de sensores y la de servicios. Esta última implementada, en este ejemplo, por una aplicación Web tradicional.

Se propuso la implementación de una red de sensores que puedan tomar en tiempo real información de las máquinas y enviarlas a una aplicación Web que se encarga de almacenarlas en su base de datos, para que luego los administradores de la planta puedan observar en tiempo real cuantas unidades llevan completas cada orden de

trabajo y el tiempo en que estuvo productiva cada máquina para poder tomar mejores decisiones de negocio.

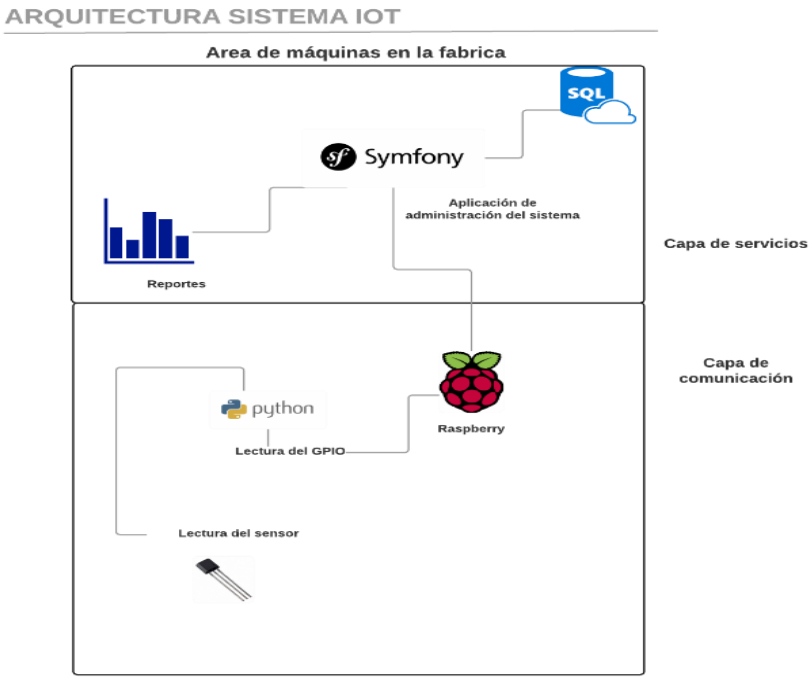


Figura 2. Solución en Fábrica de Bolsas de Plásticos – Tecnologías utilizadas

En la Figura 2 se muestran un gráfico con los elementos tecnológicos utilizados en el caso de estudio. En este caso, el middleware y la orquestación de servicios están representados por una aplicación Web tradicional.

Se describen los elementos de la Figura 2:

Raspberry PI 3 [6]: es una mini computadora del tamaño de una billetera, que a pesar de su tamaño reducido cuenta con una gran capacidad de procesamiento si la comparamos con otros dispositivos de IoT, cuenta con 1 GB de memoria RAM, un GPIO de 40 pines, 4 puertos USB, un puerto Ethernet y un módulo WIFI. Los beneficios propuestos por este tipo de placas es su gran capacidad de procesamiento, ya que además de sensar las confeccionadoras, el staff de la empresa desea a futuro poder agregarle pantallas táctiles que permitan al operario mediante un aplicativo escrito en PYTHON poder observar información de la producción y poder interactuar con el sistema de IoT en la zona de producción de la planta.

Sistema operativo Raspbedian [7]: es una versión del sistema operativo Debian optimizado para poder ejecutarse en las placas Raspberry PI. Cuenta con una intefaz de configuración completa para todos los parámetros configurables de la placa "raspi-conf".

Sensor de efecto Hall [8]: es un sensor con salida digital y retención de estado, que cambia de cero a uno ante la presencia de un campo electromagnético. Esto significa

que cada vez que un imán este dentro de su rango de sensado, el sensor cambiara su estado, de 0 a 1 o de 1 a 0 según su estado actual, reteniéndolo hasta que de nuevo un imán se encuentre dentro de su rango de sensado, al contrario de otros sensores que solo cambian su estado cuando el imán está cerca.

Symfony [9]: es un framework de programación web basado en PHP, es uno de los más utilizados en el mercado y reduce considerablemente los tiempos de desarrollo, brindando múltiples módulos de servicios reutilizables.

Python [10]: lenguaje de programación de propósito general, orientado a objetos, que también puede utilizarse para el desarrollo web.

4 Caso de estudio real – Implementación de red IoT en industria de plástico

La organización que se presenta en el caso de estudio es una fábrica de bolsas de plástico, la misma cuenta con 4 (cuatro) máquinas denominadas confeccionadoras, 2 (dos) extrusoras y 3 (tres) impresoras. Durante este trabajo nos centraremos en las máquinas confeccionadoras.

Durante la tarea de producción de la fábrica se ejecutan diferentes procesos, primero se prepara el material líquido por las máquinas extrusoras, logrando un material sólido, que luego atraviesa un conjunto de rodillos, dispuestos en impresoras, para imprimir los gráficos de las bolsas.

Las máquinas denominadas confeccionadoras se encargan de cortar por unidad las bolsas que se generaron en los procesos anteriores, para eso cuenta con una guillotina que automáticamente, según las indicaciones de medida que le ingresa el operario, realiza el corte de cada una de las unidades. Las máquinas cuentan con distintos conjuntos de rodillos por los cuales va a pasando el material.

El sector de administración crea órdenes de trabajo indicando las unidades que deben fabricarse, el tiempo estimado de trabajo y la velocidad de funcionamiento de la confeccionadora. Este proceso cuenta con el inconveniente de que la información de la producción actual, junto con el tiempo de utilización y detención de la máquina son ingresados manualmente por un operario de la planta en una planilla de papel, la que luego al finalizar su turno, es dejada en administración para su posterior procesamiento. Esto provoca que la disponibilidad de la información de producción de la planta sea lenta e insegura.

La propuesta de mejora que presenta este trabajo se centra en construir una red de sensores que realicen la lectura de la cantidad de cortes realizados por la confeccionadora.

El proceso de lectura de los sensores (Capa de Comunicación)

Como se mencionó, las confeccionadoras cuentan con una cuchilla con la cual generan el corte de cada unidad, por lo tanto se utilizó un sensor de efecto hall [8], que se activa al detectar un cambio en el campo magnético, en este caso sería cuando la cuchilla se acerca a la base de la máquina en donde está alojado el sensor para realizar un corte. En la Figura 3 se muestra el diagrama de conexión del sensor.

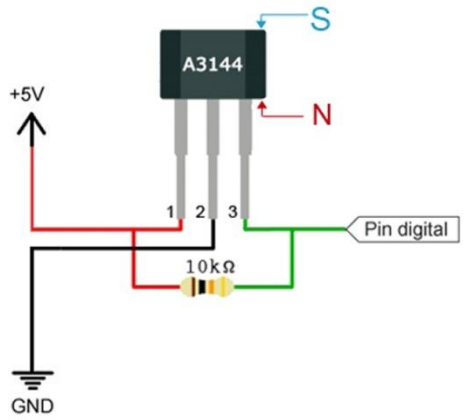


Figura 3. Conexión del sensor.

En la placa raspberry mediante codigo PYTHON se realiza la lectura del sensor, en el GPIO correspondiente al pin digital del circuito que es por donde llegara la lectura del sensor, la información de que se registro una nueva lectura es encapsulada y enviada al servidor mediante API REST, a continuación se muestra el codigo PYTHON.

```
import RPi.GPIO as GPIO
import time
import urllib
GPIO.setmode(GPIO.BCM)
GPIO.setup(23, GPIO.IN, pull_up_down=GPIO.PUD_DOWN)

print 'Sensores iniciado'


GPIO.add_event_detect(23, GPIO.BOTH)
def my_callback(dd):
    print 'Cambio en el sensor'
    url = 'http://10.0.1.2/app_dev.php/machine_uno_encendida'
    u = urllib.urlopen(url)
    data = u.read()

GPIO.add_event_callback(23, my_callback)

while True:
     time.sleep(0.5)
```

Aplicación web de administración (Midlleware y Orquestación)

El sistema web propuesto en el presente trabajo se desarrolló con Symfony [9] para administrar las órdenes de trabajo y observar la productividad generada por las máquinas confeccionadoras. La información obtenida en tiempo real del estado de producción de la fábrica, permite optimizar los recursos y el tiempo de trabajo, pudiendo balancear de forma correcta las órdenes entre las distintas máquinas confeccionadoras, según su tiempo ocioso.

A continuación mostraremos los módulos relacionados con la lectura de los sensores.

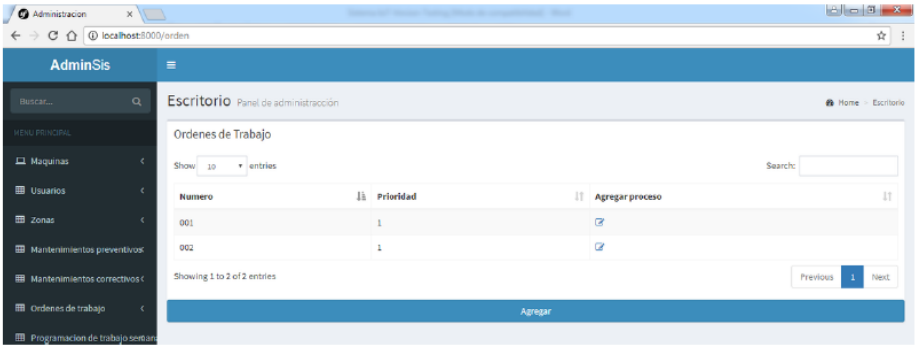


Figura 4. Listado de ordenes de trabajo

En la Figura 4 se observa el módulo donde se pueden crear nuevas órdenes de trabajo, indicando el tiempo teórico de ejecución, la cantidad de unidades y en que máquina confeccionadora se va a ejecutar, también se puede observar el listado de órdenes de trabajo creadas, editarlas o eliminarlas.

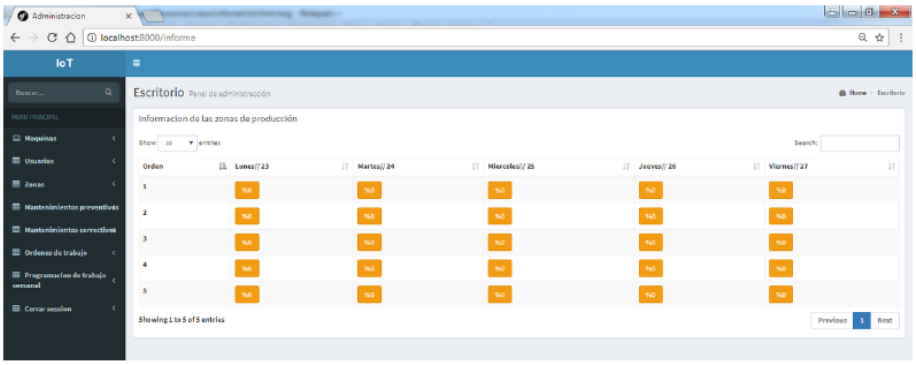


Figura 5. Proceso de ejecución por dia de cada orden de trabajo

En la Figura 5 se observa el proceso de ejecución por día de cada orden de trabajo, en la cual se muestra el crecimiento de la ejecución de las órdenes de trabajo dividido por día, de esta manera poder tener un histórico de la información registrada.



Figura 6. Porcentaje de ejecución de una orden de trabajo determinada

En la Figura 6 al ingresar en una orden de trabajo se observa el porcentaje de ejecución de la misma, pudiendo también detener la lectura de los sensores.

4 Conclusiones y Trabajos Futuros

Es necesario pensar soluciones que permitan la innovación tecnológica dentro de la industria a un bajo costo, esta innovación es indispensable para lograr que la industria nacional pueda ser competitiva con la de otros países más evolucionados tecnológicamente hablando.

Las empresas tienen que considerar a la tecnología de IoT como un gran aliado estratégico para poder fortalecer su crecimiento.

El futuro de IoT dependerá de cómo sus componentes puedan interactuar con los procesos de negocio o con las soluciones de IT que se ejecuten a su alrededor [11]. Una costosa red de sensores carece de utilidad si no se construyen soluciones de software que consuman las señales. Si ese software está conformado por componentes orquestados como procesos de negocio, es natural pensar que los elementos de IoT deban ser tenidos en cuenta desde la fase de modelado de dichos procesos de negocio [12].

En futuros trabajos se plantea cambiar en la arquitectura planteada la aplicación web por un motor de procesos de negocio, logrando la completa integración entre todas las capas de la arquitectura, también se plantea modificar el protocolo de comunicación API REST entre las placas de IoT y el modulo Web por un protocolo de comunicación más liviano como es MTTQ para evitar posibles congestiones en la red.

5 Referencias

1- Okano, M. T. (2017). IOT and Industry 4.0: The Industrial New Revolution. In International Conference on Management and Information Systems September (Vol. 25, p. 26).

2- http://image-src.bcg.com/Images/Acelerando-el-Desarrollo-de-Industria-40-en-Argentina_tcm62-184622.pdf.
3- https://www.aimplas.es/sites/default/files/optiman_entregable.pdf.
4- http://www.mincyt.gob.ar/adjuntos/archivos/000/038/0000038319.pdf.
5- Bazán, P. (2009). Un modelo de integrabilidad con SOA y BPM (Doctoral dissertation, Cap. 4, Facultad de Informática.
6- Raspberry PI 3, https://www.raspberrypi.org/products/raspberry-pi-3-model-b/ .
7- Sistema operativo Raspbedian, https://www.raspberrypi.org/downloads/raspbian/ .
8- Sensor de efecto Hall, https://www.ucm.es/data/cont/docs/76-2015-03-19-Efecto-Hall-Final.pdf.
9- Symfony, http://symfony.com/doc/current/index.html.
10- Python, http://www.python.org.ar/.
11- Meyer, S., Sperner, K., Magerkurth, C., & Pasquier, J. (2011, June). Towards modeling real-world aware business processes. In Proceedings of the Second International Workshop on Web of Things (p. 8). ACM.
12- Weske, M. (2012). Business process management architectures. In Business Process Management (pp. 333-371). Springer, Berlin, Heidelberg.

Un análisis de comportamiento entre distintos mecanismos de control de congestión ensayados sobre una topología mixta

Diego R. Rodríguez Herlein, Carlos A. Talay, Claudia N. González, Franco A. Trinidad y Luz Almada
Licenciatura en Sistemas, UNPA-UARG
Río Gallegos, Argentina
Campus universitario - Oficina B 18
{dherlein, ctalay, cgonzalez}@uarg.unpa.edu.ar, tfrancoalejandro@gmail.com, mluzalmada@gmail.com

Luis A. Marrone
L.I.N.T.I. – Universidad Nacional de La Plata
La Plata, Argentina
Calle 50 y 120 – 2 do. Piso – Edificio Bosque Oeste
lmarrone@linti.unlp.edu.ar

Resumen. En este trabajo se realiza una comparación de competencia entre distintos tipos de control de congestión desarrollados en algunas variantes del protocolo TCP. Para realizar esta comparación se recurrió a pruebas implementadas en una topología híbrida simple, en donde dos flujos de datos, que compiten por los recursos de la red, tienen como destino final un nodo inalámbrico. Para el desarrollo de las pruebas se utilizó el simulador de redes NS2, con el cual se realizaron todas las simulaciones, exponiéndose en el documento los resultados provenientes del análisis de los datos obtenidos.

Palabras clave: TCP; control de congestión; redes híbridas

1 Introducción

Prácticamente la totalidad de las aplicaciones utilizadas en Internet utilizan Transmission Control Protocol (TCP) [1] para el envío de datos en forma fiable a través de la red. Uno de los aspectos más importantes de TCP es el control de congestión, que si bien no formó parte de la implementación original, define, en gran medida, el rendimiento, efecto que se potencia en las redes que incluyen enlaces inalámbricos. Mediante el control de congestión [2], TCP regula la inyección de datos a la red utilizando una ventana de congestión, cuyo tamaño dependerá del estado de congestión.

La carencia de algún tipo de control de congestión en el estándar TCP original, llevo a lo que se denominó el colapso de congestión [3]. Las variantes del protocolo TCP implementan distintas estrategias para realizar este control, definiendo distintas

formas y momentos en los que se ajusta el tamaño de la ventana de congestión, y distintos métodos para detectar y tratar la congestión en la red.

TCP fue desarrollado y posteriormente optimizado para redes cableadas, lo que implica que la pérdida de paquetes se debe casi con exclusividad a la congestión en la red, suponiendo una tasa de error de bits en tránsito prácticamente inexistente. Con el crecimiento en la demanda de movilidad, se comenzó a utilizar en enlaces inalámbricos. Estos enlaces tienen características muy diferentes, de manera que los retrasos y las pérdidas de paquetes ya no se deben exclusivamente a la congestión, sino también al daño de paquetes en tránsito [4] y la desconexión temporal de los nodos, lo que puede ocasionar un comportamiento muy poco eficiente del protocolo, en ese entorno.

Internet es un entorno heterogéneo y compartido. El uso efectivo de la red no solo depende de que un flujo TCP pueda aprovechar sus recursos, sino también de qué tan bien interactúa con otros flujos a través de la misma red. La eficacia no es el único parámetro importante de los algoritmos de control de congestión, también deben hacer valer el uso equitativo de los recursos compartidos, en especial el ancho de banda.

Es por ello que el análisis de dos flujos de datos con distintas variantes TCP, puede aportar una visión acerca de cómo es utilizado el ancho de banda bajo distintas estrategias y ver cómo interactúan entre sí ante una competencia por la utilización de este recurso. Explorando este camino, planteamos aquí un modelo simple, con dos flujos que compiten entre sí en la misma trayectoria de red en una topología híbrida, sin otros flujos presentes.

Sobre alguno de estos protocolos se conocen trabajos específicos de análisis sobre superposición de flujos [5], incluso existen trabajos con análisis de superposición de flujo considerando ciertas variantes protocolos TCP [6]. Sin embargo, el caso de estudio es en base a una topología mixta, lo que no incorpora la complejidad que puede presentar el accionar de los distintos mecanismos de congestión que deben lidiar con una conexión de estas características.

Esto se debe a que las redes inalámbricas presentan desafíos críticos para el protocolo TCP, derivados de un ambiente complejo. Además de las pérdidas de paquetes por congestión, pueden ocurrir pérdidas de paquetes por la movilidad del nodo, por las características del canal inalámbrico y por las restricciones de potencia. La movilidad de los dispositivos en estas redes puede producir un cambio en la ruta, fallas, reordenamiento de paquetes, perdida de paquetes en tránsito y desconexiones cuando se sale del área de cobertura. Los canales inalámbricos son propensos a tener una alta tasa de error (BER). Es vulnerable a errores debido a la atenuación de la señal, la interferencia, los obstáculos y el desvanecimiento por trayectos múltiples. Estos errores pueden generar pérdida de paquetes o recepción de paquetes dañados en el receptor. Además, estos canales inalámbricos son compartidos, lo que limita la capacidad de un nodo para enviar paquetes debido a la contención. La Contención es la competencia por el uso del medio compartido por varios nodos.

La creciente utilización de las redes inalámbricas ha puesto de manifiesto la necesidad de modificar el protocolo TCP, pues supone que la congestión es la causa principal de las pérdidas. De esta manera, el TCP no puede reaccionar adecuadamente a las pérdidas de paquetes no relacionadas con la congestión. Si se pierde un paquete de datos debido a la interferencia en el canal de trasmisión, a pesar de que no hay

desbordamientos de búfer, TCP reduce la ventana de congestión en forma incorrecta. En vez de ello, debería simplemente recuperarse de la pérdida y continuar la transmisión a la misma tasa de envío. Esto produce una caída importante en el rendimiento.

Para la solución de este problema hay dos enfoques diferentes. El primer enfoque es el que mantiene el concepto original de host a host de TCP. En el presente trabajo solo se abordaran algunas soluciones que mantengan este enfoque.

El segundo enfoque, renuncia a la idea de host a host utilizando distintas estrategias. Estas pueden ser que los nodos intermedios divulguen el estado de la red, retransmisión de capa de enlace, o bien aislando las rutas de transmisión inalámbricas propensas a errores de las cableadas utilizando un host intermedio.

2 Caso de estudio

En la figura 1 se observa el modelo implementado en el simulador de eventos discretos Network Simulator 2 (versión 2.35) que representa una topología mixta simple con 2 nodos y sus enlaces cableados, una estación base y dos nodos con sus enlaces inalámbricos. Sobre esta configuración, hemos realizado una confrontación de dos protocolos por vez, para evitar que existan otro tipo de influencias en el resultado.

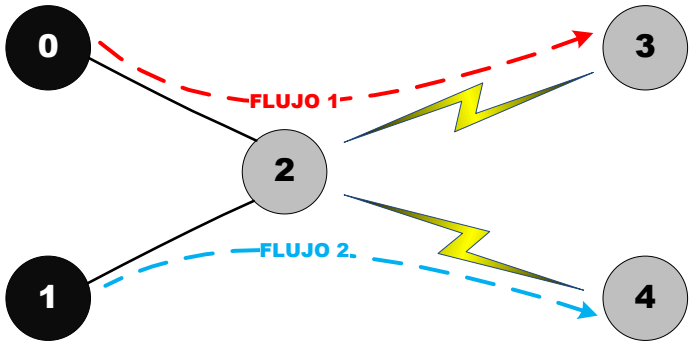


Figura 2. Modelo de estudio

Los dos enlaces cableados que conectan los nodos fijos (0) y (1) con la estación base (2) son full dúplex, de 10 Mb/s y un retardo de 2 ms. El enlace inalámbrico que vincula la estación base (2) y los nodos inalámbricos (3) y (4) se establece en 1 Mb/s, con MAC 802.11. Los nodos inalámbricos no poseen movimiento.

En cada una de las simulaciones, el flujo TCP se estableció desde el nodo (0) (emisor) al nodo (3) (receptor) para la primera variante del protocolo, mientras que para la segunda variante se configuró el nodo (1) como emisor y el nodo (4) como receptor.

Se realizaron 450 simulaciones, en las cuales se confrontaron distintas variantes del protocolo TCP, incluyendo cada una de ellas contra sí mismas. Las variantes utilizadas son: Reno, New Reno, SACK, FACK, Vegas, Veno, LP (Low Proprity),

Westwood, HS (High Speed), H-TCP, Hybla, BIC, Cubic, C-TCP (Compound) e Illinois.

Para cada par de protocolos se realizaron dos simulaciones distintas, intercambiando el orden de inicio de cada flujo, separados por 10 segundos. El primer flujo comienza a trasmitir a los 5 segundos de iniciada la simulación y termina a los 105 segundos, mientras que el segundo flujo comienza a los 15 segundos y termina a los 120 segundos.

De cada una de estas simulaciones se obtuvo el throughput promedio y el Packet Delivery Ratio (PDR) del archivo de traza correspondiente, mediante el uso de scripts AWK. El throughput promedio se obtuvo sumando todos los paquetes salientes del nodo emisor de un flujo, en el intervalo que va desde los 40 a los 100 segundos. El PDR para cada uno de los flujos, se obtuvo calculando el cociente entre los paquetes recibidos correctamente en el receptor y el total de los enviados en el mismo intervalo de tiempo. De forma análoga, se obtuvieron los valores instantáneos del throughput y del tamaño de la ventana de congestión para cada uno de esos flujos.

El objetivo de tomar las muestras solo en el intervalo de tiempo 40 a 100 segundos, es analizar la utilización de los recursos de la red, por parte de los dos flujos que compiten, en un estado estacionario relativo, y de esta manera, evitar las oscilaciones típicas de un estado transitorio, para analizar el comportamiento en un trayecto común de una comunicación establecida.

Para la comparación de las distintas variantes de TCP en competencia se utilizó una métrica resultante del producto del throughput promedio y el PDR correspondiente para cada una de las simulaciones y para cada uno de los flujos.

$$TPP = \text{Thoughput promedio} * PDR \qquad \text{(Eq. 1)}$$

3 Resultados obtenidos

Sobre la topología escogida (figura 1) se han realizado ensayos que posibilitaron conformar dos matrices que representan la confrontación de 15 protocolos. Los resultados pueden observarse en las tablas 2 y 3, donde las filas y columnas representan a las variantes del protocolo TCP analizadas.

En la primera matriz (Tabla 2), cada casilla expresa el TPP de la confrontación entre la variante del protocolo TCP de la fila, con la variante TCP de la columna, siendo la variante de protocolo TCP correspondiente a la fila quién inicia la transmisión y por consiguiente el de la columna el que comienza su transmisión en segundo término

Por ejemplo, en la simulación de los protocolos Cubic vs Hybla, el valor de TPP de Cubic, que es el protocolo que primero inicia, es de 0,289. En contrapartida, el valor obtenido en la situación Hybla vs Cubic, representa el TPP de Hybla cuando este inicia primero, y es de 0,347.

En la segunda matriz representada en la (Tabla 3), el valor de celda que está en la intersección de una fila y columna expresa el TPP del protocolo que inicia su transmisión en segunda instancia, si consideramos nuevamente la confrontación de la variante del protocolo TCP de la fila, con la variante del protocolo TCP de la columna.

En este caso, si por ejemplo tomamos la confrontación de Cubic vs Hybla, teniendo en cuenta siempre la referencia que el primer protocolo se encuentra en la fila y el segundo protocolo en la columna, el valor de TPP de Hybla comenzando la transmisión primero Cubic, es 0,347. Por otro lado, en la simulación Hybla vs. Cubic el valor de TPP de Cubic, habiendo iniciado Hybla primero, es de 0,308.

	Bic	Compound	Cubic	Fack	Highspeed	HTCP	Hybla	Illinois	LowPriority	NewReno	Reno	Sack	Vegas	Veno	Westwood
Bic	0,322	0,322	0,322	0,322	0,322	0,322	0,290	0,322	0,322	0,315	0,315	0,315	0,568	0,322	0,322
Compound	0,322	0,322	0,322	0,322	0,322	0,322	0,278	0,322	0,322	0,315	0,315	0,315	0,568	0,322	0,322
Cubic	0,322	0,322	0,322	0,322	0,322	0,322	0,289	0,322	0,322	0,315	0,315	0,315	0,568	0,322	0,322
Fack	0,322	0,322	0,322	0,322	0,322	0,322	0,253	0,322	0,322	0,315	0,315	0,315	0,568	0,322	0,322
Highspeed	0,322	0,322	0,322	0,322	0,322	0,322	0,278	0,322	0,322	0,315	0,315	0,315	0,568	0,322	0,322
HTCP	0,322	0,322	0,322	0,322	0,322	0,322	0,267	0,322	0,322	0,315	0,315	0,315	0,568	0,322	0,322
Hybla	0,345	0,353	0,347	0,376	0,353	0,351	0,355	0,343	0,353	0,377	0,348	0,367	0,608	0,362	0,345
Illinois	0,322	0,322	0,322	0,322	0,322	0,322	0,290	0,322	0,322	0,315	0,315	0,315	0,568	0,322	0,322
LowPriority	0,322	0,322	0,322	0,322	0,322	0,322	0,278	0,322	0,322	0,315	0,315	0,315	0,568	0,322	0,322
NewReno	0,313	0,316	0,316	0,315	0,316	0,316	0,251	0,315	0,316	0,318	0,318	0,318	0,494	0,314	0,316
Reno	0,313	0,316	0,316	0,315	0,316	0,316	0,276	0,315	0,316	0,318	0,318	0,318	0,494	0,314	0,316
Sack	0,313	0,316	0,316	0,315	0,316	0,316	0,270	0,315	0,316	0,318	0,318	0,318	0,494	0,314	0,316
Vegas	0,083	0,083	0,083	0,083	0,083	0,083	0,046	0,083	0,083	0,081	0,081	0,081	0,310	0,083	0,083
Veno	0,322	0,322	0,322	0,322	0,322	0,322	0,269	0,322	0,322	0,315	0,315	0,315	0,568	0,322	0,322
Westwood	0,322	0,322	0,322	0,322	0,322	0,322	0,283	0,322	0,322	0,315	0,315	0,315	0,568	0,322	0,322

Tabla 2. TPP obtenido del flujo que inicia en primer término, considerado en la fila correspondiente

	Bic	Compound	Cubic	Fack	Highspeed	HTCP	Hybla	Illinois	LowPriority	NewReno	Reno	Sack	Vegas	Veno	Westwood
Bic	0,325	0,325	0,325	0,325	0,325	0,325	0,345	0,325	0,325	0,312	0,312	0,312	0,565	0,325	0,325
Compound	0,325	0,325	0,325	0,325	0,325	0,325	0,353	0,325	0,325	0,314	0,314	0,314	0,565	0,325	0,325
Cubic	0,325	0,325	0,325	0,325	0,325	0,325	0,347	0,325	0,325	0,317	0,317	0,317	0,565	0,325	0,325
Fack	0,325	0,325	0,325	0,325	0,325	0,325	0,376	0,325	0,325	0,317	0,317	0,317	0,565	0,325	0,325
Highspeed	0,325	0,325	0,325	0,325	0,325	0,325	0,353	0,325	0,325	0,314	0,314	0,314	0,565	0,325	0,325
HTCP	0,325	0,325	0,325	0,325	0,325	0,325	0,351	0,325	0,325	0,316	0,316	0,316	0,565	0,325	0,325
Hybla	0,380	0,370	0,308	0,362	0,366	0,313	0,345	0,392	0,374	0,359	0,341	0,351	0,600	0,367	0,375
Illinois	0,325	0,325	0,325	0,325	0,325	0,325	0.343	0,325	0,325	0,313	0,313	0,313	0,565	0,325	0,325
LowPriority	0,325	0,325	0,325	0,325	0,325	0,325	0,353	0,325	0,325	0,314	0,314	0,314	0,565	0,325	0,325
NewReno	0,315	0,315	0,315	0,315	0,315	0,315	0,377	0,315	0,315	0,319	0,319	0,319	0,560	0,315	0,315
Reno	0,315	0,315	0,315	0,315	0,315	0,315	0,348	0,315	0,315	0,319	0,319	0,319	0,560	0,315	0,315
Sack	0,315	0,315	0,315	0,315	0,315	0,315	0,367	0,315	0,315	0,319	0,319	0,319	0,560	0,315	0,315
Vegas	0,084	0,084	0,084	0,084	0,084	0,084	0,049	0,084	0,084	0,082	0,082	0,082	0,311	0,084	0,084
Veno	0,325	0,325	0,325	0,325	0,325	0,325	0,362	0,325	0,325	0,314	0,314	0,314	0,565	0,325	0,325
Westwood	0,325	0,325	0,325	0,325	0,325	0,325	0,345	0,325	0,325	0,314	0,314	0,314	0,565	0,325	0,325

Tabla 3. TPP obtenido del flujo que inicia en segundo término, considerado en la columna correspondiente

Analizando las representaciones de los valores instantáneos de throughput promedio y el tamaño de la ventana de congestión, se puede apreciar, en forma cualitativa, la evolución en el tiempo de los flujos instantáneos de los datos. Por otro lado, a través de las variaciones de la ventana de congestión, podemos apreciar cómo los algoritmos de control de congestión intentan desplegar su estrategia a fin de adaptarse al ancho de banda disponible. Así se tiene una primera aproximación de la agresividad que pueden manifestar ciertos protocolos y la manera en que el algoritmo de congestión intenta obtener el mayor ancho de banda disponible para transmitir los datos.

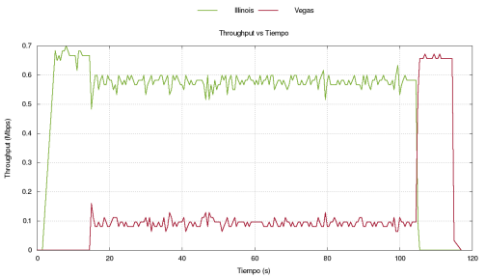


Figura 2. Throughput - Illinois Vs. Vegas

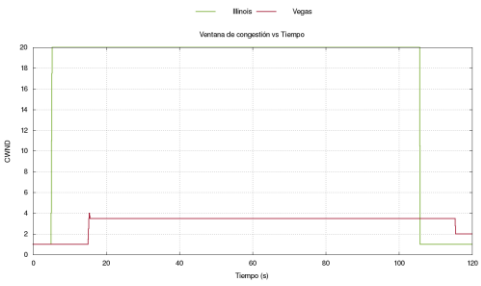


Figura 3. CWND Ilinois Vs. Vegas

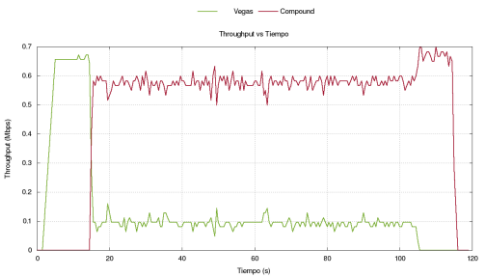


Figura 4. Throughput - Vegas vs. Compound

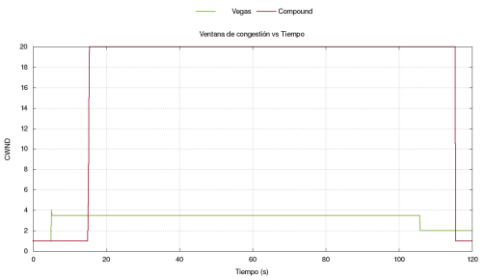


Figura 5. CWND Vegas vs. Compound

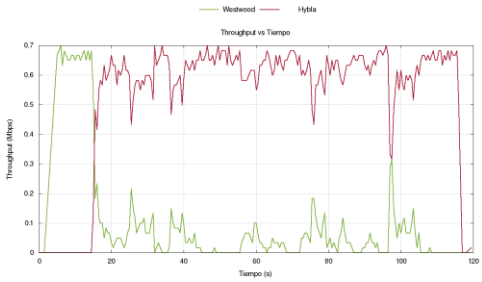


Figura 6. Throughput - Westwood Vs. Hybla

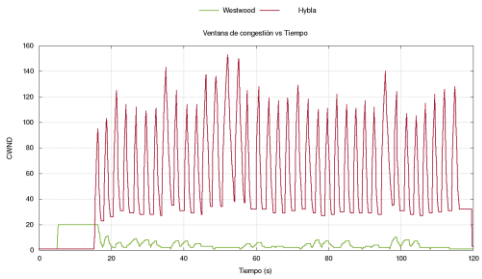


Figura 7. CWND Westwood Vs. Hybla

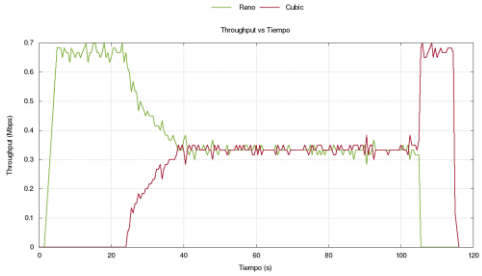


Figura 8. Throughput - Reno vs. Cubic

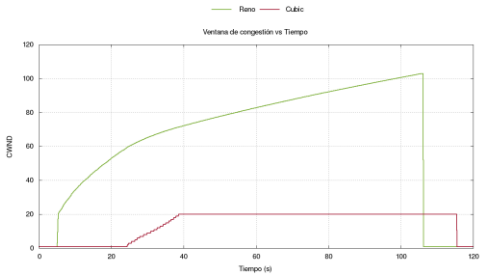


Figura 9. CWND Reno vs. Cubic

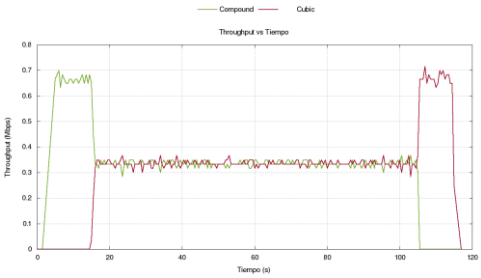


Figura 10. Throughput - Compound vs. Cubic

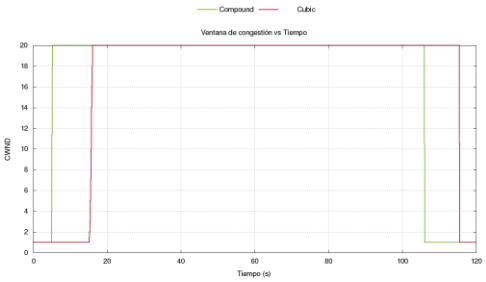


Figura 11. CWND Coumpond vs. Cubic

Como puede observarse en la tabla 2 y 3, existen grupos de datos que podrían agruparse por rangos de valores. Tomando en cuenta ello, se ha sombreado las celdas de dichas tablas en función de una distribución que comprenden a los intervalos (0 a 0,100), (0,101 a 0,300), (0,301 a 0,400) y (0,401 a 1). Por último, podemos acotar que, el mero valor de TPP no expresa en forma cabal el comportamiento de los protocolos, por ello se ha recurrido a la presentación de gráficos en donde pueden apreciarse las evoluciones de flujos instantáneos y sus respectivas variaciones de ventana de congestión, para casos que se ha considerado interesantes para el siguiente análisis.

4 Conclusiones

En los ensayos se confrontaron todos los protocolos entre sí, lo que implica una confrontación de sus respectivos controles de congestión. De esta manera podemos evaluar el comportamiento de un protocolo contra cada uno de los restantes protocolos seleccionados. A partir de la heterogeneidad de Internet y que un flujo TCP puede atravesar distintos tipos de enlaces desde el emisor hasta el receptor, resultó interesante la evaluación del comportamiento de las variantes propuestas en esta topología híbrida. Más allá de esta justificación, se puede observar que existen variantes del protocolo TCP que pueden coexistir mientras que otras no lo pueden hacer. En esta situación, es interesante ver cómo los resultados que arrojan las medidas de TPP (Tabla 2 y 3) muestran que existen variantes del protocolo TCP que se evidencian más agresivas, captando todo el ancho de banda posible para transmitir los datos, mientras que otras muestran ser particularmente vulnerables contra sus congéneres al momento de utilizar el ancho de banda.

Analizando los datos y gráficas obtenidas, podemos ver tres tipos de comportamientos básicos considerando que tenemos dos protocolos A y B, donde primero comienza el protocolo A y al cabo de un tiempo, lo sigue el protocolo B. En estas condiciones podemos determinar tres situaciones características:

1) Primer caso: al comenzar el protocolo A capta el máximo ancho de banda disponible. Al iniciar el protocolo B es relegado en la posibilidad de crecer en el uso del recurso, estabilizándose con un valor de throughput instantáneo sensiblemente

menor que el de A. Aquí el protocolo Vegas, aun siendo el que inicia, resulta particularmente sensible respecto al resto de los protocolos. Como ejemplo de este caso, Illinois vs. Vegas (figura 2) y Vegas vs. Compound (figura 4) son muy representativos de este comportamiento. En ellos se observa que Vegas, comenzando a transmitir en segundo término (figura 2) o trasmitiendo en primer término (figura 4), resigna rápidamente si participación en el uso del ancho de banda del canal.

2) Segundo caso: al comenzar el protocolo A capta el mayor ancho de banda disponible, pero al iniciar el protocolo B relega al protocolo A, obteniendo una porción mayor del ancho de banda disponible y disminuyendo sensiblemente el ancho de banda que puede utilizar A. Las simulaciones de Westwood vs. Hybla (figura 6) muestran la agresividad de Hybla para capturar el ancho de banda, refrendado por la evolución su CWND (figura 7). Este ejemplo es representativo para mostrar este tipo de comportamiento.

3) Tercer caso: el protocolo A comienza a transmitir datos y capta el mayor ancho de banda disponible. Al iniciar el protocolo B se producen fluctuaciones donde ninguno parece ejercer un predominio marcado sobre el otro y, paulatinamente se llega a una distribución equitativa del ancho de banda (figura 8 y 10). En este caso en particular, el resultado de la evolución del throughput instantáneo en la zona de análisis (de t= 40 a t= 100 segundos), determina la utilización del ancho de banda en partes casi iguales. La llegada a ese punto de equilibrio puede lograrse de manera paulatina (figura 8) o de una manera más súbita (figura 10). Este último es un caso particularmente interesante de análisis, ya que normalmente Compound se implementa en los equipos con sistema operativo Windows, mientras que los sistemas operativos Linux normalmente utilizan CUBIC.

Como vemos los protocolos que basan sus algoritmos de control de congestión en métodos reactivos (basados en pérdida de paquetes) dominan en mayor o menor medida al único protocolo proactivo seleccionado (Vegas). Si analizamos la confrontación entre los protocolos restantes, vemos que no existen importantes diferencias entre el ancho de banda que cada uno emplea para la transmisión de datos. Esto pone de manifiesto que las negociaciones por la captura del ancho de banda responden a un uso equitativo del mismo, esto podría verificarse mediante un análisis basado en el concepto de equidad.

Analizando las distintas pruebas de confrontación, vemos que el algoritmo de control de congestión que posee Vegas no le permite sostener un uso de ancho de banda cuando compite con los restantes protocolos seleccionados para realizar las pruebas. Así Vegas resigna ancho de banda y pasa a un segundo plano.

En el otro extremo, protocolos como Hybla, Illinois y aquellos destinados a redes de alta velocidad, tienden a dominar a sus contrapartes en los ensayos e intentan capturar el mayor ancho de banda posible.

Por último, tenemos el tercer caso en donde los algoritmos de control de congestión negocian su tránsito a un estado estable, distribuyendo de una manera equitativa el ancho de banda disponible. Este comportamiento se da en la gran mayoría de las confrontaciones que se observaron, dos ejemplos particulares son los mostrados en las figuras 8 y 10. Como decíamos, si bien el comportamiento es similar, no son exactamente iguales. La diferencia fundamental se da en el tiempo

empleado para llegar un estado de equilibrio. Efectivamente, en el caso de la figura 8 se observa que la transición desde que comienza el protocolo B, hasta que se llega a un estado de distribución equitativa de throughput instantáneo es de 15,5 segundos, mientras que en el caso de la figura 10, este tiempo es de 1,5 segundos. Estos dos casos se seleccionaron de una amplia gama de valores y formas, en donde puede apreciarse como las curvas de throughput muestran que llegan a un equilibrio.

5 Agradecimientos

Este proyecto fue financiado por la Universidad Nacional de la Patagonia Austral, Unidad Académica Río Gallegos.

Referencias

1. J. Postel, “RFC 793: Transmission Control Protocol”, September 1981.
2. M. Allman, V. Paxson, and E. Blanton. TCP Congestion Control. RFC5681 (Draft Standard), September 2009.
3. S. Floyd and K. Fall, “Promoting the use of end-to-end congestion control in the Internet,” IEEE/ACM Trans. Netw., vol. 7, no. 4, pp. 458–472, August 1999.
4. H. Elaarag, “Improving TCP Performance over Mobile Networks”, ACM Computing Surveys, Vol. 34, No. 3, September 2002.
5. A. Rodgers and T. Kuznetsova. “TCP Reno and Vegas co-existence.” Semantic Scholars. (2015).
6. A. Esterhuizen, A.E. Krzesinski. “TCP Congestion Control Comparison”. SATNAC, September 2012.
7. V. Jacobson, “Congestion avoidance and control,” ACM SIGCOMM, pp. 314–329, 1988
8. W. Stevens, “TCP slow start, congestion avoidance, fast retransmit, and fast re-covery algorithms”. RFC 2001, 1997.
9. V. Jacobson, “Congestion Avoidance and Control”, ACM SIGCOMM Computer Communication Review, 25(1), 1995, pp. 157-187.
10. W. Stevens, “RFC2001—TCP Slow Start, Congestion Avoidance, Fast Retransmit,” RFC, 1997
11. V. Jacobson, “Modified TCP congestion avoidance algorithm,” email to the end2end list, April 1990.
12. S. Floyd and T. Henderson, “RFC2582—the NewReno modification to TCP’s fast recovery algorithm,” RFC, 1999.
13. S. Floyd, T. Henderson, and A. Gurtov, “RFC3782—the NewReno modification to TCP’s fast recovery algorithm,” RFC, 2004.
14. M. Mathis, J. Mahdavi, S. Floyd, and A. Romanov, “RFC2018—TCP selective acknowledgment options,” RFC, 1996.
15. M. Mathis and J. Mahdavi, “Forward acknowledgement: refining TCP congestion control,” in Proc. conference on applications, technologies, architectures, and protocols for computer communications (SIGCOMM), New York, NY, USA, 1996, pp. 281–291.
16. L. Brakmo and L. Peterson, “TCP Vegas: end to end congestion avoidance on a global Internet,” IEEE Journal Selected Areas Communications. vol. 13, no. 8, pp. 1465–1480, October 1995.

17. G. Hasegawa, K. Kurata, and M. Murata, “Analysis and improvement of fairness between TCP Reno and Vegas for deployment of TCP Vegas to the Internet” in Proc. IEEE ICNP, 2000, pp. 177–186.
18. C. P. Fu and S. C. Liew, “TCP Veno: TCP enhancement for transmission over wireless access networks,” IEEE J. Sel. Areas Commun.,vol. 21, no. 2, February 2003.
19. A. Kuzmanovic and E. Knightly, “TCP-LP: low-priority service via end-point congestion control,” IEEE/ACM Trans. Netw., vol. 14, no. 4, pp. 739–752, 2006.
20. A. Kuzmanovic and E. W. Knightly, “TCP-LP: a distributed algorithm for low priority data transfer,” in Proc. IEEE INFOCOM, April 2003.
21. S. Mascolo, C. Casetti, M. Gerla, M. Y. Sanadidi, and R. Wang, “TCP Westwood: Bandwidth estimation for enhanced transport over wireless links,” in Proc. ACM MOBICOM, 2001, pp. 287–297.
22. S. Floyd, “RFC3649—HighSpeed TCP for large congestion windows,” RFC, 2003.
23. S. Floyd, “HighSpeed TCP and Quick-Start for Fast Long-Distance HighSpeed TCP and Quick-Start for fast longdistance networks” (slides), TSVWG, IETF, March 2003.
24. D. Leith and R. Shorten, “H-TCP: TCP for high-speed and longdistance networks,” in Proceedings of PFLDnet, 2004.
25. D. Leith, “H-TCP: TCP congestion control for high bandwidth-delaybproduct paths,” IETF Internet Draft, http://tools.ietf.org/html/draftleith-tcp-htcp-06, 2008.
26. C. Caini and R. Firrincieli, “TCP Hybla: a TCP enhancement for heterogeneous networks,” International J. Satellite Communications and Networking, vol. 22, pp. 547–566, 2004.
27. L. Xu, K. Harfoush, and I. Rhee, “Binary increase congestion control for fast, long distance networks,” in Proc. IEEE INFOCOM, vol. 4, March 2004, pp. 2514–2524.
28. I. Rhee and L. Xu, “CUBIC: a new TCP-friendly high-speed TCP variant,” SIGOPS Operating Systems Review, vol. 42, no. 5, pp. 64–74, July 2008.
29. K. Tan, J. Song, Q. Zhang, and M. Sridharan, “A compound TCP approach for high-speed and long distance networks,” IEEE INFOCOM. July 2005.
30. S. Liu, T. Basar, and R. Srikant, “TCP-Illinois: A loss and delay-based congestion control algorithm for high-speed networks,” in Proc. First International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS), 2006.

A Multi-objective Evolutionary Algorithms Study applied to Routing and Spectrum Assignment in EON networks

Carmelo Fretes[1], Diego Pintos Roa[1], Ysapy Mimbi Ortiz[1],

[1] Universidad Nacional de Asunción, Facultad Politécnica,
Asunción, Paraguay
{carmelofretes, dppintoroa, ysapymimbi}@gmail.com, {dpinto}@pol.una.py

Abstract. In this work a new approach based on multi-objective evolutionary algorithms (MOEA) is proposed for the routing and spectrum assignment (RSA) Problem in elastic optical networks (EON); where, given a set of unicast requests, the proposed MOEA minimizes (a) the total cost, and (b) the spectrum used, simultaneously under optical layer constraints. The test experimental indicates that the proposal is suitable for the RSA when it is compared to another MOEA of the-state-of-the-art considering different quality measures. Basically, the proposed MOEA sequences the requests to be served under random-and-cost based strategy while that considered the-state-of-the-art is just random.

Keywords: Routing and Spectrum Assignment, Elastic Optical Networks, Multi-objective Optimization, Evolutionary Algorithms.

1 Introduction

For the resolution of the numerous problems that have multiple objectives, a good meta-heuristic for this type of problems are the evolutionary algorithms (EA - Evolutionary Algorithm). Traditional EAs are customized to adapt to multi-objective problems, through the use of specialized fitness functions and the introduction of methods to promote the diversity of the solution. There are general approaches to the optimization of multiple objectives. One is to combine the individual objective functions in a single compound function or move all, except one of them for the set of constraints. The next approach is to determine a whole set of optimal Pareto solutions or a representative subset. An optimal set of Pareto is a set of solutions that are not dominated with respect to the others [8]. This last approach is more convenient for making decision over a set of trade-off best solution instead of two first approaches. In this work, the main contribution is an approach based on a Multi-objective Evolutionary Algorithms (MOEA) for the RSA problem, in which it is determined that the proposed approach improves in terms of quality from the Pareto front to the work presented in [8]. The MOEA optimizes: (a) the spectrum used, and (b) the total cost, subject to the constraints of continuity, contiguity, and spectrum conflict imposed by the EON layer.

2 Elastic Optical Networks

We can define the EON as an OTN (Optical Transport Network) where all the equipment and the control plane can handle optical channels of variable bandwidth and all the switching elements can support different granularities in the spectrum of the channels that transmit information. EONs introduce fixed granularity into the bandwidth of the channels transported through the fiber. The ITU-T G.694.1, establishes a series of fixed spectral grids, which divide the optical spectrum between 1530-1565 nm, from the C band, ranging from 12.5 GHz. (Giga Herz) to 100 GHz, where most used are those of 50 GHz and 100 GHz [5]. The important change in the EON architecture is the replacement of the fixed grid (Fixed-grid) by a new flexible grid (Flexi-grid). The ITU-T is focused on the revision of a G.694.1 standard [5], for a division of the flexible optical spectrum called flexi-grid, for which the optical spectrum of the C band (1530-1565 nm) was defined, which is divided into FS (Frequency Slots) of fixed sizes of 6.25, 12.5, 25 and 50 GHz [6] and in addition a central frequency (CF, Central Frequency) is assigned to each elastic optical path (EOP - Elastic Optical Path) that must coincide with the beginning or the end of these slots existing differences in a fixed grid scheme and a flexible grid scheme In the case of the fixed grid scheme, we can observe the inefficient use of spectrum due to the fixed division that has the 50 GHz spectrum between each CF's, and if we observe the scheme of flexible grids can be noticed the free spectrum obtained thanks to the fine granularity that it offers and that allows to assign in a flexible way only the required bandwidth. The problem of RSA in Elastic Optical Networks is similar to the problem of Routing and Wavelength Assignment (RWA) in networks based on WDM. The difference between them (RSA and RWA) is the ability to flexibly assign the frequency spectrum. The RSA is classified into two types: Online/Dynamic and Offline/ Static traffic. In the case of the offline RSA problem, the list of all transmission requests is already entered as input, in order to proceed with the analysis and resolution with this input data. For the RSA online problem, the analysis and resolution is done as the requests arrive dynamically. In the first problem are can be applied optimization strategies; while in second one are usually developed heuristics.

3 Related Work

As the RSA is considered a NP-Complete problem [7], it has been treated with several techniques, exact and heuristic, both for dynamic traffic and for static traffic. Among the exact techniques are the ILP, while among the heuristics are optimizations with Colony of Bees (BCO, Bee Colony Optimization) [10], Genetic Algorithms (GA, Genetic Algorithm) [11] [12] [13], among others [14] [15].

Different ILP models for small instances and different heuristics for more real scenarios have been used successfully to solve the RSA problem. As an example we can mention in [16] an ILP model was proposed to minimize the use of the spectrum to serve a traffic matrix in an EON. The authors propose a method that divides the problem into two sub-problems, the first is the routing and the second is the spectrum assignment and solves them sequentially, using a route-based approach. They also propose a

heuristic algorithm that serves the connections one by one sequentially. Then in [2], the authors extend their previous results including consideration of modulation level. With this new consideration, a new problem was defined routing, modulation level and spectrum assignment (RMLSA), being outside the scope of this work. Other problems such as Fragmentation Aware and Dynamic Traffic are also not considered. Another ILP formulation and the proof that the RSA problem is a NP-complete problem can be found in [7]. The work proposed in [8], presents the multi-objective RSA problem and its associated algorithm model. Each request has many possible routes, and in each routing it has several spectrum assignment options. The problem is to minimize the spectrum width to support all requests and minimize the overall cost of the spectrum in the link. The objective function for the work proposed in [8] is as following: there are two objectives associated with each solution. The first objective *f1*, is the width of the spectrum that indicates the maximum indexed slice used in the network. The second objective *f2* is the total cost of the spectrum link. Given a set of requests, the route and channel are calculated for each one. After attending each demand sequentially and without any sort of ordering, the spectrum availabilities vector of each link is updated.

In this work it is developed a pure multi-objective approach to calculate a Pareto front. This approach is an extension of the work presented in [4] which has an approach based on weighted sum. In our work, as in [8] it has many possible routes, and in each routing it has several spectrum assignment options. The problem is to minimize the spectrum used and the overall cost of the link spectrum at the same time. The same objective function is taken from [8] and the requests are handled as follows: applications are ordered from highest to lowest, defined by the highest possible cost of said request, the first 30% of said list is attended in the first place, while the remaining 70% is treated in a random manner, unlike [8] it is a random ordering. More details are given in section 7.

4 Problem Statement

Given the physical topology, the matrix of demands and a list of pre-calculated routes (as K-shortest path), we need to satisfy all the demands of source-destination connection; i.e. to determine the route and spectrum assignment for each traffic demand with optimum spectrum utilization and he total cost. The spectrum utilization is given by the maximum index FS used on all fibers in the network while the total cost is depending on the distance traveled and the FS requested. For the proposed model, the following assumptions are established: The spectral resource of each optical fiber is divided into FS; the capacity of the fiber in terms of FS is limited in all links; the connection demands are bidirectional, and a complete end-to-end optical path must be found for each demand; A set of K specific route is given for a connection in advance; the request is represented by three tuples (s, d, α_{sd}), including the source node s, the destination node d, and the bandwidth / data rate demanded α considered in the quantity of FS requested.

4.1 Multi-Objective Formulation Problem

Given:

G	: Network topology, which represents an EON
E	: Set of links, in G
V	: Set of vertices, in G
GB	: Amount of FS for Band Guard
F_{total}	: Amount of FS available in each fiber
P	: Set of K routes for each demand
K	: Number of available routes
SD	: Quantity of demands

The notations and the formulation are presented below:
Constants:

$dist_max$	: Maximum distance traveled considering the longest routes.
$espectrum_max$	: Maximum FS index available.
$cost_max$	: Total cost of applications considering their maximum distances.
$dist_p^{sd}$	: Distance of the route p
α_{sd}	: Quantity of FS requested by the application where $s, d, \in V$

Indexes:

sd	: Demand index, $sd \in \{1, 2, \ldots, SD\}$
p	: Route index, $p \in \{1, 2, \ldots, SD\}$
mn	: Directional link index, $m \neq n$

Variables:

x_p^{sd}	: 1 if the path p is used to meet the request sd, Otherwise
Λ_{sd}	: First FS assigned to the request sd, $sd \in \{0,\ldots, F_{total} - 1\}$
$\Delta_{sd, s'd'}$	: Indicator that is equal to 0 if $\Lambda_{s'd'} < \Lambda_{sd}$, and 1 in otherwise.

Objective function:
Minimize $f(x) = [f1, f2]$

Subject to:

- The Spectrum use:

$$f_1 = \frac{\max_{\forall sd}(\Lambda_{sd} + \alpha_{sd})}{spectrum_{max}} \quad (1)$$

- The total cost:

$$f_2 = \frac{\sum_{sd} \sum_p \left(\alpha_p^{sd} * dist_p^{sd} * x_p^{sd}\right)}{cost_{max}} \quad (2)$$

$$\sum_{p \in P_{sd}} x_p = 1, \forall (s,d) \quad (3)$$

$$\Lambda_{sd} + \alpha_{sd} * x_p^{sd} + GB - \Lambda_{s'd'} \leq (Ftotal + GB) * \left(1 - \delta_{sd,s'd'} + 2 - x_p^{sd} - x_{p'}^{s'd'}\right) \quad (4)$$

$$\Lambda_{s'd'} + \alpha_{s'd'} * x_{p'}^{s'd'} + GB - \Lambda_{sd} \leq (Ftotal + GB) * \left(1 - \delta_{s'd',sd} + 2 - x_p^{sd} - x_{p'}^{s'd'}\right) \quad (5)$$

$$\delta_{sd,s'd'} + \delta_{s'd',sd} = 1 \quad (6)$$

$$\Lambda_{s'd'} - \Lambda_{sd} < F_{total} * \delta_{sd,s'd'} \quad (7)$$

$$\Lambda_{sd} - \Lambda_{s'd'} < F_{total} * \delta_{s'd',sd} \quad (8)$$

The objective function (1) represents the maximum spectrum used, and (3) represents the total cost. On the other hand, we have that, for all request *sd, s'd'* and the paths $p \in P_{sd}$ and $p' \in P_{s'd'}$ with p and p' sharing at least one common link *mn* the constraints (3), (4), (5), (6), (7) and (8) represents the total cost. Restrictions (3), (4) and (5) ensure that the portions of spectrum that are assigned to connections that use paths that share a common link do not overlap and are adjacent. Also, for all requests *sd, s'd'* that have $p \in P_{s'd'}$, with p and p' sharing at least one common link ($\exists$ *mn* : $nm \in p \wedge mn \in p'$), the constraints (6), (7) and (8) ensure that either $\delta_{sd,s'd'} = 1$ means that the initial frequency Λ_{sd} is smaller than the initial frequency $\Lambda_{s'd'}$, that is, $\Lambda_{sd} < \Lambda_{s'd'}$, o $\delta_{s'd',sd} = 1$, in which case $\Lambda_{sd} > \Lambda_{sd}$. Note that Λ_{sd} and $\Lambda_{s'd'}$ are always bounded superiorly by F_{total}, and that therefore their difference will always be less than F_{total}.

5 NSGA II Implementation

Our algorithm, which is an extension of the algorithm MOEA presented in [4], begins with the creation of the initial population. This MOEA is called Non-dominated Sorting Genetic Algorithm II, NSGAII. The best solutions are found over several generations. Operators such as crossing and mutation explore other possible solutions. In this implementation, the objective is to find the route and the set of FS for each request, such that the total distance traveled, the maximum FS used and the total cost are minimized; all this complying with the respective RSA restrictions. The implementation of the NSGAII is described below in Algorithm 1.

Algorithm 1: NSGAII

INPUT: Route table P; Total amount of FS; List of demands; Size of the population; Probability of mutation; Stop

```
Criterion; FS Assignment Algorithm; Total Distance, Maximum
FS, Maximum Cost
  OUTPUT: ParetoFront
  1: Initialize Population (P)
  2: While the stop criterion is not met
  3:       Q = generate individual (P) by selection, crossing
               and mutation
  4:       Q = Q U P
  5:       R = Construct the Pareto front from Q based
               in dominance
  6:       Build Pareto fronts (R)
  7:       Calculate Distance of Crowding (R)
  8:       P = [0]
  9:       while P < PopulationSize
 10:              Include the solution R in population P
 11:       End while
 12: End while
 13: return ParetoFront (P)
```

In the NSGA II presented in this work, the chromosome represents a set of requests attended. Basically, the chromosome is a compound vector in which each gene represents an attended request. Each element of said vector contains: the index of the assigned route (taken from the table of pre calculated routes), and the index of the assigned FS of the request. The steps of the algorithm procedure can be described below:

Initial Population. The first step is to initialize the population. The NSGA II begins with an initial population of chromosomes, defined as explained below. The Algorithm deals with the requests in a determined order, which was taken from a paper presented in [1]. At work, the order is defined as follows: orders are ordered from highest to lowest, defined by the highest possible cost of said request, the first 30% of said list is attended in the first place, while the remaining 70% is attended at random. This order is represented by the positions of the genes in the chromosome and is maintained throughout the execution of the algorithm. Then, randomly assign the routes and FS to the demands, taking into account the previously defined order. Each chromosome encodes a valid solution.

Selection of chromosomes for the next generation. The NSGA II algorithm shows us that the cycle begins with the selection of individuals, in step 3. The stochastic universal sampling method is used to select two parents to produce new individuals for the next generation [9].

Crossover operator. In this work we used the two-point cross operator [9] through which two cut points are randomly generated in each player, using the same points generated, assigning intercalary each segment generated from the parents to each child. In algorithm 1 is applied in step 6. In Figure 1, we can observe the crossing procedure in which the cut points generated randomly were 1 and 2, dividing the player into 3 segments. The first segment of player 1 is assigned to the first segment of descendant 1, so the first segment of player 2 is assigned to the first segment of descendant 2. Then, the second segment of player 1 is assigned to the second descendant, while the second segment Player segment 2 is assigned as the second segment of the first descendant.

Then the last segments are interspersed, resulting in both descendants shown in figure 4. This process is repeated until crossing the entire current population.

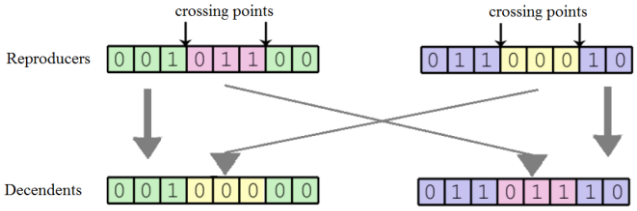


Figure 1: Crossing of 2 reproducers

Mutation. This procedure is applied after crossing, in each individual independently, in step 7 of algorithm 1. For the individual selected, according to the mutation probability obtained, a position of the vector is chosen randomly to change the route used in said position. Selecting a route from those available for said position, you have a higher probability of generating a feasible solution.

Pareto dominance. In step 4 the union of the two populations Q = Q ∪ P is performed, in step 5 and 6 the population is classified into categories (ranking) on the basis of non-dominance. Each solution is assigned a fitness value equal to its non-domain range (rank 0 is the best). Then the newly formed population is classified into categories (rank) according to their domain relation, and then, as explained in step 7, calculate the Crowding Distance of each individual, and then select the best ones in the next cycle that begins in the step 8, select the individuals with the best rank and crowding distance to fill the size of the population, as seen in steps 9, 10 and 11 of algorithm 1. Therefore, the algorithm starts all over again, from the election of breeders, until it reaches the stop condition.

Spectrum assignment. The algorithm used in this NASGA II is Random Fit, which randomly assigns the free FS found that complies with the constraints of the problem.

Stop criterion. A maximum execution time is used as stopping criterion.

6 Experimental Tests and results.

In this section we present the difference with the work proposed in [8] and the work presented by us, in addition the results of the experimental tests are presented and analyzed. The work proposed in [8], presents the multi-objective RSA problem and its associated algorithm model. Each request has many possible routes, and in each routing it has several spectrum assignment options. The problem is to minimize the spectrum width to support all requests and minimize the overall cost of the spectrum in the link. The objective function for the work proposed in [8] is as follows: there are two objectives associated with each chromosome. The first objective *f1*, is the width of the spectrum that indicates the maximum indexed slice used in the network. The second objective *f2* is the total cost of the spectrum link. Given a chromosome, the route and

channel are calculated for each demand. After attending each demand sequentially and without any sort of ordering, the spectrum availabilities vector of each link is updated.

In this developed work, which is an extension of the work presented in [4] which has an approach based on weighted sum, a pure multi-objective approach with Pareto fronts is presented. In our work, as in [8] it has many possible routes, and in each routing it has several spectrum assignment options. The problem is to minimize the spectrum width to support all requests and minimize the overall cost of the link spectrum. The same objective function is taken from [8] and the requests are handled as follows: applications are ordered from highest to lowest, defined by the highest possible cost of said request, the first 30% of said list is attended in the first place, while the remaining 70% is treated in a random manner, unlike [8] it is a random ordering. The tests carried out considering different types of traffic load, on the NSF topology of 14 nodes, different K values (paths) and different amounts of demands, try to replicate various possible scenarios of the problem to solve. The experimental tests carried out show that our proposal for the ordering of the requests presents promising results.

6.1 Testing environment

The experiments were performed on a computer with an Intel Core i3 processor (3.40 GHz) and 8 GB of RAM. The implementation and execution of the MOEAs were carried out with JAVA 8. The traffic loads used were of the all-to-all type, that is, each node of the network makes a transfer request to all others in the network. In addition, the type of traffic load was random. The loads are divided into 3 categories, 50, 100 and 150 (low, medium, high), that is to say that for the category of 50 FS, for each demand a random value between 1 and 50 was generated as a requested quantity of FS; For category 100, for each demand a random value between 1 and 100 was generated as the requested quantity of FS and for category 150, a random value of 1 and 150 was generated as requested quantity of FS. Another variant that was taken was the number of shortest routes pre-calculated, that is, the value K. They were made with the following values of k = 2, 3, 4 and 5 for the network. For the executions of the NSGA II, the values shown in Table 1 were used as evolutionary parameters. The metric used for the comparison of the algorithms are hyper-volume and coverage [4].

Table 1. Parameters used for the execution of the NSGA II

Parameters	Value
Size of the population	50
Probability of mutation	0.1
Stop Criterion (in minutes)	5
Number of independent runs	15
Size of the population	50

6.2 Hyper-volume Metric and Coverage Metric

For the hyper-volume and coverage metric you can see the figure number 2, for load type 50 (low), with the number of paths k = 2, our proposed algorithm of order 30/70 obtains better results before the algorithm without ordering. With k = 3 paths, again our algorithm with order 30/70, exceeds the algorithm without ordering. With k = 4, the algorithm without ordering obtained better results with our algorithm 30/70. For k = 5, our algorithm 30/70 obtained good results. For the hyper-volume and coverage metric for load type 100 (average) for k = 2, the algorithm without ordering obtained better results, with k = 3, our algorithm with order 30/70, has better results before the algorithm without ordering, for k = 4, our 30/70 sorting algorithm improves the results before the algorithm without ordering. For k = 5, we obtained very good results with respect to the algorithm without ordination. For the hyper-volume and coverage metric for load type 150 (high) with k = 2, our sort algorithm 30/70 obtained better results compared to the algorithm without ordering. In k = 3, the algorithm without ordination obtained good results. Our 30/70 sorting algorithm got better results when k = 4 compared to the unordered algorithm. The unordered algorithm had better results when k = 5, compared to our 30/70 sorting algorithm.

		Hypervolume (NORMALIZED VALUES)		COVERAGE	
Type of load (low, mid, High)	Number of roads (K)	Sorting Algorithm 30/70	Unordered Algorithm	Sorting Algorithm 30/70	Unordered Algorithm
50	2	**0,00450575727952277000**	0,00373184833987698000	**0,6**	0,3
	3	**0,03004763727709010000**	0,00114619949446014000	**1,0**	0,0
	4	0,00107277586229790000	**0,00608913898555240000**	0,0	**1,0**
	5	**0,01524133420666540000**	0,01404235292493280000	**0,5**	0,0
100	2	0,000000000401372065 52	**0,00000000428130203222**	0,0	**1,0**
	3	**0,00235590742252450000**	0,00192187663358532000	**0,3**	0,0
	4	**0,00000000625575249301**	0,00000000039930335062	**1,0**	0,0
	5	**0,00000000560063877525**	0,00000000040004562680	**1,0**	0,0
150	2	**0,00000000437845029918**	0,00000000027365314370	**1,0**	0,0
	3	0,00004759974150804820	**0,00063478877339381300**	0,0	**1,0**
	4	**0,00036203376622539800**	0,00003620284610853910	**1,0**	0,0
	5	0,00004690367773651360	**0,00171967667778795000**	0,0	**1,0**

Figure 2: Comparison of algorithms, hyper-volume metric and Coverage metric.

8 CONCLUSIONS AND FUTURE WORK

According to the exposed results, we can conclude that our algorithm with ordering obtains better Pareto Fronts, with respect to the algorithm without ordination. Likewise we conclude that if we give a treatment to the table of requests, ordering them from highest to lowest, defined by the highest possible cost of said request, and we divide the table of requests into two groups, one group of seniors and another group of random attendance we get better Pareto Fronts. As future work to develop we can mention several opportunities: study the performance of other spectrum assignment algorithms, consider other strategies of sorting the request to be served, extend this approaches considering other issues as modulation level assignment or coded assignment.

References

1. Joao HL Capucho, Leandro C Resendo: “ILP model and effective genetic algorithm for routing and spectrum allocation in elastic optical networks”, Microwave & Optoelectronics Conference (IMOC), 2013 SBMO/IEEE MTT-S International, pp. 1—5, 2013.
2. Konstantinos Christodoulopoulos, Ioannis Tomkos, EA Varvarigos: “Elastic bandwidth allocation in flexible OFDM-based optical networks”, Journal of Lightwave Technology, pp. 1354—1366, 2011.
3. Biswanath Mukherjee: Optical WDM networks. Springer Science & Business Media, 2006.
4. Ysapy Mimbi Ortiz, Carmelo Fretes y Diego P. Pinto. “Routing and Spectrum Allocation in Elastic Networks. An approach based on Multi-objective Evolutionary Algorithms”. En: 5th International Conference on Engineering Optimization (ENGOPT). Federal University of Rio de Janeiro. 2016.
5. G Recommendation. 694.1: Spectral grids for WDM applications: DWDM frequency grid”. En: International Telecommunications Union, Tech. Rep. (2012).
6. Fatai Zhang y col. “Requirements for GMPLS control of fexible grids”. En: IETF Internet Draft 1 (2011).
7. Yang Wang, Xiaojun Cao y Yi Pan. “A study of the routing and spectrum allocation in spectrum-sliced elastic optical path networks”. En: INFO-COM, 2011 Proceedings IEEE. IEEE. 2011, pags. 1503-1511.
8. D. T. Hai, "Multi-objective genetic algorithm for solving routing and spectrum assignment problem”, 2017 Seventh International Conference on Information Science and Technology (ICIST), Da Nang, 2017, pp. 177-180. doi: 10.1109/ICIST.2017.7926753
9. Prashant K Gupta y Pranab K Muhuri. “Multi-objective linguistic optimization: Extensions and new directions using 2-tuple fuzzy linguistic representation model”. En: Fuzzy Systems (FUZZ-IEEE), 2017 IEEE International Conference on. IEEE. 2017, pags. 1-6.
10. Goran Z Markovic. “Routing and spectrum allocation in elastic optical networks using bee colony optimization”. En: Photonic Network Communications (2017), pags. 1-19.
11. Miros law Klinkowski. “A genetic algorithm for solving RSA problem in elastic optical networks with dedicated path protection”. En: International Joint Conference CISIS'12-ICEUTE12-SOCO12 Special Sessions. Springer. 2013, pags. 167-176.
12. Feng Wang, Yuan Man y Lichun Man. “Intelligent optimization approach for the k shortest paths problem based on genetic algorithm”. En: Natural Computation (ICNC), 2014 10th International Conference on. IEEE. 2014, pags. 219-224.
13. Dao Thanh Hai y Kha Manh Hoang. “An ecient genetic algorithm approach for solving routing and spectrum assignment problem”. En: Recent Advances in Signal Processing, Telecommunications & Computing (Sig-TelCom), International Conference on. IEEE. 2017, pags. 187-192.
14. Ying Wang y col. “ACO-based routing and spectrum allocation in fexible bandwidth networks”. En: Photonic Network Communications 25.3 (2013), pags. 135-143.
15. Roza Goscien, Miroslaw Klinkowski y KrzysztofWalkowiak. “A tabu search algorithm for routing and spectrum allocation in elastic optical networks”. En: Transparent Optical Networks (ICTON), 2014 16th International Conference on. IEEE. 2014, pags. 1-4.
16. Kostas Christodoulopoulos, Ioannis Tomkos y Emmanouel A Varvarigos. “Routing and spectrum allocation in OFDM-based optical networks with elastic bandwidth allocation”. En: Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE. IEEE. 2010, pags. 1-6.

Análisis de rendimiento de protocolos de Publicación/Suscripción en comunicación con una Red de Sensores Inalámbricos Zigbee

Carlos Hervas Parra, Luis Marrone.
Facultad de Informática, Universidad Nacional de La Plata
La Plata – Argentina
hervasc@espoch.edu.ec, lmarrone@linti.unlp.edu.ar.

Resumen. En esta investigación se realizó una revisión del estado del arte de una Red de Sensores Inalámbricos basada en el estándar IEEE 802.15.4 y su operación a través de la tecnología Zigbee, así como también el estudio de las comunicaciones M2M mediante los protocolos de publicación/suscripción MQTT-SN y CoAP. Proponiendo la construcción de un escenario de pruebas para la implementación de estos protocolos, con el fin de realizar un análisis de rendimiento en su operación con la WSN mediante la medición de consumo de ancho de banda, tasa de entrega y pérdida de publicaciones dentro de la red. Finalmente, se expone los resultados obtenidos a través del análisis de cada protocolo mediante gráficos con sus correspondientes discusiones y conclusiones principales del trabajo.

Palabras Clave: WSN, Zigbee, Publicación/Suscripción, MQTT-SN, CoAP.

1 Introducción

Mediante este trabajo se pretende investigar a fondo el funcionamiento de una Red de Sensores Inalámbricos bajo la tecnología Zigbee y como esta se puede interconectar hacia la red tradicional de servicios (Red LAN o Internet), para su interacción con la misma a través de un protocolo de comunicación que brinde de manera más eficiente la comunicación de extremo a extremo entre las dos redes. Los protocolos de comunicación que se estudian y utilizan a lo largo de la investigación son: el protocolo Message Queue Telemetry Transport Sensor Network (MQTT-SN) y el Protocolo de Aplicación Restringido (CoAP).

Para el desarrollo de la investigación se planteó la utilización de un escenario de red de pruebas compuesto por una Red de Sensores Inalámbricos Zigbee y un Gateway de interconexión hacia un servidor de gestión y almacenamiento de datos dentro de la red TCP/IP como se visualiza en la figura 1.1. Toda la infraestructura de red de pruebas será proporcionada por el laboratorio de informática del Instituto Tecnológico Superior Riobamba.



Figura1.1 Escenario de Red
Fuente: El Autor

Mediante este escenario, se implementó la operación de cada protocolo en estudio para la obtención de mediciones de los parámetros de consumo de ancho de banda, tasa de entrega y pérdida de publicaciones a nivel de capa de aplicación en la comunicación M2M entre la WSN y el servidor de gestión de datos. El análisis de rendimiento se basó en realizar mediciones de estas métricas en función del número de nodos que publiquen información dentro de la red, teniendo un rango de 10 a 40 nodos sensoriales.

A través de la observación de los resultados del análisis de cada protocolo en el escenario de red de pruebas, se realizó una comparativa para la obtención del protocolo de comunicación M2M basado en

publicación/suscripción más adecuado en el diseño de un sistema de comunicación extremo a extremo entre la Red de Sensores Inalámbricos Zigbee y la red de servicios tradicional.

2 Estado del Arte

Desde la aparición de las WSN (Wireless Sensor Networks / Redes de Sensores Inalámbricos), en los últimos años, este tipo de redes ha ido ganando cada vez más atención, desde el punto de vista técnico hasta el punto de vista comercial, debido a su potencial en la implementación de soluciones novedosas y de bajo costo en áreas tales como: automatización industrial, domótica, transporte, monitoreo de eventos físicos y ambientales, sensores para medir o detectar temperatura, presión, humedad, niveles de contaminación o cualquier parámetro crítico de un escenario determinado, es así que el Instituto Tecnológico de Massachusetts (MIT) identificó a las Redes de Sensores Inalámbricos como una de las diez tecnologías que tienen alto impacto en el progreso tecnológico mundial [14].

En las tendencias actuales como son los ambientes del Internet de las Cosas (IoT), se propone que los dispositivos se conecten y puedan detectar/comunicar datos de su entorno. Se espera que esto genere redes con cantidades de nodos mucho más grandes, por lo tanto, es importante utilizar un método eficiente y escalable para recopilar información de Redes de Sensores Inalámbricos de altas densidades. De igual manera, muchas de estas implementaciones requieren que los datos recopilados por los sensores sean enviados hacia aplicaciones que se encuentran en una infraestructura de red tradicional, por ejemplo: redes LAN, redes empresariales, Internet, etc. Este motivo implica que la integración de las WSN a redes TCP/IP sea un tema abierto para su desarrollo tomando en cuenta que las Redes de Sensores Inalámbricos cuentan con recursos limitados en hardware y software en sus nodos finales, por lo que la utilización de protocolos TCP/IP a nivel de capa de red en los nodos no es siempre factible.

Para atender esta necesidad, en la actualidad la integración de una WSN se encuentra resuelta bajo dos enfoques, a través de un Gateway de interconexión o a través de Redes Overlay. En el primer caso se considera que los nodos sensoriales no cuentan con direccionamiento IP, siendo la estación base de la WSN la que posee el aprovisionamiento a nivel de capa de red y actúa como gateway de interconexión para la salida de la WSN hacia la red TCP/IP. Este enfoque es aplicable en las WSN que no permiten cargas extras de protocolo en los nodos sensoriales, como por ejemplo en redes que utilizan Zigbee [1], definidas por el IETF como dispositivos de "Clase 0" que no pueden ejecutar una pila IP segura debido a limitaciones de recursos, bajo nivel de procesamiento y memoria.

Mientras que para el segundo caso se considera incluir parte o toda la pila TCP/IP en los nodos sensoriales mediante el protocolo 6LoWPAN [15], el cual permite la implementación del stack IPv6 sobre las capas PHY y subcapa MAC del estándar IEEE 802.15.4 para que cualquier dispositivo que lo utilice pueda tener conectividad a nivel de capa de red logrando de esta manera que un nodo final o dispositivo sensorial dentro de la WSN tenga conexión directa hacia internet donde se publican sus datos, cada nodo es visible desde la red tradicional externa a través de IPv6 con lo cual la conectividad hacia entornos de IoT se encuentra resuelta.

Por otro lado, en la ciudad de Riobamba a cargo del Instituto Tecnológico Superior Riobamba se encuentra en producción una Red de Sensores Inalámbricos compuesta por un conjunto de motas sensoriales que operan a través de la plataforma Xbee Pro, la cual se encarga de la recolección de parámetros ambientales para su estudio por parte de instituciones públicas de la ciudad, que acceden a la información a través de un servidor que se encuentra instalado de forma local en la red. Con el fin de mejorar y optimizar el acceso y visualización de los datos generados por la red, dada su importancia de operación en la ciudad, se aprovecha el uso de su infraestructura como principal motivación de esta investigación para el análisis de un mecanismo de comunicación que vincule la operación de este tipo de dispositivos de redes sensoriales hacia la red tradicional de servicios (Red LAN o Internet), para su interacción con la misma. Debido a la limitación de recursos de la red, se utiliza el enfoque de integración para la WSN a través de un Gateway de interconexión mediante Zigbee.

También se consideró que la actual aplicación que gestiona los datos de la red, de manera general se encuentra enviando consultas continuamente hacia todos los nodos sensoriales requiriendo una gran cantidad de información que en muchos casos no es utilizada o analizada, con lo cual la obtención de información de la WSN se realiza de forma ineficiente y conduce a un desperdicio de recursos valiosos dentro de la red como son: la energía, procesamiento, memoria y el ancho de banda, que son muy limitados en este tipo de dispositivos. Esta problemática puede superarse con la utilización de un enfoque de comunicación centrado en los datos, en el cual la información se entrega a los consumidores no en función de los identificadores de red, sino más bien en función de sus contenidos e intereses. El modelo de Publicación/Suscripción [2], es un ejemplo muy claro de comunicación centrada en datos y es ampliamente utilizado en redes empresariales, principalmente debido a que

permite una mayor escalabilidad y flexibilidad, además de proporcionar una topología de red más dinámica. Estas son características deseables para una Redes de Sensores Inalámbricos y tendencias actuales que exige el Internet de las Cosas que hacen que el modelo de publicación/suscripción sea muy adecuado para este propósito.

Existen varios estudios sobre el uso de protocolos de transporte confiables en las WSN [12-13]. Es así como algunos protocolos bajo el modelo de publicación/suscripción ya se han propuesto para este tipo de redes [3-4], la mayoría de los cuales sin embargo son protocolos cerrados, es decir, están diseñados para una aplicación específica sin una estandarización universal o su desarrollo ya no es soportado. Por otro lado, los protocolos abiertos han tenido un mejor auge en su desarrollo e intentos de estandarización [5]. Entre estos protocolos, se encuentra MQTT (Message Queuence Telemetry Transport), con su especificación MQTT-SN (Message Queuence Telemetry Transport Sensor Network) [8], como una versión de MQTT adaptada a las peculiaridades de un entorno de comunicación inalámbrica para sensores, siendo ampliamente utilizado para aplicaciones de monitoreo remoto [6], aplicaciones de mensajería [7], y una gama de aplicaciones de automatización y domótica.

También se encuentra el protocolo CoAP (Constrained Application Protocol) [9], desarrollado por el grupo de trabajo IETF CORE [10]. CoAP es un protocolo de transferencia de estado representacional (RESTful) con funcionalidades web optimizado para redes con recursos limitados, está enfocado hacia aplicaciones M2M en entornos del Internet de las cosas, se basa en una arquitectura REST en la que los recursos se ponen a disposición mediante un proceso de aplicación y se visualizan mediante identificadores universales de recursos (URI), lo cual lo hace ideal para su funcionamiento en conjunto con redes de sensores restringidas, tiene como objetivo operar en nodos de bajo nivel de procesamiento dentro de las WSN. CoAP también proporciona un modelo de publicación/suscripción llamado modelo de observador [11].

De esta manera tanto MQTT-SN como CoAP son los dos principales protocolos abiertos que se podrían utilizar dentro de redes de sensores restringidas en recursos que utilicen Zigbee. Debido a que el uso de servicios web en Internet se ha vuelto omnipresente en la mayoría de las aplicaciones actuales, es de vital importancia el estudio de este tipo de protocolos en su función como intermediarios en la unificación de la red de sensores inalámbricos con la red tradicional donde se alojan los servicios de dichas aplicaciones.

2.1 Interconexión de una Red Zigbee hacia la Red Tradicional

Desde su creación una WSN con Zigbee fue concebida como una red donde cada nodo sensorial recolecta datos de su entorno físico y estos son destinados hacia el nodo coordinador el cual mediante una conexión serial permite la transferencia de esta información hacia una aplicación o servidor local para su tratamiento y almacenamiento.

En la actualidad con el surgimiento de IoT, desde una perspectiva de comunicación de extremo a extremo, una WSN con Zigbee puede ser vista como la unificación de dos subredes:

- Una subred que conecta uno o más nodos sensoriales, quienes enrutan los datos hasta el nodo coordinador el cual se interconecta con un gateway utilizando protocolos propios especificados por Zigbee.
- Otra subred de conexión TCP/IP bajo el paradigma de comunicación M2M entre el gateway y un servidor o intermediario de servicios de fondo.

A través de un protocolo de comunicación de datos fiable la puerta de enlace debe agregar todos los datos recibidos de varios nodos sensoriales y luego encargarse de enviarlos hacia un servidor intermediario, para que estos recursos sean consumidos por parte del cliente.

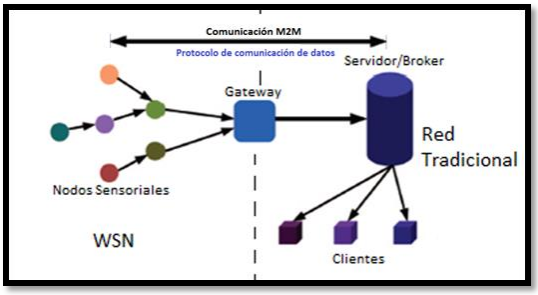


Figura 2.1. Interconexión de una Red Zigbee hacia la Red Tradicional
Fuente: El Autor

Según lo visualizado en la figura 2.1, se puede determinar los elementos que cumplen el papel de intermediarios de interconexión entre las dos subredes que forman parte de la arquitectura de una WSN con visibilidad hacia la red tradicional, siendo los siguientes:

- **Gateway de interconexión**: Es el único punto de acceso a la WSN, permite tomar los datos de los nodos Zigbee y enviarlos por la red tradicional logrando el acoplamiento de las dos subredes ofreciendo los servicios de capas superiores para que los nodos sensoriales con recursos limitados no requieran cargas extras de protocolo.
- **Comunicación M2M:** Proporciona la arquitectura de comunicación entre los dispositivos de la red sensorial y el servidor que gestione los datos de la red.
- **Protocolo de comunicación de datos**: Las WSN a diferencia de las redes inalámbricas tradicionales, están compuestas por nodos que tienen recursos acotados de cómputo, memoria y energía. Estas restricciones implican el surgimiento de protocolos para la comunicación de sus datos diseñados específicamente para este tipo de tecnología.

3 Construcción del escenario de red de pruebas

En la Figura 3.1, se observa la representación gráfica del escenario de red implementado para realizar las pruebas y mediciones, el cual consta de una WSN con un total de 40 nodos sensoriales inalámbricos conectados hacia un nodo coordinador. La comunicación inalámbrica se realiza mediante la tecnología Zigbee con una topología de red tipo malla. Los Nodos de la WSN se encuentran dispersos en el Campus del Instituto Tecnológico Superior Riobamba en la ciudad de Riobamba y en su operación se encargan de la recolección de parámetros ambientales (temperatura, humedad y presencia de gases).

La arquitectura de interconexión hacia la red tradicional es centralizada mediante un dispositivo gateway, el cual se encuentra interconectado hacia el nodo coordinador de la WSN mediante una conexión serial por puerto USB y por una interfaz ethernet se encuentra conectado hacia la red LAN, para su comunicación con el servidor en donde se deben gestionar los datos de la WSN. Adicionalmente en el segmento de la red LAN se encuentra interconectado un ordenador portátil que se encarga de la captura de tráfico generado entre la WSN y el servidor para la medición de los parámetros estipulados en el análisis.

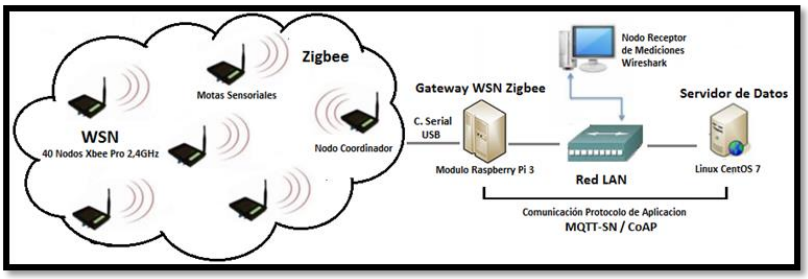


Figura 3.1. Escenario de Red de Pruebas
Fuente: El Autor

Para el análisis de rendimiento de cada protocolo se dispuso de 4 escenarios de operación los cuales varían según el número de nodos de la WSN, es decir las métricas de rendimiento (consumo de ancho de banda, tasa de entrega, retransmisión y pérdida de publicaciones) de cada protocolo fueron medidas en función al número de nodos que trasmitan información por la red, teniendo en cuenta escenarios con 10, 20, 30 y 40 nodos sensoriales transmitiendo datos a través del nodo coordinador y su gateway para su conexión hacia el servidor de gestión de datos en la red LAN.

3.1 Emulación de una comunicación a través de Internet

Se implementó la emulación de una red restringida o amplia dentro del escenario de red de pruebas que permita insertar parámetros de retraso, pérdida, duplicación, corrupción y reordenamiento de paquetes con el objetivo de que se pueda obtener el comportamiento de una red con características similares a las que se presentan en ambientes de comunicación amplios como internet, con la ventaja de que mediante el uso de la emulación los parámetros mencionados anteriormente puedan ser manipulados en función del tipo y cantidad de restricciones que se desea emular en el comportamiento de la red. Por tal razón, se utilizó la herramienta NetEM en su versión 2.6 para emular un ambiente de red restringida entre el gateway y el servidor de gestión de datos.

Para la emulación de una ruta de Internet con pérdida de paquetes variables, latencia, reordenamiento y jitter se utilizó como base la configuración obtenida de los ejemplos de NetEM [16], que emula las características de conexión de un enlace de Internet desde un ISP de bajo nivel operando en Sudamérica hacia los servidores de un ISP de la costa este de los Estados Unidos, con lo cual se establecieron los siguientes parámetros:

- El retardo de latencia se establece en 100ms con una variación de +/- 40ms con un 25% de correlación con el último paquete enviado.
- La pérdida de paquetes se establece en un 9% de que los paquetes se eliminen aleatoriamente y se pierdan, cada probabilidad sucesiva depende en un 25% de la última.
- La duplicación de paquetes se estable en 1% del total de paquetes enviados.
- La corrupción de paquetes se establece en el 0.1% del tráfico, con lo cual se introduce un solo error de bit en el desplazamiento aleatorio del paquete.
- El reordenamiento de paquetes se establece en que el 5% de los paquetes con una correlación del 50% se enviará inmediatamente.

3.2 Implementación de MQTT-SN en el escenario de red de pruebas

Para la implementación del protocolo MQTT-SN dentro del escenario de pruebas se utilizó la arquitectura de gateway agregado ofrecida por el protocolo. En la cual los nodos de la WSN se comunican hacia el gateway a través de conexiones con MQTT-SN y posteriormente el gateway establece su conexión hacia el servidor de gestión de datos mediante MQTT. En la figura 3.2, se visualiza el esquema de comunicación entre todos los elementos de la red. Los nodos sensoriales actuarán como clientes MQTT-SN (agentes de publicación de información). El gateway de interconexión opera como agente intermediario de comunicación traduciendo los mensajes MQTT-SN de la WSN a mensajes MQTT para su envío a través de la red TCP/IP hacia el servidor de gestión de datos en el cual se implementó un servicio de Broker MQTT y un agente de suscripción para recibir los datos generados por la red de sensores inalámbricos.

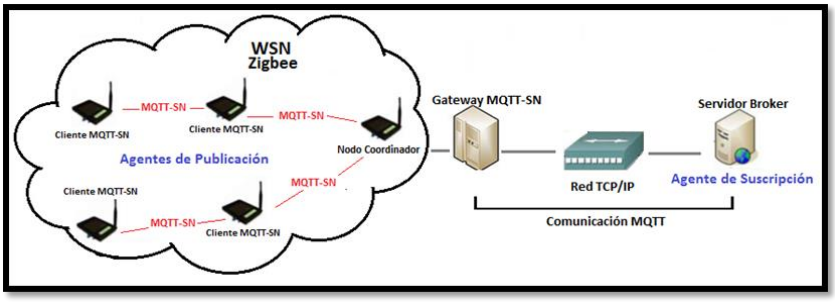


Figura 3.2. Implementación MQTT-SN en escenarios de red de pruebas
Fuente: El Autor

La comunicación de extremo a extremo entre el cliente MQTT-SN o agente de publicación y el agente de suscripción implementado en el servidor de gestión de datos se llevará a cabo mediante la especificación de un "topic" o tema en común gestionado por el protocolo, el cual será utilizado por los nodos sensoriales para publicar sus datos, así como también lo utilizará el servidor para recibir toda la información. La información fue publicada mediante los mecanismos de confiabilidad de QoS 0 y 1 brindados por el protocolo.

3.3 Implementación de CoAP en el escenario de red de pruebas

Para la implementación del protocolo CoAP dentro del escenario de pruebas se utilizó la arquitectura del modelo "observador" ofrecida por el protocolo. En la cual se facilita un mecanismo de interacción publicación/suscripción, donde el protocolo CoAP permite a un cliente observar constantemente los eventos de un sujeto registrando su interés en sus eventos por medio de una solicitud GET extendida y la opción OBSERVE. Para la obtención de datos de la red de sensores inalámbricos, el cliente observador registra su interés en los cambios de eventos dentro de la WSN, los nodos sensoriales cumplen el papel de sujetos de observación publicando sus recursos a través de un proxy intermediario que realiza la tarea de unificación de los nodos sensoriales de la red inalámbrica hacia la red tradicional TCP/IP donde se encuentra el cliente observador. En la figura 3.3, se visualiza el esquema de comunicación entre todos los elementos de la red, donde los nodos sensoriales actúan como sujetos de observación (agentes de publicación de eventos), el servidor de gestión de datos opera como nodo observador (agente de suscripción de eventos) y el gateway de interconexión funciona como proxy intermediario de comunicación encargándose de multiplexar la información de los sujetos de observación con el interés del nodo observador a través del protocolo CoAP.

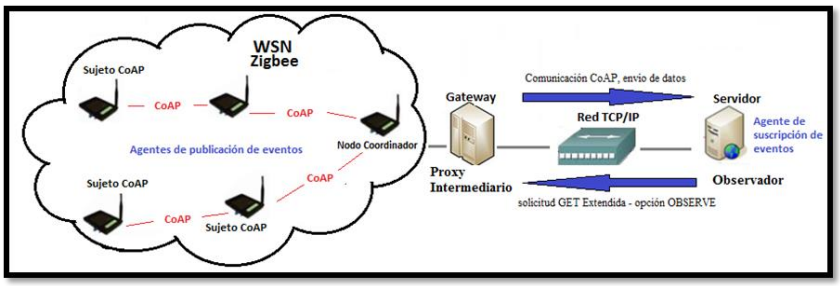


Figura 3.3. Implementación CoAP en escenarios de red de pruebas
Fuente: El Autor

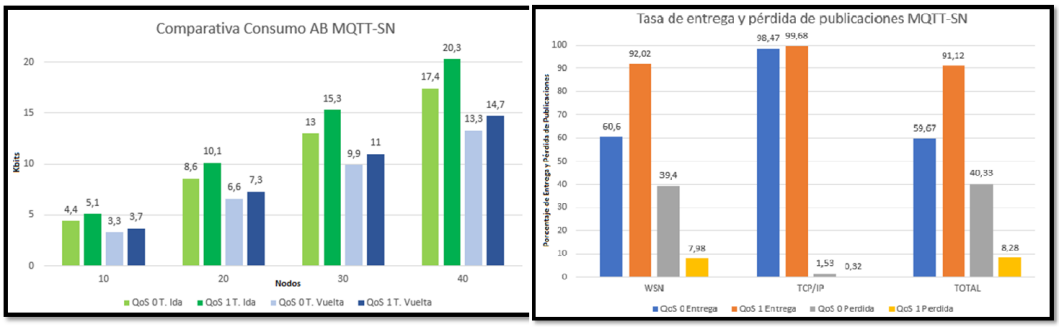
La información fue publicada mediante los mecanismos de confiabilidad de CoAP a través de mensajes no confirmables de tipo NON y mensajes confirmables de tipo CON brindados por el protocolo según cada escenario de prueba.

4 Resultados

En la gráfica 4.1, se puede observar la comparativa de consumo de ancho de banda promedio tanto para el tráfico de ida como el tráfico de vuelta de los mecanismos de confiabilidad QoS 0 y 1 de MQTT-SN. Teniendo como resultado que con el mecanismo de confiabilidad QoS 1 se tiene un aumento en el consumo de

ancho de banda para el tráfico de Ida de un 16,93% promedio por cada diez nodos a comparación del tráfico de ida del mecanismo QoS 0. Mientras que, para el tráfico de Vuelta, el mecanismo de QoS 1 presenta un aumento del consumo de ancho de banda del 11,09% promedio por cada diez nodos a comparación del tráfico de vuelta del mecanismo QoS 0.

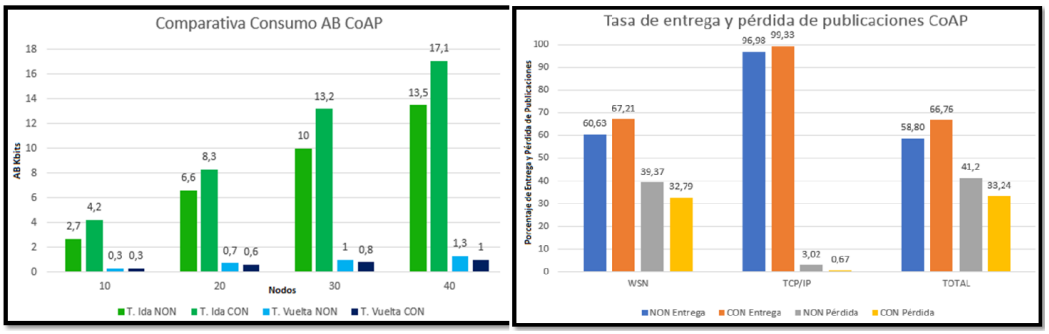
También se puede observar la comparativa del porcentaje de entrega y pérdida de publicaciones entre los mecanismos de QoS 0 y 1 de MQTT-SN. Teniendo como resultado que con el mecanismo de confiabilidad QoS 1 se tiene un mejor porcentaje de entrega de publicaciones a nivel general, habiendo un aumento significativo del porcentaje de entrega del 31,42% en la WSN, 1,21% en la red TCP/IP y un 31,45% en el total de publicaciones entregadas a comparación del mecanismo QoS 0.



Gráfica 4.1. Comparativa de variables MQTT-SN QoS 0 y 1
Fuente: El Autor

En la gráfica 4.2, se puede observar la comparativa de consumo de ancho de banda promedio tanto para el tráfico de ida como el tráfico de vuelta de los mecanismos de entrega de mensajes tipo NON y CON de CoAP. Teniendo como resultado que con el mecanismo de confiabilidad de entrega de mensajes CON se tiene un aumento en el consumo de ancho de banda para el tráfico de Ida de un 34,99% promedio por cada diez nodos a comparación del tráfico de ida del mecanismo de entrega de mensajes NON. Mientras que, para el tráfico de Vuelta, el mecanismo de entrega de mensajes NON presenta un aumento del consumo de ancho de banda del 17,92% promedio por cada diez nodos a comparación del tráfico de vuelta del mecanismo de entrega de mensajes CON.

También se puede observar la comparativa del porcentaje de entrega y pérdida de publicaciones entre los mecanismos de entrega de mensajes NON y CON de CoAP. Teniendo como resultado que con el mecanismo de entrega de mensajes tipo CON se tiene un mejor porcentaje de entrega de publicaciones a nivel general, habiendo un aumento del porcentaje de entrega del 6,58% en la WSN, 2,35% en la red TCP/IP y un 7,96% en el total de publicaciones entregadas a comparación del mecanismo de entrega de mensajes tipo NON.

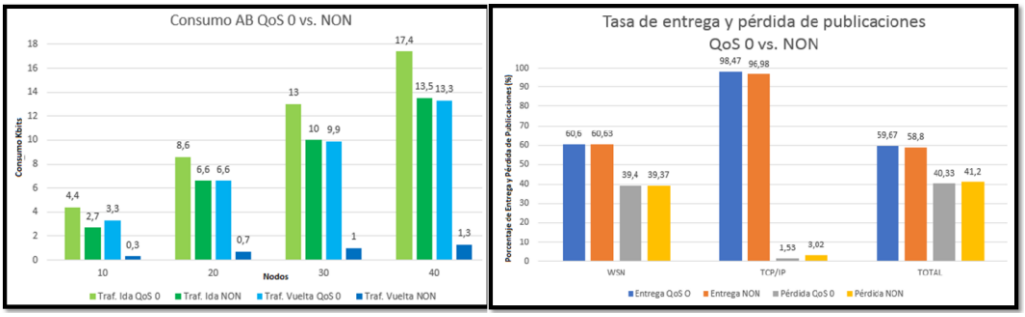


Gráfica 4.2. Comparativa de variables CoAP NON y CON
Fuente: El Autor

En la gráfica 4.3, se puede observar la comparativa de consumo de ancho de banda promedio tanto para el tráfico de ida como el tráfico de vuelta de los mecanismos de confiabilidad QoS 0 y NON de cada protocolo. Teniendo como resultado que con el mecanismo de confiabilidad NON se tiene un consumo de ancho

de banda menor de un 26,85% promedio para el tráfico de ida y de un 90,11% promedio para el tráfico de vuelta por cada 10 nodos publicadores a comparación del mecanismo de confiabilidad QoS 0.

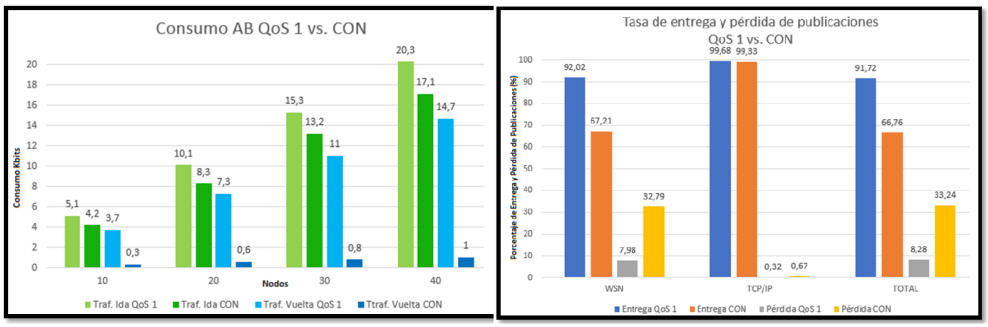
También se puede observar la comparativa del porcentaje de entrega y pérdida de publicaciones entre los mecanismos de confiabilidad QoS 0 y NON de cada protocolo. Teniendo como resultado que con los dos mecanismos de entrega de mensajes se tienen porcentajes de entrega de publicaciones similares a nivel de la WSN, habiendo un ligero aumento del porcentaje de entrega por parte del mecanismo QoS 0 del 1,49% en la red TCP/IP y un 0,87% en el total de publicaciones entregadas a comparación del mecanismo de entrega de mensajes tipo NON.



Gráfica 4.3. Comparativa de variables QoS 0 y NON
Fuente: El Autor

En la gráfica 4.4, se puede observar la comparativa de consumo de ancho de banda promedio tanto para el tráfico de ida como el tráfico de vuelta de los mecanismos de confiabilidad QoS 1 y CON de cada protocolo. Teniendo como resultado que con el mecanismo de confiabilidad CON se tiene un consumo de ancho de banda menor de un 16,24% promedio para el tráfico de ida y de un 92,40% promedio para el tráfico de vuelta por cada 10 nodos publicadores a comparación del mecanismo de confiabilidad QoS 1.

También se puede observar la comparativa del porcentaje de entrega y pérdida de publicaciones entre los mecanismos de confiabilidad QoS 1 y CON de cada protocolo. Teniendo como resultado que con los dos mecanismos de entrega de mensajes se tienen porcentajes de entrega de publicaciones similares a nivel de la red TCP/IP habiendo un ligero aumento del porcentaje de entrega por parte del mecanismo QoS 1 del 0,35%, mientras que en la WSN se tiene un denotado mayor porcentaje de entrega con un 24,81% y un 24,96% en el total de publicaciones entregadas a comparación del mecanismo de entrega de mensajes tipo CON.



Gráfica 4.4. Comparativa de variables QoS 1 y CON
Fuente: El Autor

Conclusiones

En base a los resultados del análisis de rendimiento de los protocolos de Publicación/Subscripción MQTT-SN y CoAP en términos de consumo de ancho de banda, tasa de entrega y pérdida de publicaciones en su operación con una Red de Sensores Inalámbricos Zigbee, se han determinado las siguientes conclusiones:

El uso de los mecanismos de confiabilidad con acuse de recibo (QoS 1 y CON) a comparación de los mecanismos best-effort (QoS 0 y NON) de cada protocolo produce una disminución en el rendimiento del ancho de banda, visibilizándose un mayor consumo de este. En cambio, se obtiene un mejor rendimiento en el porcentaje de la tasa de entrega de publicaciones en general para todos los escenarios de pruebas con diferentes valores dependiendo del rendimiento de cada protocolo.

La implementación del mecanismo de acuse de recibo de publicaciones en el protocolo de aplicación MQTT-SN produce un impacto poco significativo en el aumento de consumo de ancho de banda de la comunicación. Teniendo según los resultados que con el mecanismo QoS 1 se tiene un aumento del 16,93% y 11,09% promedio del tráfico de ida y vuelta respectivamente a comparación del mecanismo QoS 0. Mientras que para la tasa de entrega de publicaciones existe un aumento significativo del porcentaje de entrega del 31,42% en la WSN, 1,21% en la red TCP/IP y un 31,45% en el total de publicaciones entregadas con QoS 1 a comparación de QoS 0.

La implementación del mecanismo de acuse de recibo de publicaciones en el protocolo de aplicación CoAP produce un impacto poco significativo en el aumento de consumo de ancho de banda de la comunicación. Teniendo según los resultados que con el mecanismo CON se tiene un aumento del 34,99% promedio del tráfico de ida a comparación del mecanismo NON. Mientras que para la tasa de entrega de publicaciones existe un aumento mínimo poco significativo del porcentaje de entrega del 6,58% en la WSN, 2,35% en la red TCP/IP y un 7,96% en el total de publicaciones entregadas a través de CON a comparación de NON.

En la interconexión de la WSN con tecnología Zigbee hacia la red tradicional, la arquitectura de gateway agregado brindada por MQTT-SN destaca sobre la arquitectura de proxy intermediario que brinda el protocolo CoAP, ya que esta ofrece el apoyo del protocolo durante toda la comunicación de extremo a extremo, obteniendo un mayor nivel de independencia que tiene cada nodo sensorial en publicar su información por separado directamente hacia el servidor de gestión a comparación de la arquitectura de CoAP en la cual el protocolo no cuenta con soporte para una comunicación de extremo a extremo con este tipo de tecnologías, produciendo una falta de acompañamiento por parte del protocolo que afecta directamente al rendimiento de este en el manejo de las publicaciones de los nodos WSN, las cuales deben ser gestionadas de manera centralizada a través del proxy como agente intermediario.

Los mecanismos de confiabilidad QoS 0 y 1 de MQTT-SN presentan un mayor consumo de ancho de banda que los mecanismos NON y CON de CoAP debido a la cantidad de bits que utilizan para sus procesos de señalización en la comunicación de extremo a extremo con la WSN, recalcando que MQTT-SN utiliza TCP como protocolo de transporte por lo cual requiere una mayor señalización tanto para el tráfico de ida como de vuelta a comparación del protocolo CoAP que opera sobre UDP reduciendo la cantidad de paquetes enviados o recibidos durante la comunicación. En los resultados se visibiliza que para comunicaciones sin acuse de recibo a nivel de aplicación el mecanismo NON tiene un consumo de ancho de banda menor de un 26,85% y 90,11% promedio para el tráfico de ida y vuelta respectivamente a comparación del mecanismo QoS 0 y para comunicaciones con acuse de recibo a nivel de aplicación el mecanismo CON presenta un consumo de ancho de banda menor de un 16,24% y 92,40% promedio para el tráfico de ida y vuelta respectivamente a comparación del mecanismo QoS 1. Mientras que en la tasa de entrega de publicaciones el uso de TCP como protocolo de transporte por parte de MQTT introduce un pequeño aporte en el éxito de entrega de publicaciones dentro de la red tradicional a comparación de UDP utilizado por CoAP.

En base al aumento del número de nodos que realizan publicaciones dentro de la WSN se debe considerar que, aunque MQTT-SN demuestre mayores consumos de ancho de banda a comparación de CoAP, el protocolo MQTT-SN presenta un rendimiento más equilibrado en el uso del ancho de banda, teniendo picos de consumo máximo menores al 50% del valor de consumo de ancho de banda promedio en función al aumento del número de nodos WSN que realizan publicaciones con los mecanismos de QoS 0 y 1. Mientras que el protocolo CoAP presenta picos de consumo máximo desproporcionales mayores al 80% del valor de consumo de ancho de banda promedio con el mecanismo NON y picos de consumo máximo desproporcionales mayores al 150% del valor de consumo de ancho de banda promedio con el mecanismo CON en función al aumento del número de nodos WSN que realizan publicaciones.

En cuanto a la tasa de entrega, retransmisión y perdida de publicaciones, la evaluación de los resultados muestra que el mecanismo de confiabilidad con acuse de recibo de extremo a extremo juega un papel importante en todos los escenarios de red evaluados, principalmente en el mejoramiento de la tasa de entrega de publicaciones en el segmento de red perteneciente a la WSN, en la cual se evidencia el mayor porcentaje de pérdida. Teniendo como resultado que el mecanismo QoS 1 de MQTT-SN tiene el mejor porcentaje de entrega de publicaciones con un

91,72% sobre el 66,76% de entrega del mecanismo CON de CoAP, denotándose que la principal desventaja del mecanismo CON de CoAP es no contar con soporte de acuse de recibo por parte del protocolo en el segmento de Red que presenta mayores pérdidas.

La diferencia existente en el rendimiento de cada protocolo se puede aprovechar para mejorar el rendimiento en general de diferentes tipos de implementaciones de red, ya que en base a esta información se puede decidir que protocolo utilizar según las condiciones de la red predominantes. Dentro de las comunicaciones M2M, las aplicaciones pueden requerir optimizar el consumo de ancho de banda o a su vez dar prioridad a la entrega de publicaciones sobre otras métricas.

Finalmente, se concluye que el protocolo de Publicación/Subscripción más adecuado para el diseño e implementación de un sistema de comunicación extremo a extremo entre una Red de Sensores Inalámbricos Zigbee y la red de servicios tradicional bajo las métricas de consumo de ancho de banda y tasa de entrega de publicaciones es el protocolo MQTT-SN bajo su mecanismo de confiabilidad QoS 1 por sus resultados de rendimiento en los cuales se denoto un consumo de ancho de banda equilibrado y de crecimiento proporcional en función al aumento de nodos publicadores, así como también ser el mejor mecanismo para lograr una mayor tasa de entrega de publicaciones, disminuyendo significativamente la perdida de publicaciones dentro del segmento de la WSN, considerada la parte más sensible de la comunicación de extremo a extremo.

Referencias

1. Zigbee Alliance, TECNOLOGIA ZIGBEE. [en línea], ultima fecha de acceso: 28/05/2016. Disponible en: http://www.zigbee.org/
2. Eugster P. T., Felber P. A., Guerraoui R., Kermarrec A., (2003). “The many faces of publish/subscribe”. ACM Comput. Surv. vol. 35, no. 2, Pp. 114–131
3. Souto E., Guimarães G., Vasconcelos G., Vieira M., Rosa N., Ferraz C., (2006). “A publish/subscribe middleware for sensor networks”. Personal and Ubiquitous Computing., vol 10, no. 1, Pp. 37–44.
4. Hunkeler U., Truong H. L., Stanford-Clark A., (2008). “MQTT-SN — A publish/subscribe protocol for Wireless Sensor Networks". Communication Systems Software and Middleware and Workshops. COMSWARE 2008. 3rd International Conference, Bangalore. Pp. 791-798.
5. MQ Telemetry Transport Standardization Announce. (2011). [en línea], ultima fecha de acceso: 28/05/2016. Disponible en: http://mqtt.org/2011/08/open-invitation-to-join-the-mqtt-standardization-discussion
6. Using WebSphere MQ Telemetry and Pachube to connect to Remote Sensors and Devices. (2012). [en línea]. Ultima fecha de acceso: 28/05/2016. Disponible en:
https://www.ibm.com/developerworks/websphere/library/techarticles/1106_maynard/1106_maynard.html
7. Building Facebook Messenger. (2011). [en linea], ultima fecha de acceso: 28/05/2016. Disponible en: http://www.facebook.com/notes/facebook-engineering/building-facebook-messenger/10150259350998920
8. Stanford-Clark A., Truong H. L., (2013). “MQTT for sensor networks (MQTT-SN) Protocol Specification Version 1.2”. Disponible en: http://mqtt.org/new/wp-content/uploads/ 2009/06/MQTT-SN_spec_v1.2.pdf
9. Core Working Group., (2012). “Constrained Application Protocol (CoAP), Draf-Ietf-Core-Coap-11”. [en linea], última fecha de acceso: 28/05/2016. Disponible en: http://tools.ietf.org/html/draft-ietf-core-coap-11
10. Constrained RESTful Environments (core). [en linea], ultima fecha de acceso: 28/05/2016. Disponible en: http://datatracker.ietf.org/wg/core/charter/.
11. Internet Engineering Task Force (IETF). (2015). “Observing Resources in the Constrained Application Protocol (CoAP)”. ISSN: 2070-1721., [en linea], última fecha de acceso: 28/05/2016. Disponible en: https://tools.ietf.org/html/rfc7641.
12. Wang C., Sohraby K., Li B., Daneshmand M., Hu Y. M., (2006). “A survey of transport protocols for wireless sensor networks”. IEEE Network, vol. 20, no. 3, pp. 34-40.
13. Jones J., Atiquzzaman M., (2007). “Transport protocols for wireless sensor networks: State of the art and future directions”. International Journal of Distributed Sensor Networks. Pp. 119–133.
14. Escobar Días M. S., (2009). “Wireless Sensors Network, Estado del arte e Investigación”, Computer Science and Engineering Department, Universidad Carlos III de Madrid, [en línea], ultima fecha de acceso: 28/05/2016. Disponible en: http://arcos.inf.uc3m.es/~sescolar/index_files/presentacion/wsn.pdf
15. Bormann C., Mulligan G., Arkko J., Townsley M., Schumacher C., (2007). “IPv6 over Low power WPAN (6lowpan)”, IETF Working Group. Disponible en: https://datatracker.ietf.org/wg/6lowpan/documents/
16. NetEM Network Emulator, Network Latency and Packet Loss Emulation, [en linea], última fecha de acceso: 28/05/2016. Disponible en: https://wiki.linuxfoundation.org/networking/netem

Gestión de Calidad Móvil: Contratos, QoS y QoE

Juan Enriquez[1], Antonio Collazo García[2], Ariel Machini[1], Sandra Casas[1]

[1]Instituto de Tecnología Aplicada, Universidad Nacional de la Patagonia Austral,
[2]Centro de Investigaciones y Transferencias Santa Cruz, CONICET
Piloto Rivero S/N, Río Gallegos, Argentina
{jenriquez,scasas}@unpa.edu,ar

Abstract. La gestión de la calidad, la medición y evaluación, el monitorio, la optimización de la QoS / QoE de los servicios móviles en sus diversos aspectos, es un problema actual. Este trabajo explora y analiza la integración de QoS y QoE móvil en contratos, que permitan especificar las necesidades y/o preferencias de los usuarios de las aplicaciones y a la vez puedan ser ejecutados en los dispositivos. Para ello presentamos mecanismos para la especificación e implementación de los contratos y la composición de los mismos con las aplicaciones móviles. En particular se cubre la plataforma Android.

Keywords: Aplicaciones Móviles, Calidad de Servicio, Calidad de Experiencia, Contratos.

1 Introducción

Los dispositivos móviles se vuelven cada vez más populares y potentes, esto se debe a múltiples factores, entre ellos la capacidad de ejecutar aplicaciones de contenidos multimediales y por otro lado, a partir del avance de las tecnologías de redes inalámbricas de banda ancha, lo que también ha sido denominado como multimedia móvil.

Sin embargo, las aplicaciones y/o servicios móviles están expuestos a diversos y múltiples factores que afectan no solo su desempeño (calidad de servicio – QoS), sino también la satisfacción del usuario final (calidad de experiencia - QoE). Ante un inaceptable desempeño en la ejecución de las aplicaciones móviles y una experiencia no satisfactoria del usuario, resulta común que no se utilicen o se abandonen las aplicaciones. Algunos factores ampliamente reconocidos como causantes de los problemas de desempeño se encuentran, canales físicos poco fiables y con ancho de banda limitado, movilidad de nodos, enrutamiento, limitaciones de recursos (la CPU, la batería y memoria de los dispositivos, tamaño de pantalla y métodos de entrada), métricas de evaluación, condiciones climáticas [1][2][3].

En este escenario, sumamente heterogéneo y cambiante, una de las áreas de mayor interés entre la comunidad científica, las prestadoras de servicios, los desarrolladores de aplicaciones móviles y los usuarios, es la calidad. La gestión de la calidad, la medición y evaluación, el monitorio, la optimización de la QoS / QoE de los servicios móviles en sus diversos aspectos, es un problema actual.

La literatura y la industria han hecho distinciones entre QoS y QoE, no obstante, se proponen enfoques de tratamiento más integrales [4][5][6][7] que desde la hipótesis de que una QoS degradada generará un QoE inaceptable, admiten sus estrechas relaciones causales, ya que la QoS ofrece la infraestructura subyacente de base para la medición de QoE, es así que la QoS define las características no funcionales de un sistema, afectando la calidad percibida de los resultados.

Una visión más amplia de QoS, la determina como el conjunto de aspectos no funcionales del sistema que determinan el nivel de satisfacción de los usuarios de la funcionalidad ellos proveen [8], con el fin de cumplir con el objetivo general de obtener niveles de QoS que coincidan con las necesidades de los usuarios y así impacten en la QoE. Esta forma de entender la QoS implica una traducción de nivel superior de los requisitos de QoS a nivel de usuario a los requisitos de QoS de nivel inferior.

En esta perspectiva entran a jugar mayor cantidad de factores, propiedades, métricas, parámetros e indicadores, que atraviesan varias de capas del ecosistema móvil. Esta necesidad se convierte en una tendencia por definir mecanismos para permitir que las aplicaciones declaren sus necesidades en términos de QoE (objetivo), resultando coherente con los requisitos 5G expresados por NGMN Alliance en [9], que afirma que el próximo sistema 5G debe ser consciente de los usuarios finales y las necesidades de QoE de las aplicaciones, midiéndolas donde se percibe (es decir, en los dispositivos) [10].

Este trabajo explora y analiza la integración de QoS y QoE móvil en contratos, que permitan especificar las necesidades y/o preferencias de los usuarios de las aplicaciones y a la vez puedan ser ejecutados en los dispositivos. Para ello presentamos mecanismos para la especificación e implementación de los contratos y la composición de los mismos con las aplicaciones móviles. En particular se cubre la plataforma Android.

2 QoS y QoE Móvil

Los antecedentes de gestión de la calidad, remiten casi unívocamente a la QoS, ámbito en el cual se definen con mayor precisión funciones, fases, elementos y componentes. La gestión de QoS se puede considerar como un área especializada de la administración de sistemas distribuidos, y refiere a la supervisión y el control necesarios para garantizar que se alcancen y se mantengan las propiedades de calidad de servicio deseadas, que se aplica tanto a las interacciones continuas de los medios como a las interacciones discretas [11]. En el contexto de la computación móvil, la Gestión de la QoS se refiere al conjunto de actividades dedicadas a la vigilancia y control de los recursos involucrados en la provisión de un nivel de QoS adecuado. La Gestión de QoS es una tarea global en el sentido de que tiene en cuenta todo tipo de recursos que soportan el servicio (s) proporcionado (s) por un sistema: recursos de red, recursos del sistema, recursos de middleware, aplicaciones, etc., con el fin de cumplir con el objetivo general de obtener niveles de QoS que coincidan con las necesidades de los usuarios. Esta forma de entender la QoS se conoce como “end-to-end QoS” e implica una traducción de nivel superior de los requisitos de QoS a nivel de usuario a los requisitos de QoS de nivel inferior [8].

Los diversos aspectos de la interacción y los tipos de garantías requeridos, deben sintetizarse en una especificación de requisitos y en las relaciones de compensación para permitir la gestión de la QoS entregada [12], estas especificaciones a menudo constituyen "contratos". La gestión de QoS se suele dividir en funciones o fases, entre las cuales básicamente se distinguen dos fases o tipos de funciones: estáticas (definición del contrato que consiste en la especificación de requerimientos, capacidades, indicadores, parámetros, de carácter principalmente cuantitativos, especificación de las reglas de negociación/re-negociación: acciones que se ejecutarán ante diversas circunstancias, y reserva de recursos en los casos que se requiera) y dinámicas (los aspectos dinámicos de la gestión de la calidad de servicio responden al cambio dentro del entorno, permitiendo que un contrato se cumpla de forma continua). Las funciones estáticas, aplicadas al inicio de una interacción, y funciones dinámicas, aplicadas según sea necesario durante una interacción (sesión u ejecución del sistema).

En cuanto a la QoE, no existe un andamiaje de "gestión" como conjunto de funciones, sino a tareas centradas en la evaluación, monitorio y/o predicción, basándose en métricas objetivas y/o subjetivas. Existen varios métodos para cuantificar las métricas subjetivas [13][14][15][16]. Las métricas subjetivas se encuentran relacionadas a la opinión del usuario, evaluando la calidad de la aplicación o servicio a partir de su experiencia con la aplicación [15], dependiendo de factores subjetivos como el contexto y expectativas del usuario sobre la aplicación. En general, las métricas subjetivas se basan en la realización de encuestas y cuestionarios a los usuarios para la obtención de información, presentando la dificultad de ser costosas en tiempo y dinero [4][17].

Con respecto a las métricas objetivas se destaca su carácter sistemático, exacto y repetible, y refieren a distintas propiedades como, el tiempo de presentación de los datos, que mide el tiempo transcurrido entre la acción del usuario y la respuesta de la aplicación, el consumo de los datos móviles, consumo de energía, interfaz de usuario y presentación del contenido [15]. Varios estudios analizan la QoE en estrecha relación con la QoS, en tanto que una degradación en la QoS provoca una inaceptable QoE. Parámetros de QoS como retardo, inestabilidad, tasa de pérdida, tasa de error, ancho de banda y tasa de éxito de la señal se usan para determinar el valor de la QoE, dependiendo únicamente del cálculo de la QoS [7]. La QoS se compone de un conjunto de métricas relacionadas a los parámetros de rendimiento de la red que representan un impacto en la QoE. La cantidad de datos que se recibe del servicio como respuesta, el tiempo transcurrido entre el envío y la recepción de los datos y la pérdida de información, son algunas de las métricas asociadas a la QoS [4].

3 Integrando QoS y QoE en contratos

En esta Sección se presenta el enfoque para la gestión de calidad, a partir de la integración de QoS y QoE en contratos. La misma cubre las fases estática y dinámica de la gestión.

3.1 Definición de Contratos de Calidad

La definición del contrato, estará supedita al concepto subyacente de enfoque de calidad que se pretende administrar. La Figura 1, representa el enfoque adoptado. La calidad como un todo (parte izquierda), que a la vez reconoce como partes componentes, la QoS y la QoE, cada una de estas propone sus propias propiedades de análisis (parte derecha).

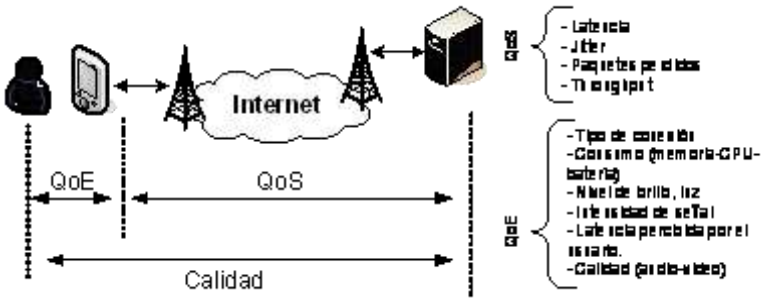


Fig.1: Enfoque de Gestión de Calidad

Así en términos generales y genéricos (Fig.2), un contrato es una lista o conjunto de tuplas, cada tupla establece una relación entre una propiedad de calidad (QoS o QoE) requerida y una acción a realizar si la propiedad no se cumple. Una propiedad requerida, y un valor (un valor específico o en un rango) que se define estáticamente pero se controla dinámicamente.

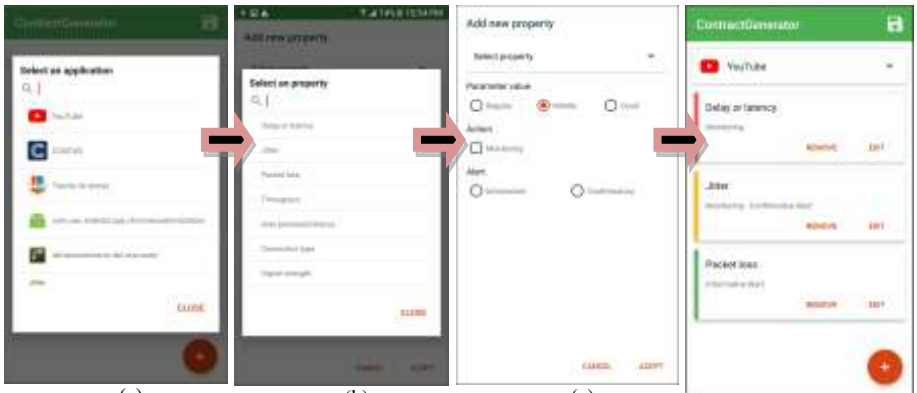
```
MPropiedades  =  {latencia,  jitter,  paquetes_perdidos,  conexion,
                 consumo_memoria,    consumo_bateria,    consumo_CPU,
                 nivel_brillo,      nivel_luz,      intensidad_señal,
                 latencia_percibida, calidad_audio, calidad_video,…}
MAcciones = {no_accion, monitorear, controlar_flujo, almacenar_datos,
            restringir_calidad, reservar_recurso, reasignar_recurso,
            enviar_advertencia, suspender_act, abortar_otras_act}
MContrato = {(p, a): p ∈ MPropiedades y a ∈ MAcciones}
```

Una lista posible de acciones a realizar se han tomado de [18], ninguna acción, monitorear la QoS/QoE lograda, controlar el flujo de información, actuar para restringir la QoS/QoE lograda, reservar o reasignar recursos, advertencia de que se están cruzando límites o umbrales, suspender o abortar la actividad, suspender o abortar otras actividades que compiten por los recursos.

3.2 Implementación de Contratos

La especificación del contrato se realiza desde una aplicación móvil instalada en el dispositivo, la cual es independiente del resto de las aplicaciones de usuario. La Figura 2, presenta la secuencia del proceso. Primero, debe elegirse la aplicación sobre

la cual se definirá el contrato (2.a). El contrato es particular para la aplicación que se pretende controlar, en razón que distintas aplicaciones requieren distintas propiedades (parámetros y valores) y a la vez diferentes acciones. A continuación, se selecciona las propiedades que formará parte del contrato, para lo cual, se presenta una lista de las mismas (2.b). A medida que el usuario selecciona las propiedades, indica la cualidad o desempeño que debe cumplir (regular, medio, bueno) y la acción a realizar si el valor no se logra (monitoreo, alerta informativa y alerta confirmatoria) (2.c).



(a) (b) (c) (d)

Fig. 3. App ContractGenerator.

La especificación de los valores de las propiedades, se realiza a partir de categorías cualitativas, que tienen su correspondiente valor numérico asignado o un indicador, en la Tabla 1 se presenta un subconjunto de las 11 que han sido aplicadas. Mientras que solo un valor de propiedad puede ser asignado a una propiedad, más de una acción puede ser relacionada a la propiedad. En (2.d) se presenta un contrato definido para YouTube que especifica las propiedades Latencia, Jitter y Paquetes perdidos.

Tabla 1. Propiedades, Valores cualitativos, indicadores y umbrales.

Propiedad	Bueno	Medio	Regular
Latencia	<100 ms	<= 200 ms	>200 ms
Jitter	< 50ms	<= 60 ms	>60
Paquetes perdidos	< 0.001	<= 0,002	>0,002
Tipo de Conexión	4G - WiFi	3G	GSM
Intensidad de Señal	40 a 60 dBm	60 a 80 dBm	>80dBm
Latencia percibida por el usuario	<100 ms	<=200 ms	>200 ms
Consumo de memoria	<100 um	<=200 um	>200

La especificación se almacena en formato XML, la Figura 4, presenta la estructura del contrato y una instanciación en la que se ha especificado para una aplicación diversas propiedades, valores y acciones.

<?xml version="1.0" encoding="utf-8"?> <contract app="com.google.android.youtube" date="2018-07-02 17:19:37">	<?xml version="1.0" encoding="utf-8"?> <contract app="com.google.android.youtube" date="2018-07-02 17:19:37">

```
<property name="ScreenBrightness |
                ConnectionType |
                SignalStrength | Jitter |
                Latency |
                UserPerceivedLatency |
                LightEnvironment |
                PacketLoss | Proximity |
                MemoryConsumption |
                CPUConsumption ">
    <range> regular | middle | bad </range>
    <actions>
      <monitoring/>
      <informative/>
      <confirmatory/>
    </actions>
  </property>
</contract>
```

```
<property name="ScreenBrightness">
    <range>regular</range>
     <actions>
       <confirmatory/>
     </actions>
</property>
<property name="Latency">
    <range>middle</range>
     <actions>
       <monitoring/>
     </actions>
</property>
<property name="Proximity">
    <range>bad</range>
    <actions>
       <monitoring/>
       <informative/>
    </actions>
</property>
</contract>
```

Fig. 4. Contrato e instanciación.

La composición del contrato, es el mecanismo que permite que un contrato se aplique o ejecute durante la interacción del usuario con la aplicación, es decir, permite llevar adelante la fase dinámica de la gestión de la calidad. Para ello, se ha adoptado un enfoque de separación de concerns y la modularización de las tareas correspondientes a la fase dinámica se han diseñado e implementado en AspectJ[19]. AspectJ es una extensión de Java que permite implementar requerimientos no funcionales y/o de calidad, como en este caso, a partir de una serie de mecanismos tales como pointcuts y advices y un proceso de tejido. De esta forma, un aspecto puede interceptar eventos (por ejemplo métodos) de clases correspondientes a la funcionalidad central de un sistema para incorporar nueva funcionalidad. En este caso, el aspecto ContractCheck, encapsula toda la lógica de la verificación de las propiedades y valores definidos en el contrato y la ejecución de las acciones definidas si corresponde (Fig. 5).

```
@Aspect
public class ContractCheck
```

```
@Pointcut("execution(* *.*.onCreate(..))")
  public void metodo() {}
```

```
@Before("metodo()")
     // Leer el contrato
     // Para cada propiedad definida calcular el valor actual
    // Verificar valores del contrato con los valores calculados
    // Ejecutar la acción si no se cumple la restricción
```

Fig. 5. Aspecto ContractCheck.

Para ello, un pointcut intercepta cada vez que se crea una actividad y previamente a partir de un aviso “before” (fragmento de código que se ejecutará antes que la creación de la actividad se ejecute), realiza las operaciones necesarias: leer el contrato, analizar que propiedades y valores establecidos, y comparar con el cálculo

dinámico de dichas propiedades. Para los casos que no se cumpa la especificación, se dispara la acción indicada. Para el cálculo del valor de cada propiedad en ejecución, que es contrastada con los valores de las propiedades definidas en el contrato, hemos desarrollado una biblioteca de propiedades, que implementa los diversos algoritmos.

El listado de la Figura 6 es un recorte del fichero XML generado en una de las pruebas que hemos realizado, cuyo contrato especificaba la acción monitorear para las propiedades latencia, latencia percibida por el usuario, tipo de conexión, etc. Se registra la propiedad controlada, la fecha y hora del control y el valor calculado, el cual difiere del especificado.

```
<property id="latencia" date="2018-07-05 20:22:15">150.0</property>
<property id="latenciaPercibidaSegundos" date="2018-07-05 20:22:15">0</property>
<property id="brilloPantalla" date="2018-07-05 20:22:42">12</property>
<property id="tipoConexion" date="2018-07-05 20:22:42">3G</property>
<property id="brilloPantalla" date="2018-07-05 20:26:47">12</property>
<property id="tipoConexion" date="2018-07-05 20:26:47">3G</property>
<property id="latenciaPercibidaSegundos" date="2018-07-05 20:26:47">0</property>
<property id="brilloPantalla" date="2018-07-05 20:30:59">12</property>
<property id="tipoConexion" date="2018-07-05 20:30:59">GSM</property>
<property id="latenciaPercibidaSegundos" date="2018-07-05 20:31:00">1242</property>
<property id="brilloPantalla" date="2018-07-05 20:34:21">12</property>
<property id="tipoConexion" date="2018-07-05 20:34:21">GSM</property>
<property id="latenciaPercibidaSegundos" date="2018-07-05 20:34:21">0</property>
<property id="luzEntornoLx" date="2018-07-05 20:34:21">0.0</property>
<property id="brilloPantalla" date="2018-07-05 20:35:01">12</property>
```

Fig. 6. Monitoreo de un Contrato.

5. Trabajos Relacionados

A continuación se detallan varias herramientas (librerías y/o framework) que dan soporte a la medición de QoS o QoE móvil. Estas se centran en una de las dimensiones, proponen propiedades limitadas y en menor medida la posibilidad de definir acciones. Aunque algunas de estas propuestas consideran la participación del usuario final, el concepto de contrato no es explícito, principalmente el objetivo de la mayoría es realizar mediciones para el análisis posterior de las dimensiones.

El trabajo de [20] propone un framework para Android para el desarrollo de aplicaciones QoS distribuidas con capacidades de auto-reconfiguración. Las funcionalidades se incorporan mediante la extensión de sus clases (Activities, Services e Intens). Este framework se utiliza para desarrollar estrategias para múltiples parámetros de QoS de aplicaciones. El esquema QoS adoptado define, para cada dominio de aplicación, los atributos y valores de calidad, así como también permite al usuario expresar las preferencias personales de QoS en una solicitud de servicio. Esto posibilita que los usuarios y los proveedores de servicios sean capaces de definir los requisitos y las propuestas del servicio a fin de llegar a un acuerdo sobre la prestación del mismo.

Netradar [https://www.netradar.org/] es un sistema de software para mediciones de calidad de internet móvil que soporta tanto datos móviles como redes WiFi y los resultados obtenidos se pueden observar en un mapa. Proporciona clientes de

medición para múltiples sistemas operativos y los resultados de la medición se almacenan de manera centralizada. Las propiedades que se miden son, velocidad de bajada y subida de datos, latencia e intensidad de la señal.

Fastah Network Kit [https://getfastah.com/] es una biblioteca para Android para llevar a cabo la estimación de calidad de una red. Es útil para aplicaciones que requieren adaptar la experiencia y comportamiento de usuario en respuesta a diferentes condiciones de la red (4G, 3G, 2G y WiFi). La latencia y la congestión se miden utilizando servidores de baja latencia. La información se almacena en los servidores. Para realizar la estimación de calidad mediante la librería se requiere agregar como dependencia la misma a la aplicación que se va a controlar y configuraciones. Si se necesita un control manual sobre las mediciones, la integración se debe realizar además sobre el código fuente de la aplicación móvil.

Existen varias herramientas para la plataforma Android, entre ellas Speedtest.net, nPerf, Speed Test Pro, Speedcheck, que apuntan a medir la calidad de la red, sin embargo son muy limitadas y están enfocadas básicamente en considerar solamente el ancho de banda de la red, proporcionan valores absolutos sin realizar ningún tipo de análisis desde el punto de vista de los usuarios finales. Estas herramientas son aplicaciones que se instalan en los dispositivos móviles para monitorizar la calidad de servicio de la conexión, experimentada en el dispositivo.

Para la medición objetiva de la QoE en forma automática existen diferentes herramientas. Las herramientas Monsoon Power Monitoring y PowerTutor [15] se basan en la medición de métricas referentes al consumo de energía de los dispositivos móviles. Prometheus [21] se enfoca en el uso de métricas objetivas con el fin de estimar la QoE en aplicaciones móviles, prestando atención en métricas asociadas al tráfico de la red como rendimiento, volúmenes de datos, tasa de perdida y otras como la intensidad de la señal. AppInsight [22] ayuda a diagnosticar fallos de rendimiento experimentados por sus aplicaciones, indicando las optimizaciones necesarias para mejorar la experiencia de usuario. Analiza las app móviles interponiéndose en los manejadores de eventos del dispositivo, teniendo en cuenta la latencia percibida por el usuario. AppInsight se enfoca en el monitoreo analizando métricas como el porcentaje de transferencia de datos extras y la cantidad de bytes por transmisión. Timecard [23] es una herramienta para estimar la QoE a partir de los tiempos de respuesta de las aplicaciones a los usuarios, basándose en el análisis de métricas como el tiempo transcurrido desde que el usuario inicia la solicitud y la estimación del tiempo que tarda en transmitir la respuesta del servidor al cliente y procesar la respuesta en el cliente. Panappticon [24] es una herramienta de monitoreo de eventos para Android que permite identificar problemas de rendimiento y diagnosticar las razones del bajo rendimiento percibido por el usuario en la realización de transacciones que inician por un toque de pantalla o pulsación de tecla y finalizan por una actualización de pantalla, midiendo el intervalo de tiempo entre la acción del usuario y la respuesta mostrada por la aplicación. Proteus [25] es una herramienta de recopilación de información sobre el rendimiento de la red, enfocada en aplicaciones de comunicación en tiempo real para predecir el rendimiento futuro de la red y la ocurrencia de pérdida de paquetes dentro de un intervalo de tiempo, analizando métricas como la tasa de pérdida, retardo unidireccional y rendimiento. QoE Doctor [17], la cual se basa en la medición de métricas objetivas como la latencia percibida por el usuario, el consumo de datos móviles y consumo de energía. Además cuenta con un conjunto de métricas para aplicaciones específicas como Facebook, Youtube y navegadores. En [26] se

describe una herramienta de evaluación capaz de estimar la QoE en términos de MOS para el servicio de YouTube. La misma permite mapear la QoS en QoE respecto de las sesiones de YouTube. Además, analiza posibles causas que conducen a una baja QoE y proporciona algunas sugerencias para mejorarla. En este trabajo, proponen un procedimiento genérico para estimar la calidad percibida por el usuario final siguiendo tres pasos: estimar o medir la QoS de la red (rendimiento, tiempo de ida y vuelta, tasa de pérdida, etc.); convertir métricas QoS de red en métricas QoS de la aplicación; y mapear la QoS a la QoE del usuario final (en términos de MOS).

6. Conclusiones

En este trabajo se presenta una propuesta para integrar QoS y QoE en contratos para aplicaciones móviles, con el objetivo de gestionar la calidad en las capas finales del ecosistema móvil. Algunas características de nuestra propuesta son, la especificación de los contratos es flexible en razón que pueden incluir distintos tipos de propiedades a controlar y acciones, se pueden realizar diversas combinaciones, atendiendo a los diferentes contextos (aplicaciones, usuarios, dispositivos); la composición de los contratos con las aplicaciones con aspectos además de garantizar una mejor modularización y evitar propiedades no deseables como el código invasivo, permite un más fácil desacople de los contratos.

Este trabajo exploratorio abre varias líneas de trabajo que continuaremos, identificar las limitaciones de los contratos, en cuanto a las propiedades y acciones que no puedan ser controladas y/o ejecutadas desde los dispositivos; analizar e incorporar las declaraciones de requerimientos de las aplicaciones, en la especificación de los contratos, resulta otra vía de trabajo complementaria; estimar el costo de la gestión de contratos, en términos de rendimiento, consumo de memoria y energía; y por último, analizar métodos de composición alternativos a AspectJ, para llevar la propuesta a otras plataformas móviles.

Referencias

1. Luo H. Shyu M.: Quality of service provision in mobile multimedia - a survey - Human-centric Computing and Information Sciences 1:5 Springer. (2011)
2. Devashish G.: Mobile Computing, International Journal of Advanced Research in Computer Science and Software Engineering, vol 3, num 9, pp 846-855 (2013)
3. Hefeeda M, Hsu C-H.: Mobile Video Streaming in Modern Wireless Networks. Proceedings of the ACM International Conference on Multimedia 1779–1780 (2013)
4. Kim H., Lee D., Lee J., Lee K., Lyu W., y Choi S: The QoE Evaluation Method through the QoS-QoE Correlation Model, en 4th International Conference on Networked Computing and Advanced Information Management, vol. 2, pp. 719-725. (2008)
5. Kosinski J., Nawrocki P., Radziszowski D., Zielinski K., Zielinski S., Przybylski G., Wnek P.: SLA Monitoring and Management Framework for Telecommunication Services, Networking and Services, LCNS 2008. 4th International Conference on, pp. 170–175, IEEE Computer Society (2008).

6. Moor K., Ketyko I., Joseph W., Verdejo J., Pessemier T., Marez L.:Evaluating Quality of Experience of mobile applications and services in a Living Lab setting. Mobile Networks and Applications, vol. 15 (2010).
7. Nawrocki P. y Śliwa A.:Quality of Experience in the context of mobile applications, Comput. Sci., vol. 17, n.o 3, p. 371 (2016).
8. Asensio J., Villagra V., Lopez de Vergara J., Berrocal J.: UML Profiles for the Specification and Instrumentation of QoS, Management Information in Distributed Object-Based Applications, Technical University of Madrid, Spain. (2001).
9. GMN Alliance: 5g white paper, Tech. Rep. (2015).
10. Viti R., Pecorellla T.,Brilli L., Fantacci R., Carlini C.: Enhance the QoE for Smartphone Applications: an API approach for QoS control in LTE-A, AEIT International Annual Conference. (2015)
11. Blair G., Stefani J.: Open Distributed Processing and Multimedia – Addison-Wesley, 1997.
12. Chalmers D., Sloman M.: A survey of quality of service in mobile computing environments, Communications Surveys & Tutoriales, IEEE, 2(2):2-10 (1999).
13. Hektner J., Schmidt J., y Csikszentmihalyi M.: Experience Sampling Method: Measuring the Quality of Everyday Life. SAGE (2007).
14. Kahneman D., Krueger A., Schkade D., Schwarz N. y Stone A.:A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method, Science, vol. 306, n.o 5702, pp. 1776-1780 (2004).
15. Ickin S.: Quality of Experience on Smartphones: Network, Application, and Energy Perspectives, Tesis Doctoral - Diva Portal (2015).
16. Streijl R., Winkler S, y Hands D.: Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives, Multimed. Syst., vol. 22, n.o 2, pp. 213-227(2016).
17. Haokun Luo Q. y Sanae Rosen Z.: QoE Doctor: Diagnosing Mobile App QoE with Automated UI Control and Cross-layer Analysis, IMC Proceedings de Conference on Internet Measurement, Canada. Pp 151-164 (2014)
18. ITU-T Recommendation X.641- Information Technology – Quality of Service: Framework - International Telecommunication Union (1997)
19. Kiczales G., Hilsdale E., Hugunin J., Kersten M., Palm J., Griswold W.: An Overview of AspectJ, Proceeding de 15th European Conference on Object-Oriented Programming pp 327-353 UK(2001)
20. Gonçalves J., Ferreira L.:A Framework for QoS-Aware Service based, Mobile Systems (2010).
21. Aggarwal V., Halepovic E., Pang J., Venkataraman S., y Yan H.: Prometheus: Toward Quality-of-Experience Estimation for Mobile Apps from Passive Network Measurements, HotMobile Proceedings de 15th Workshop on Mobile Computing Systems and Applications. Article No. 18 (2014).
22. Ravindranath L., PadhyeJ., Agarwal S., Mahajan R., Obermiller I. y Shayandeh S.: AppInsight: Mobile App Performance Monitoring in the Wild, en Proceedings of the 10th USENIX Conference on Operating Systems Design and Implementation, USA, pp. 107–120. (2012)
23. Ravindranath L., Padhye J., Mahajan R., y Balakrishnan H.: Timecard: Controlling User-perceived Delays in Server-based Mobile Applications, en Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles, USA, pp. 85–100.(2013)
24.Zhang L., Bild D., Dick R., Mao Z., Dinda P.: Panappticon: Event-based tracing to measure mobile application and platform performance, International Conference on Hardware/Software Codesign and System Synthesis (2013).
25. Xu Q., Mehrotra S., Mao Z., y Li J.: PROTEUS: Network Performance Forecast for Real-Time, Interactive Mobile Applications, Microsoft Res. (2013).
26. Gómez G., Hortigüela L., Pérez Q., Lorca J., García R., Aguayo-Torres M.: YouTube QoE evaluation tool for android wireless terminals. EURASIP J. Wireless Communications and Networking, vol. 164, pp. 1-14 (2014).

Sub-graph based Multicast Protection in WDM Networks A Multi/Many-Objective Evolutionary Algorithms approaches

Diego Pinto and Rodrigo Lugo

Universidad Nacional de Asunción. Facultad Politénica

Abstract. In this paper is addressed the multicast routing-and-protection, and wavelength assignment (MRPWA) problem which is critical for the success of applications point-multipoint in WDM networks. Basically, it is proposed the design of the primary and protection multicast routes, where the resources protection are based on sub-graph protection strategy subject to the quality requirements of the QoP protection: dedicated (1 + 1), shared (M: N) and better effort (without protection). In this way, NSGA-II and NSGA-III, evolutionary algorithms, are applied to MRPWA considering multi- and many-objectives optimization context, respectively. The evolutionary algorithms optimize simultaneously: (i) the total number of links used, (ii) the number of wavelength converters, (iii) the number of splitter nodes, and (iv) the number of destinations served-and-protected. Considering Hyper-volume measure, the experimental tests on a set of instances indicate that the protection approach based on sub-graph proves to be promising in comparison to the dual-tree protection strategy. On the other hand, the evolutionary technique oriented to many-objectives (NSGA-III) is more convenient than the oriented towards multi-objectives (NSGA-II) in the study problem.

1 Introduction

The growth of Internet traffic caused to the use of applications that need high bandwidth, such as multimedia applications among others [21, 20, 2, 18, 8]. Wavelength Division Multiplexing (WDM) technology is always the main candidates for transport these large traffic volumes. It is advantageous the enormous bandwidth of these networks, however, the failures of links and optical nodes are catastrophic for the significant losses caused even in small time lapses [18]. Therefore, the development of techniques based on the protection, recovering and restoration of traffic are critical issues. The protection problem can be divided into (a) the routing and wavelength allocation (RWA) problem for the primary path, and (b) protection problem for the secondary path. This problem becomes more complex when considering requests multicast and is called the Multicast Routing-and-Protection, and Wavelength Assignment (MRPWA) problem [8].

Since the multicast protection is very costly in terms of 'optical and computational resources', the contribution proposes in this work is as follows: (a) Sub graph-based protection scheme [24], (b) Protection level assigned to the quality of protection (QoP) requirements [17], (c) Multi-objective mathematical formulation [16], and (d) Solution based on multi- and many-objective evolutionary algorithms [5, 4]. The approach proposed in this work has not yet been presented in the literature in our best knowledge.

2 SURVIVABILITY IN OPTICAL NETWORKS

The Quality of Protection (QoP) is associated to the level that the resources (links with wavelengths) assigned to protection are shared. Zhong and Jaekel [24] proposed to apply the protection concept to achieve three levels of QoP where an alternative light-path (or path + wavelength) is assigned as a dedicated or shared backup according to the QoP requirements [24]. The previous work proposes the levels of protection to traffic unicast. These approaches were extended by Rodas and Pinto [17] considering several levels of protection quality. Similar to [24] we propose three QoP levels:

1. Level 1 dedicated protection (1: 1): protection resources they are exclusive of the request;
2. Level 2 shared protection (M: N): the resources of protect are assigned to several requests;
3. Level 3 best effort: does not assign any protection to application.

The multicast protection protection is more complex than those oriented to protect point-to-point type transmissions. To apply dedicated protection (1:1) to multicast tree is very expensive. One of the first approaches considers the concept of dedicated protection (1:1) where a 'dedicated secondary tree' protects a primary tree [8]. This is an approach based on dual-trees strategy which is very expensive because the protection needs, on average, at least the same amount of resources.

In dual-tree approach if a link of the tree fails, transmission to all destinations is interrupted and re-transmitted over the secondary tree. This solution offers 100% protection at the cost of a major interruption in the destination nodes which could be unacceptable for a good quality of service and depending on the characteristics structures of a topology. Further, there may be configurations of 'unprotected trees'. Equally, considering sharing resources depending on the level of protection, we would only decrease the use of these, but not the recovery time. An alternative approach is to consider a protection oriented to the destination nodes, that is, for each destination it is required a protection path. Under the previous concept, in [23, 10, 15] it has been proposed to expand the tree to a bi-connected sub-graph. This sub-graph is composed of a 'main tree' and of extra links that are activated at the time of a failure. An example of a sub-graph is provided in Figure 1. In the sub-graph approach, a failure to cause only a partial

interruption, that is, the failure only generates interruption to some destination nodes before a new tree is configured.

The literature reports several types of failures: link, node [22, 1] and shared risk groups [12]. In this paper we focus on link failures, considering only one failure at a time (simple failure).

3 MULTI AND MANY OBJECTIVE OPTIMIZATION

A Multi- or Many-objective Optimization Problem (MOP) usually consists of a set of n decision variables, a set of k functions objectives and a set of ω constraints [3]. The objective functions and restrictions are functions of the decision variables. Therefore, a MOP generally is optimized:

$$z = f(x) = (f_1(x), f_2(x), , f_k(x)) \tag{1}$$

sujeto a

$$g(x) = (g_1(x), g_2(x), , g_\omega(x)) \geq 0 \tag{2}$$

where $x = (x_1, x_2, ..., x_n) \in X$ is a decision vector, X denotes the decision space of $f(x)$, $z = (z_1, z_2, ..., z_k) \in Z$ is an objective vector, while Z denotes the objective space of $f(x)$.

For several years, various evolutionary algorithms have been proposed for multi-objective optimization, several of which have proven to be competitive when optimizing problems with 2 or 3 objectives. However, they have concluded that the performance of most of these algorithms is significantly degraded with the increase in the number of objectives to be optimized [14, 9, 6]. Because of this, it is proposed the term Many Objective to refer to the subset of problems where several objectives (more than three) require be optimized [7]. In the present work, for the search of solutions are used the evolutionary algorithms NSGA-II [5], which is a multi-objective algorithm and NSGA-III [4], which is a many objective algorithm, being the most referenced in the literature. Since the problem considers more than 3 objectives, we want to determine which of these algorithms it is the most appropriate for the problem in question.

4 MRPWA FORMULATION

For a better reading of this work, below the following nomenclature is indicated that will be used in formulating the MRPWA problem:

$\|.\|$	Indicates cardinality of a set;
G	Topology representing the optical network;
V	Node set of topology G;
E	Link set of topology G;
Λ	Set of wavelengths supported by the optical system, where $\Lambda = \{\lambda_1, \lambda_2, .., \lambda_{\|\Lambda\|}\}$;

(i,j) Optical link from node i to node j, where $i,j \in V$ and $(i,j) \in E$;
m Multicast request $m = \{s,(D),q\}$ with source nodes $s \in V$, a set of destination nodes $D = \{d_1,d_2,...,d_{|D|}\} \subset V$ and its quality requirement of protection $q \in \{1,2,3\}$; where 1 =*dedicated*, 2 =*shared* and 3 =*better effort, no protection*;
M set of multicast request, where $M = \{m_1,m_2,...,m_{|M|}\}$;
(i,j,λ) Light-link with start node i, destination node j and channel λ ;
$path_{sd}$ Unicast path with source node s and destination node d, where $path_{sd} = \{(i_1,j_1),(i_2,j_2),...,(i_p,j_p)\}$ with $i_1 = s$ and $j_p = d$; it is also understood that $j_1 = i_2, j_2 = i_2, ..., j_{p-1} = i_p$;
$l\text{-}path_{sd}$ Light-path with source node s and destination node d of a request m, where $l\text{-}path_{sd}= \{(i_1,j_1,\lambda_1),(i_2,j_2,\lambda_2),...,(i_p,j_p,\lambda_p)\}$ with $i_1 = s$ y $j_p = d$; it is also understood that $j_1 = i_2, j_2 = i_2, ..., j_{p-1} = i_p$;
t_m Light-tree primary for the multicast request m;
T_M Primary trees for the set M, where $T_M = \{t_{m_1},t_{m_2},...,t_{m_{|M|}}\}$;
p_m Protective links for the tree t_m, where $p_m = \{(i_1,j_1,\lambda_1),(i_2,j_2,\lambda_2),...,(i_{|p|},j_{|p|},\lambda_{|p|})\}$;
S_m Sub-graph or light-graph for the form m application by t_m and p_m, this is $S_m = t_m \cup p_m$;
S_M Multicast protection for the M; where $S_M = \{s_{m_1},s_{m_2},...,s_{m_{|M|}})\}$;
X_i^m Binary variable. If node i performs a conversion of wavelength for s_m then the variable is set to 1 otherwise 0. Here it is necessary equip node i with a length converter of wave for the request m;
Y_{ij}^m Binary variable. If the link (i,j) is used by S_m then the variable is set to 1 otherwise 0;
B_d^m Binary variable. If the destination node $d \in m$ is about the light-tree t_m then $B_d^m = 1$, in another case $B_d^m = 0$;
Z_i^m Binary variable. If node i bifurcates wavelength for s_m then the variable is set to 1 otherwise 0. This implies that in node i a splitter is needed to the request m;
H_d^m Binary variable. If the destination node $d \in t_m$ is found protected in the light-graph s_m, then $H_d^m = 1$, otherwise $H_d^m = 0$;

Given a network topology G and a set of requests multicast M, the problem consists to calculate a set of ligth-graph S_M such that they simultaneously optimize the following objective functions:

1. Minimize the total hop number: minimize the sum of the links used for each request m.
$$z_1 = \sum_{(i,j)\in E} \sum_{m\in M} Y_{ij}^m \tag{3}$$
2. Minimize the number of wavelength conversions: we seek that only some nodes have capacity of wavelength conversion (scarce conversion) since a converter will bring more costs to the components of the network.
$$z_2 = \sum_{i\in V} \sum_{m\in M} X_i^m \tag{4}$$

3. Minimize the number of wavelength bifurcation: minimize the number of splitter nodes. The number of amplifiers is minimized implicitly.

$$z_3 = \sum_{i \in V} \sum_{m \in M} Z_i^m \tag{5}$$

4. Minimize the number of blocked destinations: minimize the sum of the destinations that could not be achieved in each request. This calculation is done about the primary tree.

$$z_4 = \sum_{i \in V} \sum_{m \in M} B_i^m \tag{6}$$

5. Minimize the number of unprotected destinations: minimize the sum of the destinations that could not be protected. This calculation is done about the alternative paths.

$$z_5 = \sum_{i \in V} \sum_{m \in M} H_i^m \tag{7}$$

Subject to the following restrictions:

1. No overlapping: Given the light-graphs S_{m_1} and S_{m_2} , they cannot use the same wavelength λ on the same link (i, j):

$$\lambda_{ij}^{m_1} \neq \lambda_{ij}^{m_2} \tag{8}$$

2. Channel capacity: The total number of light-graphs that use the same link cannot be greater than the number of wavelengths supported by the system:

$$\sum_{(i,j) \in E} \left(\sum_{m \in M} Y_{ij}^m \leq |\Lambda| \right) \tag{9}$$

3. Bifurcation capacity: The number of bifurcations in which an input wavelength is divided cannot be greater than the specific capacity of the splitter C_S and the number of exit links θ of the node:

$$\sum_{i \in V} \sum_{m \in M} B_i^m \leq C_S, \theta_i \tag{10}$$

As we can see in the problem specification, we need find the routes for the traffic, protect these routes and assign wavelengths to each link that forms part of the light-graph, looking to optimally use the resources of a WDM network as in [13]. For address this is proposed to divide into two sub-problems the MRPWA problem: 1) Calculation of the primary ligth-tree t_m, that satisfies the multicast request m, and 2) Calculation of ligth-paths secondaries p_m, extending the light-tree t_m to a light-graph S_m. These two steps will be observed in the approach of the next section.

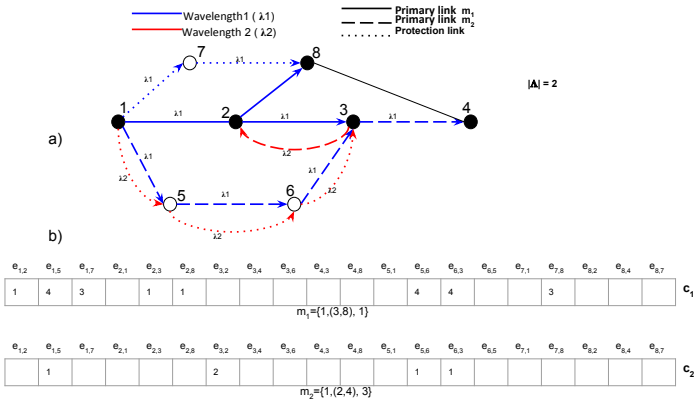


Fig. 1. Representation of a solution $C = \{c_1, c_2\}$. a) Solution on a ligth-graph, b) Chromosome. Note that nodes 2 and 3 are splitter nodes while node 3 is, in addition, a wavelength converter. The ligth-links used by protection ligth-paths have a number of $\lambda > |\Lambda|$ to differentiate them from the primary ligth-links.

5 PROPOSED APPROACHES

Chromosome representation. The representation of the individual for the MRPWA problem is formed by a vector of M elements $C = \{c_1, c_2, ..., c_M\}$ which represents a S_M solution.

Each element $c_m = \{c_{m1}, c_{m2}, ..., c_{m|2E|}\}$ is at same time a vector which represents a ligth-graph S_m; of length $|2E|$, due each link is direct. Each c_{m_e} is a whole number associated with the link of the network e and a S_m. A c_{m_e} element is a whole number indicating which wavelength in the link e is assigned to the sub-graph S_m. The states that can adopt me are: (a) $c_{m_e} = 0$ if the link is not used, (b) $c_{m_e} = [1, |\Lambda|]$ if the link is primary, and (c) $c_{m_e} = [|\Lambda|+1, 2*|\Lambda|]$ if it is secondary.

If $c_{me} \in [1, |\Lambda|]$ then the ligth-graph S_m is assigned the wavelength c_{me} on the link e and corresponds to the primary tree. If $c_{me} \in [|\Lambda|+1, 2*|\Lambda|]$ then the sub-graph S_m is assigned the wavelength $c_{me} - |\Lambda|$ at link e and corresponds to a secondary link. Figures 1a and 1b reflect the representation of a solution on a graph G and as a chromosome respectively.

Evolutionary Operators. The selection operator is the tournament binary [4]. The quality of an individual is determined according to the schemes proposed by NSGA-II [5] and NSGAIII [4]. The crossing operator was designed according to the structure of the chromosome representation while the mutation operator was not necessary since the crossing already introduces randomness in offspring.

We call TreeCrossover to the crossover operator designed. Given two individuals C and C', TreeCrossover works with each S_m and in particular with the genes of the main tree, that is, with $t_m \in S_m$. Basically, TreeCrossover generates a new individual corresponding to the primary ligth-tree t_m. The crossing is done

on sub-graphs corresponding to the same multicast request. If a link is used by the same primary trees this is automatically inherited in the new individual (the link and wavelength assigned to them if they are equal). It is clear that with the highest probability the new primary tree is incomplete, that is, it does not have a structure of tree. In this circumstance, new links are attached randomly to form a valid primary tree. Each link is added if it has at least one length of valid wave. Finally, an algorithm assigns wavelength to the tree, trying to assign the same wavelength to each *light-path*$_{sd}$ to minimize the use of converters.

6 EXPERIMENTAL TESTS

Experimental Environment. Characteristics of the environment in which they were made tests: The simulations were carried out in a Intel (R) Core (TM) i7-6700HQ CPU 2.60 GHz CPU, RAM8 GB, 64-bit operating system, WINDOWS 10. The JMetal java framework was modified [11] for the implementation of evolutionary algorithms. The version of the used jdk is 1.8.0 71. WDM technologies with 8 wavelengths have been considered per fiber optic and two fibers per link, one for each direction. For the use of the wavelength converters we opted for sparse conversion, i.e. some nodes are conversion capability. Of the same way we consider spare splitters capability. To keep operational the entire network we use as a method of survival the protection, introducing protection quality for each request. We consider simple link failures. Population size N: 100 individuals, and run time 15 minutes. The used topology is the NSF network [17].

6.1 Test Stages

Generation of Applications For the generation of requests we need to define:

- The origin node s for each request. It is set up request that starts from each node, i.e., each $v \in V$ can be a source node.
- Destinations D and the amount of destination nodes for each request. To set up the number of destinations $|D|$, it is defined the set $\partial = \{20, 40, 60, 80, 100\}$, where each ∂_α ($\alpha = 1, ..., |\partial|$) indicates the percentage of the total number of destination nodes. To calculate each number of destination nodes we use the number of nodes $|V|$ and each percentage ∂_α as it is observed in equation (11).

$$|D|_{\partial_\alpha} = \frac{(|V| - 1) \cdot \partial_\alpha}{100} \tag{11}$$

  The set of D nodes for a request m whose origin is v, is complete as follows: they are selected the $|D|_{\partial_\alpha}$ nodes more distant from v.
- The QoP levels are defined by the set $q = \{1, 2, 3\}$

Traffic Load We define a traffic load b by the following combination:

1. Number of destinations for each request ($|D|_{\partial_\alpha}$).
2. Quantity of requests (γ) that depart from each node $v \in V$, where $\gamma = \{4, 8, 12, 16, 20\}$.

This combination indicates that γ_φ ($\varphi = 1, ..., |\gamma|$) requests that have $|D|_{\partial\alpha}$ destinations each, corresponds to a traffic load. For a given traffic load $b = (|D|_{\partial_\alpha}, \gamma_\varphi)$, the requests defined by it and that start from the node v will be all the same, that is, the same application that the number of destinations $|D|_{\partial_\alpha}$ defined by the traffic load, but we make the assumption that they transmit information to different clients of the destination nodes.

6.2 Experimental Schemes

For the analysis of the algorithms, a series of simulations are carried out to obtain the Pareto fronts and calculate its quality. To measure the quality of evolutionary algorithms we consider two aspects: minimize the distance of the Pareto front obtained by the algorithm in front of Exact Pareto of the problem (convergence) and, maximize the extension of solutions on the front so that the distribution be as uniform as possible (diversity). For this end we consider the Hyper-volume measure [25].

To obtain the Pareto Fronts and their hyper-volumes, the following experiment is performed:

- Set of traffic loads B. Each traffic load $b = (|D|_{\partial_\alpha}, \gamma_\varphi)$ defines a set of M_k requests ($k = 1, ..., |V|$) that will depart from each node v and the union of the same $M = \bigcup_{M_k}^{M_{|V|}}$ they form an instance.
- Set of algorithms A. In view of what we have two types of protections, based on Dual-Tree (DT) and based on Sub-graph (SGM), we combine algorithms NSGA-II and NSGA-III with them to have 4 algorithms: A = {SGM-NSGA-II, DT-NSGA-II, SGM-NSGA-III, DT-NSGA-III}.
- Number of runs that an algorithm will be executed for a given instance is set up to 30.
- Amount of minutes that is will execute an algorithm for a given instance, which we establish to 15, since the solutions converge in that time according to a series of simulations made to de
ne this parameter.

7 DISCUSSION

In this experiment, we seek to answer the following questions:

1. It is really better a graph-based protection than a protection based on dual tree?; i.e. sub-graph vs dual-tree protection.
2. Which of the algorithms has a better performance for the problem raised?; i.e. NSGA-II vs NSGA-III algorithm

Table 2. Comparison types of protection

NSGAII-DT vs NSGAII-SGM					NSGAIII-DT vs NSGAIII-SGM				
b	$\overline{HV}_{DT}$	$\overline{HV}_{SGM}$	Dif. Average	p-value	b	$\overline{HV}_{DT}$	$\overline{HV}_{SGM}$	Dif. Average	p-value
20,4	0.2323603698	0.2573254477	0.02496507	6.79E-27	20,4	0.2361064607	0.2573744383	0.0212679776	0
40,4	0.0244993712	0.0586544169	0.03415504	0	40,4	0.0275746089	0.0581370340	0.0305624251	0
60,4	0.0122738375	0.0369586866	0.02468484	0	60,4	0.0136550414	0.0372864462	0.0236314048	0
80,4	0.0067977764	0.0180158842	0.01121810	0	80,4	0.0077292608	0.0179260846	0.0101968237	0
100,4	0.0033107388	0.0122170437	0.00890630	0	100,4	0.0035325561	0.0121891873	0.0086566312	0
20,8	0.0598880466	0.0285227265	−0.03136532	0	20,8	0.0306797659	0.0598904203	0.0292106544	0
40,8	0.0060680632	0.0142355676	0.00816750	0	40,8	0.0063766420	0.0140736478	0.0076970058	0
60,8	0.0029525178	0.0092203052	0.00626778	0	60,8	0.0033442638	0.0092508493	0.0059065855	0
80,8	0.0010823532	0.0039336016	0.00285124	0	80,8	0.0012435180	0.0039078280	0.0026643100	0
100,8	0.0001726355	0.0005029431	0.00033030	2.15E-25	100,8	0.0001812103	0.0005194902	0.0003382799	0
20,12	0.0110262050	0.0254325232	0.01440631	0	20,12	0.0119395854	0.0255742734	0.0136346879	0
40,12	0.0023972031	0.0062027011	0.00380549	0	40,12	0.0027619120	0.0062022490	0.0034403370	0
60,12	0.0013936591	0.0040742890	0.00268062	1.01E-50	60,12	0.0016096991	0.0041669509	0.0025572518	0
80,12	0.0004536063	0.0017335382	0.00127993	9.47E-43	80,12	0.0005303039	0.0017506655	0.0012203616	0
100,12	0.0000022927	0.0000022225	−0.00000007	0.159606	100,12	0.0000022598	0.0000022605	0.0000000008	0.988
20,16	0.0072013978	0.0143556035	0.00715420	2.36e-33	20,16	0.0079485627	0.0142777841	0.0063292214	0
40,16	0.0007585924	0.0022885193	0.00152992	1.42E-53	40,16	0.0016106009	0.0035083686	0.0018977677	0
60,16	0.0002574399	0.0009864438	0.00072900	2.88E-47	60,16	0.0008637138	0.0023070869	0.0014433731	0
80,16	0.0002574399	0.0009864438	0.00072900	8.52E-43	80,16	0.0002934669	0.0009859702	0.0006925033	0
100,16	0.0000012437	0.0000012638	0.00000002	0.211740	100,16	0.0000012446	0.0000012510	0.0000000064	0.730
20,20	0.0049414549	0.0090157573	0.00407430	9.55E-36	20,20	0.0056295233	0.0089880571	0.0033585339	0
40,20	0.0008892322	0.0022212129	0.00133198	1.30E-46	40,20	0.0009736031	0.0022119132	0.0012383100	0
60,20	0.0004932688	0.0014340629	0.00094079	3.91E-50	60,20	0.0005549143	0.0014227901	0.0008678758	0
80,20	0.0001703874	0.0006149684	0.00044458	1.38E-49	80,20	0.0001849825	0.0006146430	0.0004296604	0
100,20	0.0000008086	0.0000007790	−0.00000002	0.081407	100,20	0.0000008051	0.0000008050	0.0000000000	0.999

Sub-grafo vs Dual Tree Protection. To answer question 1, the averages of the hyper-volume values are compare considering (a) DT-NSGA-II and SGM-NSGA-II approaches (see Table 2 left part) and (b) DT-NSGA-III and SGM-NSGAI-II (see Table 2 right part).The idea is to compare the same evolutionary algorithm but using different types of protection, we see what kind of protection got better average of the hyper-volume for each instance.

According to the T-test [19], if there are significant differences ($p - value \leq 0.05$) between the averages, we have enough evidence to conclude that they are different.

In table 2 left part, the *Dif. Average* $= \overline{HV}_{SGM} - \overline{HV}_{DT}$ is greater than zero in 88% of the total instances existing (average sub-graph is greater). At 12% of instances DT-NSGAII is better, which is in cases where you have the maximum number of destinations for each request ($|D|_{100\%}$). Numericaly, the average of the hyper-volume was higher for SGM-NSGAII in the majority of the cases. In the same table, we see that the column *p-value* is always less than 0.05, so we have enough evidence that there are significant differences between the averages between the two protection approaches.

In table 2 right part, the *Dif. Average* column is greater or equal to zero always (average of sub-graph is greater or equal). For traffic loads where the maximum amount of destinations, it cannot be concluded which of the two pro-

Table 3. Evolutionary Algorithms comparison

NSGAIII-DT vs NSGAII-DT					NSGAII-SGM vs NSGAIII-SGM				
b	$\overline{HV}_{NSGAII}$	$\overline{HV}_{NSGAIII}$	Dif. Average	p-value	b	$\overline{HV}_{NSGAII}$	$\overline{HV}_{NSGAIII}$	Dif. Average	p-value
20,4	0.2323603698	0.2361064607	0.0037460909	0	20,4	0.2573254477	0.2573744383	0.0000489906	0.977
40,4	0.0244993712	0.0275746089	0.0030752377	0	40,4	0.0586544169	0.0581370340	−0.0005173829	0.4
60,4	0.0122738375	0.0136550414	0.0013812038	0	60,4	0.0369586866	0.0372864462	0.0003277596	0.512
80,4	0.0067977764	0.0077292608	0.0009314844	0	80,4	0.0180158842	0.0179260846	−0.0000897997	0.741
100,4	0.0033107388	0.0035325561	0.0002218172	0.115	100,4	0.0122170437	0.0121891873	−0.0000278564	0.893
20,8	0.0285227265	0.0306797659	0.0021570394	0	20,8	0.0598880466	0.0598904203	0.0000023736	0.995
40,8	0.0060680632	0.0063766420	0.0003085788	0.019	40,8	0.0142355676	0.0140736478	−0.0001619198	0.235
60,8	0.0029525178	0.0033442638	0.0003917460	0.002	60,8	0.0092203052	0.0092508493	0.0000305442	0.802
80,8	0.0010823532	0.0012435180	0.0001611648	0	80,8	0.0039336016	0.0039078280	−0.0000257736	0.617
100,8	0.0001726355	0.0001812103	0.0000085748	0.302	100,8	0.0005029431	0.0005194902	0.0000165471	0.373
20,12	0.0110262050	0.0119395854	0.0009133804	0.001	20,12	0.0254325232	0.0255742734	0.0001417502	0.299
40,12	0.0023972031	0.0027619120	0.0003647089	0	40,12	0.0062027011	0.0062022490	−0.0000004521	0.994
60,12	0.0013936591	0.0016096991	0.0002160400	0	60,12	0.0040742890	0.0041669509	0.0000926620	0.079
80,12	0.0004536063	0.0005303039	0.0000766977	0	80,12	0.0017335382	0.0017506655	0.0000171273	0.547
100,12	0.0000022927	0.0000022598	−0.0000000329	0.527	100,12	0.0000022225	0.0000022605	0.0000000381	0.429
20,16	0.0072013978	0.0079485627	0.0007471649	0	20,16	0.0143556035	0.0142777841	−0.0000778194	0.271
40,16	0.0014173188	0.0016106009	0.0001932822	0	40,16	0.0035381197	0.0035083686	−0.0000297511	0.452
60,16	0.0007585924	0.0008637138	0.0001051214	0	60,16	0.0022885193	0.0023070869	0.0000185676	0.535
80,16	0.0002574399	0.0002934669	0.0000360271	0	80,16	0.0009864438	0.0009859702	−0.0000004736	0.971
100,16	0.0000012437	0.0000012446	0.0000000009	0.960	100,16	0.0000012638	0.0000012510	−0.0000000129	0.574
20,20,	0.0049414549	0.0056295233	0.0006880683	0	20,20,	0.0090157573	0.0089880571	−0.0000277002	0.735
40,20	0.0008892322	0.0009736031	0.0000843709	0.001	40,20	0.0022212129	0.0022119132	−0.0000092997	0.695
60,20	0.0004932688	0.0005549143	0.0000616455	0	60,20	0.0014340629	0.0014227901	−0.0000112729	0.439
80,20	0.0001703874	0.0001849825	0.0000145952	0.012	80,20	0.0006149684	0.0006146430	−0.0000003255	0.969
100,20	0.0000008086	0.0000008051	−0.0000000036	0.837	100,20	0.0000007790	0.0000008050	0.0000000260	0.096

tections is better ($p - value > 0.005$), in contrast to the charges remaining $p - value \leq 0.005$.

Given that, in most cases there are significant differences between the averages in the two tables treated and the same is in favor of the protection based on sug-graph, we have enough evidence to conclude: Protection based on Sub-graph is better than Dual-Tree protection for requests that do not have the maximum number of destinations.

NSGAII vs NSGAIII. To answer question 2, the averages of hyper-volume obtained by NSGAII-DT and NSGAIII-DT are compared firstly (see Table 3 left part) and then NSGAII-SGM vs NSGAIII-SGM is performed (see 3 rigth part).

In table 3 left part, the column *Dif. Average* $= \overline{HV}_{NSGAIII} - \overline{HV}_{NSGAII}$ is greater than zero in most cases (NS-GAIII is better) and $p-value \leq 0.05$, except for high traffic load. Note that, for this last case the differences are not significant ($p - value > 0.05$). In table 3 rigth part, for subgraph-based protection, we see that the *p-value*> 0.05 for all cases, i.e. there is a weakly significant difference in terms of statistics between the averages of the hyper-volume.

We conclude that, by using Dual-Tree protection we have enough evidence that the NSGAIII algorithm will get a better performance. On the other hand, with sub-graph protection, NSGAIII is better than NSGAII; However, it is necessary to carry out more experimental tests to confirm this result.

8 CONCLUSION

In this paper the problem of MRPWA has been addressed considering an dual-tree and sub-graph protection as also multi- and many-objective optimization approaches. Experiments indicate that there is evidence that the NSGAIII performs better than the NSGAII and the sub-graph protection uses the resources better. As future work it is propose to extended the application on optical elastic networks and compare with other approaches of protection.

References

1. C. Bellino, A. Vazquez, and D. P. Pinto Roa. Wdm network design with node protection. In *XX Congreso Argentino de Ciencias de la Computación (Buenos Aires, 2014)*, 2014.
2. I. Chlamtac, A. Ganz, and G. Karmi. Lightpath communications: An approach to high bandwidth optical wan's. *IEEE transactions on communications*, 40(7):1171–1182, 1992.
3. C. A. C. Coello, G. B. Lamont, D. A. Van Veldhuizen, et al. *Evolutionary algorithms for solving multi-objective problems*, volume 5. Springer, 2007.
4. K. Deb and H. Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints. *IEEE Trans. Evolutionary Computation*, 18(4):577–601, 2014.
5. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.
6. M. G. Fabre. *Optimización de problemas con más de tres objetivos mediante algoritmos evolutivos*. PhD thesis, Doctoral dissertation, Masters thesis, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Ciudad Victoria, Tamaulipas, México), 2009.
7. M. Farina and P. Amato. On the optimal solution definition for many-criteria optimization problems. In *Fuzzy Information Processing Society, 2002. Proceedings. NAFIPS. 2002 Annual Meeting of the North American*, pages 233–238. IEEE, 2002.
8. A. Fei, J. Cui, M. Gerla, and D. Cavendish. A "dual-tree" scheme for fault-tolerant multicast. In *Communications, 2001. ICC 2001. IEEE International Conference on*, volume 3, pages 690–694. IEEE, 2001.
9. V. Khare, X. Yao, and K. Deb. Performance scaling of multi-objective evolutionary algorithms. In *EMO*, volume 2632, pages 376–390. Springer, 2003.
10. S. Khuller and R. Thurimella. Approximation algorithms for graph augmentation. *Journal of Algorithms*, 14(2):214–225, 1993.
11. A. J. Nebro and J. J. Durillo. jmetal 4.3 user manual. *available from Computer Science Department of the University of Malaga*, 2013.
12. J. V. Paez, A. Talia, E. D. Gimenez, and D. P. Roa. Optimal selection of p-cycles on wdm optical networks with shared risk link group independent restorability using genetic algorithm. *IEEE Latin America Transactions*, 10(1):1385–1390, 2012.
13. D. P. Pinto-Roa, C. A. Brizuela, and B. Barán. Multi-objective routing and wavelength converter allocation under uncertain traffic. *Optical switching and networking*, 16:1–20, 2015.

14. R. C. Purshouse and P. J. Fleming. Evolutionary many-objective optimisation: An exploratory analysis. In *Evolutionary Computation, 2003. CEC'03. The 2003 Congress on*, volume 3, pages 2066–2073. IEEE, 2003.
15. R. Ravi. *Approximation algorithms for Steiner augmentations for two-connectivity*. Department of Computer Science, Univ., 1992.
16. D. P. Roa, B. Barn, and C. A. Brizuela. Wavelength converter allocation in optical networks: An evolutionary multi-objective optimization approach. In *2009 Ninth International Conference on Intelligent Systems Design and Applications*, pages 414–419, Nov 2009.
17. M. D. Rodas-Brítez and D. P. Pinto-Roa. Quality of protection on WDM networks: A quantitative paradigm based on recovery probability. In *2014 XL Latin American Computing Conference (CLEI)*, pages 1–11, Sept 2014.
18. G. N. Rouskas. Optical layer multicast: rationale, building blocks, and challenges. *IEEE network*, 17(1):60–65, 2003.
19. D. Simon. Evolutionary optimization algorithms: biologically-inspired and population-based approaches to computer intelligence. hoboken, 2013.
20. J. H. Siregar, Y. Zhang, and H. Takagi. Optimal multicast routing using genetic algorithm for wdm optical networks. *IEICE transactions on communications*, 88(1):219–226, 2005.
21. A. Somani. *Survivability and traffic grooming in WDM optical networks*. Cambridge University Press, 2006.
22. A. Vazquez, D. P. Pinto-Roa, and E. Davalos. Optical multicast with protection against node failure an approach based on moaco. In *Computing Conference (CLEI), 2013 XXXIX Latin American*, pages 1–9. IEEE, 2013.
23. Y. Zhang and D. Sidhu. A multicast protection algorithm for optical wdm networks. In *Computer Communications and Networks, 2004. ICCCN 2004. Proceedings. 13th International Conference on*, pages 315–320. IEEE, 2004.
24. S. Zhong and A. Jaekel. Optimal priority-based lightpath allocation for survivable wdm networks. In *Computer Communications and Networks, 2004. ICCCN 2004. Proceedings. 13th International Conference on*, pages 17–22. IEEE, 2004.
25. E. Zitzler and L. Thiele. Multiobjective optimization using evolutionary algorithmsa comparative case study. In *international conference on parallel problem solving from nature*, pages 292–301. Springer, 1998.

OBD-II vehicle data capture and monitoring system prototype

Claudio Aciti[1,2], Mauricio Urraco[1], and Elías Todorovich[1,3]

[1] Universidad Nacional del Centro de la Pcia. de Buenos Aires, Tandil, Argentina, caciti@exa.unicen.edu.ar
[2] Universidad Nacional de Tres de Febrero, Departamento de Ciencia y Tecnología, Caseros, Argentina
[3] Universidad FASTA, Facultad de Ingeniería, Mar del Plata, Argentina

Abstract. In this work, a prototype of a system for capturing vehicle parameters according to OBD-II technology was designed and built. The prototype is able to read different magnitudes generated by a car through an ELM-327 USB interface that allows communication with the vehicle on-board computer, and the data obtained through a USB GPS. A 3G modem and a USB Wifi adapter were added to the developed prototype. The system was implemented on a Raspberry Pi B platform with the Raspbian operating system. The 3G modem was used to transmit information collected over the Internet to a remote device or exchange. The USB WiFi adapter allowed the Raspberry Pi to be configured as an Access Point. In this way, it could be connected to a mobile device that supports the Android operating system to visualize the data read. The tests were conducted on a Chevrolet Corsa Wagon Life Gls 1.4 4p model 2009.

Keywords: OBD-II, Monitor, Raspberry Pi

1 Introduction

The on-board diagnostic system (OBD) in vehicles is an integrated system that, by means of sensors, provides the driver and technician with information about the vehicle's engine and operation [11]. The advance in the electronic technologies embedded by the manufacturers in motorized vehicles has brought a series of benefits associated with motor performance. Numerous vehicles have used electronic control systems to increase the efficiency of both the fuel injection system and the ignition system. At the same time, different ways to diagnose the problems associated with these new electronic devices have been developed, and this is how an on-board computer currently controls sensors and actuators that keep the engine running under favorable conditions. The set of actuators, sensors and diagnostic software is called On-Board Diagnostic (OBD) System [8].

The US manufacturers, government, and entities concerned about the environment began the standardization process in the 1960s.In 1970, a series of

standards and requirements of gradual emissions was set for the maintenance of vehicles for prolonged periods of time. To comply with these standards, manufacturers turned to electronic controls of fuel supply and ignition systems. The sensors measured the performance of the motor and actuators adjusted the systems to provide optimum performance. In addition, these sensors could de accessed to obtain an early diagnosis of the vehicle [12]. In 1988, a standard connector and a set of diagnostic test signals were defined. This resulted in the first generation of on-board diagnostic system requirements, called OBD-I. The second version of the on-board diagnostic system, called OBD-II, is an improvement in both capacity and standardization with respect to the initial regulation of OBD. The OBD-II standard specifies the type of diagnostic connector and its pins, the available electrical signaling protocols, and the format of the messages [14,16].

In order to carry out this monitoring, a scanner is connected to the diagnostic interface and it allows access to vehicle information. There are basic and advanced scanners. The basic scanners show error codes and information stored by the vehicle, while the most advanced ones have an interface that displays the status of the sensors in real time, graphs and stores data, which greatly facilitates the diagnosis by the user. These advanced scanners have a high cost, which hinders access by mechanics and companies.

Nowadays, many companies have vehicles fleets for internal use or to provide transportation services. An information collection system allows them to control the use of vehicles to reduce investment and maintenance expenses, prevent failures and accidents, detect bad driving habits of their employees, reduce expenses associated with fuel, tires, etc. In South America, some companies provide services based on vehicle metrics. For example, Sitrack covers the so called bi-oceanic corridor and offers a fleet control system for monitoring vehicle use and drivers' actions. The system sends the information about what is happening in real time to a remote computer [5]. Another regional company based in Buenos Aires, Ful-Mar, provides a communication interface with the on-board computer that allows online reading and recording of the information captured from the vehicle in operation. [2]. Many companies develops fleet management software and hardware worldwide. For example, Fleetio [1], Plug-N-Track [3], and Titan GPS [6]. They offer devices and a cloud-based system for tracking services that provides satellite geo-localization, real-time alerts, driver behavior, vehicle health, maintenance, and reporting and analytics. Although it is a constantly growing market, there is still much to be done with this technology.

The main goal of this work was to develop a prototype of a data collection system for automobiles. This system registers what is happening in situ what happens and sends the captured information to a remote computer. In addition, an Android application was developed to view the captured data through a mobile device.

Thus, the objective was to develop a prototype that collects data from cars and is able to:

1. Monitor the on-board system on vehicles that comply with the OBD-II standard to capture engine parameters such as speed, RPM, throttle position, coolant temperature, torque, etc.
2. Obtain the geo-localization of the vehicle through satellite monitoring.
3. Connect to the Internet using telephone networks and send data remotely when the connection is available; otherwise, store the data in the internal memory until connection is available.
4. Design and implement an Android application that allows the user to connect to the data capture system to visualize the data that are being collected.

1.1 Limitations

The GPS has the limitation that in some areas it may not work, or it may work incorrectly, with considerable delays in obtaining the coordinates. A similar problem may occur with the 3G modem, because Argentina still does not have good coverage signal. Therefore, in certain areas the Internet connection may fail. In addition, the devices connected to the Raspberry Pi consume more current than the one the computer can deliver through its USB ports. For example, when the 3G modem is trying to establish the connection, it can consume peaks of up to 800mA, which can cause electrical problems that consequently end up causing a permanent failure in the Raspberry Pi. One solution to this problem is the use of a USB hub with independent power.

2 Related work

There are a number of references on OBD-II applications related to this work. For example, [7] presents a vehicle data acquisition system for the automatic management of automobile fleets using OBD, GPS, and WiFi technologies. The authors develop algorithms for decision making and suggest interesting applications for traffic control. In addition, the methods for remote diagnosis of the state of the vehicles were useful for the development of our work.

The positioning method presented in [13] corrects and replaces the inertial speed provided by a positioning system such as GPS by means of the speed value of the vehicle OBD-II system. This work makes positioning available in multiple conditions within a city. Although the work presented in our article also relates measurements from a GPS and an OBD-II system, the focus here is on monitoring.

The results of a system such as the one proposed here could serve as an input in works such as [9]. Such work proposes a method to calculate a route in order to save fuel and reduce the emissions of a vehicle. The method takes into account the driving style of the driver, the characteristics and components of the possible routes, and traffic information in real time. To evaluate the system, data sets were used, including OBD-II data of taxis with associated GPS positions. [17] also focuses on fuel savings by taking into account car driving styles. Consumption and environmental impact is obtained from OBD-II parameters. The authors

develop a system that finally offers suggestions to drivers on how to save fuel. To reduce the emission of gases, in [18] they had already presented a survey of the factors that influence the emission of pollutants through experimental studies using OBD.

In [15] the objective was to monitor and analyze land vehicle fleets. This system measures speed, distance, and fuel consumption by integrating OBD-II and GPS data. Then, it transmits the information via WiFi to a remote server.

3 System development

The objective of this paper was to develop an ad-hoc system capable of communicating with the vehicle on-board OBD-II computer in order to obtain different metrics related to its operation. These data collected from the vehicle's engine are complemented with the geographical position provided by a GPS-USB and pre-processed to be sent to a remote computer. Internet connectivity is provided through a 3G modem. The functional requirements for this vehicle OBD-II data capture and monitoring system prototype are as follows:

- Setting up the connection with the OBD-II system on board.
- Sending messages and processing the obtained OBD-II answers.
- Reading a series of PIDs and decoding the answers according to the formula corresponding to each magnitude.
- Obtaining the coordinates of the vehicle position from the GPS.
- Connecting to the Internet through mobile networks.
- Storing the collected information and sending it by mobile networks to a remote computer when the Internet is available. Otherwise, the information must be stored until it can be sent.
- Providing a graphical interface through a mobile device (tablet or smarthphone) to visualize the data that is being collected.

A Raspberry Pi was used since this computer has limited hardware capacities, which allows it to be produced and sold at low prices. Its size and energy consumption are also small. A version of Linux Debian runs On this platform, called Raspbian [4], but Raspberry Pi supports other OS. These characteristics favor projects like the one presented here. We used a Raspberry Pi B, which has two USB ports and 512 MB of RAM.

Fig. 1 shows an overview of the system and the way in which the Raspberry Pi interacts with the USB devices to capture the vehicle's engine magnitudes and the GPS data, sets up and maintains the Internet connection, and uses the WiFi interface in order to become an access point. In addition, it can be seen that the microcomputer invokes a web service to send the captured data to a remote computer. In addition, a user can use an Android mobile device to visualize the captured data in real-time.

Firstly, all devices are connected to the Raspberry Pi: the ELM-327 interface [10], the USB GPS, the WiFi adapter and the 3G modem. Once the vehicle is running or with its engine on, the power cable from the Raspberry Pi is connected

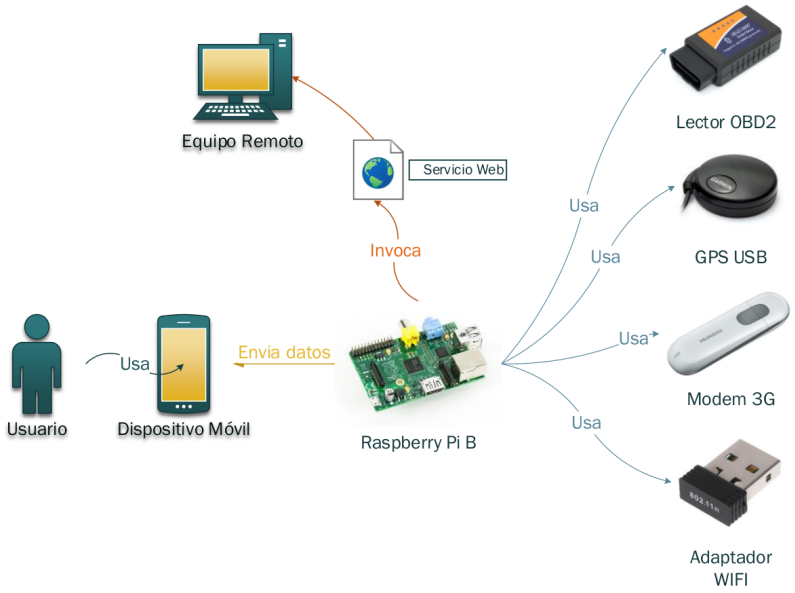


Fig. 1. System overview

to the vehicle power source. When the microcomputer has started the operating system, the data collector module, which is the main thread of the system, starts automatically. This module attempts to establish a communication with the previously mentioned ELM-327 interface to read the vehicle on-board computer. If an error occurs during this communication set-up or if the set-up fails, then the entire system will fail. On the other hand, if the connection is established successfully, then the data collector module initiates an infinite loop where the following tasks are performed:

1. The GPS reading module is implemented as a Python script. It is invoked and it obtains the geolocation and time information. If no data are received, then the system continues its execution normally but does not report any value for latitude, longitude and time.
2. The metrics of the OBD-II computer are read. If any of the readings fail, then values for that particular magnitude are not reported.
3. The data collected from the vehicle and the GPS are written in parallel to a data file and a specific TCP port.
4. Every 60 seconds the data sending module is invoked. This module is responsible for reading the data written in the file, sending it to a remote computer, and emptying the read file.

Along with this data collection process, the server module reads the data sent through the TCP port by the previous process and makes them available as a

web page through another TCP port. A person with a smartphone or tablet with an Android operating system and WiFi capability can connect to the Raspberry Pi as an access point, and open the mobile application developed in this work to visualize the data.

Then, the system consists of the following modules:

1. Vehicle data collector module
2. Data sending module
3. Data server for the mobile application
4. GPS reader module
5. Mobile application

These modules are implemented in such a way that they are independent from each other to facilitate the testing and integration stages.

3.1 Data collector module

This is the central module of the system. On the one hand, it establishes communication with the on-board computer of the vehicle (OBD-II) to request the different parameters that are to be monitored. To do this, it initializes the ELM-327 embedded system using AT commands and then requests and interprets the different responses through Mode 1. In addition, it invokes the GPS reader modules and the client that sends data to a remote computer.

The ELM-327 embedded system detects and interprets nine different low-level OBD protocols. It also translates and returns the information in a unique format for any of them, allowing the independence with the internal protocol used by the vehicle. To configure the ELM-327 system, AT commands are used. Among other options, it is possible to deactivate the echo of the messages, deactivate the header in the response, or force the communication protocol to be interpreted. Below are some examples of AT commands:

- AT Z → Resets the device and shows its identification string.
  Type: at z
  Answer: ELM327 v1.5
- AT DP → Shows the current OBD vehicle protocol.
  Type: at dp
  Answer: AUTO, ISO 14230-4 (KWP FAST)

Although the OBD-II system has several modes of operation, this work focuses on mode 1 or data flow, which allows access through a PID (Parameter Identification Data) to the Engine Control Unit (ECU) analog and digital inputs and outputs. There are two sets of PIDs: the standard ones and those provided by the manufacturers. Most of the OBD-II PIDs in use are not standard, but for the most modern vehicles, have many more functions compatible with the OBD-II interface that are covered by the standard PIDs, and there is relatively less overlap between vehicle manufacturers for private PIDs. Although there are

more than thirty standard PIDs, through which it is possible to read different metrics of the vehicle, in this work we consider only four to show the functioning of the system and the parameter reading method.

Speed: The PID corresponding to the speed request is 0x"0d" and the answer is one byte wide. The valid range is from 0 to 255 km/h and the formula to decode the speed is A, with A being the response value.

RPM (Revolutions per minute): The PID corresponding to the motor RPM request is 0x"0c" and the response is two bytes wide. The valid range is from 0 to 0 16,383.75 rpm and the formula for decoding the RPM is $((A*256)+B)/4$, with A and B being the response's first and second byte respectively.

Throttle position: The PID corresponding to the throttle position request is 0x"11" and the response is one byte wide. The valid range is from 0 to 100% and the formula for decoding it is $A*100/255$, with A being the response value.

Mass air flow (MAF) rate: The PID corresponding to the MAF request is 0x"10" and the response is two bytes wide. The valid range is from 0 to 655.35 grams per second (gr/s) and the formula for decoding it is $((A*256)+B)/100$, with A and B being the response's first and second byte respectively.

3.2 Mobile Application

The server module ensures that the data read from the vehicle OBD-II computer and the GPS are available to any user in the vehicle with an Android mobile device with WiFi capability. The idea is that the user connects via Wi-Fi to the Raspberry Pi as an AP (Access Point). The user sets an IP address and port from the phone's browser or from an application with an embedded browser.

The Android application is basic. It has an embedded browser that is automatically routed to a TCP port assigned with the IP address of the connected Raspberry Pi. For the development of this application, a framework called Ionic-Framework is used which allows applications to be implemented for Android and iOS from HTML5 code and JavaScript code. It also supports the use of AngularJS, which facilitates the development. The application has only one screen which shows the values of the captured data: speed, rpm, throttle position, air mass flow, latitude, longitude and time, in this order.

4 Case study

The developed system was tested on a Chevrolet Corsa Wagon Life Gls 1.4 4p 2009. Initially, the modules were tested independently, and then, the system test was performed. Several problems of an electrical nature were faced. The Raspberry Pi B can deliver up to 500mA of current through the USB ports and the 3G modem alone consumes peaks of up to 800mA. Thus, it is impossible to power the modem via the Raspberry Pi. Therefore, two solutions are possible: one of them is to use an independently powered USB hub, and the other one is to use the new computer, such as Raspberry Pi 2B, which can be configured to deliver up to 1.2A among all its USB ports. Therefore, with a source of 2A

or more, it is possible for the 3G modem and the other connected devices to work properly without the need for an independent USB hub. In this work, we decided to use an additional power source.

The screenshot in Fig. 2 shows the values of the different parameters while the vehicle is stopped and when it is driven at low speed in the city. The engine is running at almost 2200 RPM, while the speed is around 30 km/h. The throttle position is a value greater than zero, but rather low, because during the driving the throttle is slightly pressed.

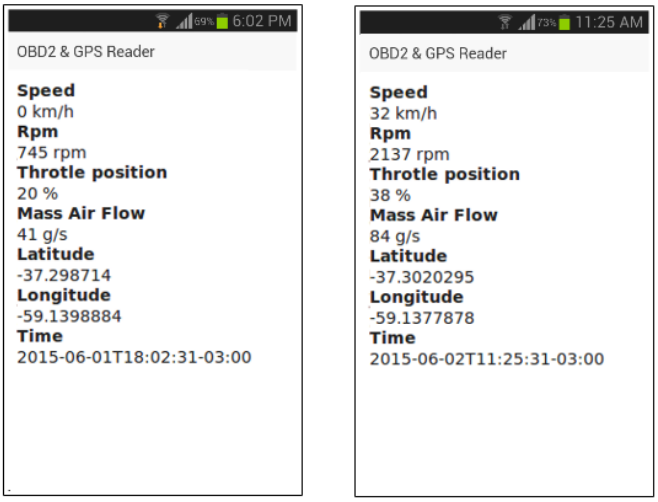


Fig. 2. Vehicle information when the vehicle is stopped and when it is driven at low speed

5 Conclusions and Future work

It was possible to develop and implement a prototype of a system that collects data and sends it to a remote computer through the Internet. This prototype can be extended both in hardware and software .

The prototype allows the user to monitor the on-board system in vehicles that comply with the OBD-II standard to capture different metrics. It also allows the user to obtain the geo-localization of the vehicle through satellite monitoring using a GPS.

To send the data, the system connects to the Internet using telephone networks when the connection is available; otherwise, the system stores the data in the internal memory until it is reconnected.

A mobile application was developed to allow the user to connect to the data capture system and to visualize the data that are being collected.

Particularly, in this work, the prototype must be electrically connected inside a vehicle with its engine on. Finding and adapting a reliable and sufficient power source can be a problem in the context of a vehicle. As explained above, in this work an additional independent power source was used.

While this work focuses on developing a prototype, from this basic platform, it is possible to implement extensions such as:

- improving the data visualization in the mobile application;
- capturing a greater number of metrics of the vehicle engine, at best, all standardized metrics for any vehicle that supports the OBD-II protocol;
- an independent battery that could also be used to power the prototype and thus replace the power coming from the car- this would prevent the prototype from shutting down if the vehicle is stopped or the power supply is cut off- and;
- developing a software that based on different algorithms, which analyze the captured information and infer more complex data such as car wear, average consumption, reduction of gas emissions, among others.

References

1. Fleetio. https://www.fleetio.com, accessed: 2018-06-22
2. Ful-mar sa. http://www.ful-mar.com, accessed: 2018-06-21
3. Plugntrackgps. https://www.plugntrackgps.com, accessed: 2018-06-22
4. Raspbian. https://www.raspbian.org/, accessed: 2018-06-27
5. Sitrack. https://www.sitrack.com, accessed: 2018-06-21
6. Titan gps. https://titangps.ca, accessed: 2018-06-22
7. Aljaafreh, A., Khalel, M., Al-Fraheed, I., Almarahleh, K., Al-Shwaabkeh, R., Al-Etawi, S., Shaqareen, W.: Vehicular data acquisition system for fleet management automation. In: Vehicular Electronics and Safety (IC-VES), IEEE International Conference. pp. 130–133 (20011)
8. Corvalán, R., Osses, M., Urrutia, C., Barrientos, V.: Control de emisiones de fuentes móviles, informe final para la comisión nacional del medio ambiente. Tech. rep., Gobierno de Chile, Santiago de Chile (diciembre 2000)
9. Ding, Y., Chen, C., Zhang, S., Guo, B., Yu, Z., Wang, Y.: Greenplanner: Planning personalized fuel-efficient driving routes using multi-sourced urban data. In: 2017 IEEE International Conference on Pervasive Computing and Communications (PerCom). pp. 207–216 (March 2017)
10. Elm Electronics Inc.: Elm327 datasheet. Tech. rep., Elm Electronics Inc. (2008)
11. Henderson, B., Haynes, J.: OBD-II & Electronic Engine Management Systems. Haynes Repair Manuals, Haynes Manuals N. America, Inc. (2006)
12. Landin, C.A.M., Jimenez, U.Y.F.V.: Scanner Automotriz Interfaz PC. Master's thesis, Instituto Politécnico Nacional, Mexico D.F. (2010)
13. Lim, J., Choi, K.H., Kim, L., Lee, H.K.: Land vehicle positioning in urban area by integrated gps/beidou/obd-ii/mems imu. In: 2016 IEEE International Conference on Intelligent Transportation Engineering (ICITE). pp. 176–180 (Aug 2016)

14. Lopes, J.C.O.: Diseño de un escáner automotriz OBDII Multi-protocolo. Master's thesis, Universidad de San Carlos de Guatemala (2014)
15. Malekian, R., Moloisane, N.R., Nair, L., Maharaj, B.T., Chude-Okonkwo, U.A.K.: Design and implementation of a wireless obd ii fleet management system. IEEE Sensors Journal 17(4), 1154–1164 (Feb 2017)
16. McCord, K.: Automotive Diagnostic Systems: Understanding OBD-I & OBD-II. S-A Design Workbench Series, CarTech (2011)
17. Meseguer, J.E., Toh, C.K., Calafate, C.T., Cano, J.C., Manzoni, P.: Drivingstyles: a mobile platform for driving styles and fuel consumption characterization. Journal of Communications and Networks 19(2), 162–168 (April 2017)
18. Ortenzi, F., Campbell, F., Zuccari, F., Ragona, R.: Experimental measurement of the environmental impact of a euro iv vehicle in its urban use. In: SAE Technical Paper 2007-01-0966 (2007)

Análisis para Despliegue de una Red de Sensores Heterogénea

Medina Santiago[1]Romero Fernando[1], De Giusti Armando[1,2], Tinetti Fernando G.[1,3]

[1]Instituto de Investigación en Informática III-LIDI, CEA-CIC, Facultad de Informática
Universidad Nacional de La Plata, 50 y 120 2do piso, La Plata, Argentina
[2]CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas
[3]CIC – Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires
{smedina, fromero, degiusti, fernando}@lidi.info.unlp.edu.ar

Abstract. Una red inalámbrica de sensores o Wireless Sensor Network (WSN) en inglés, es un conjunto de nodos compuestos por un microcontrolador, diversos sensores, dispositivos de comunicación y en algunos casos actuadores, que permiten recoger y trasmitir datos del entorno físico en cual se aplican e influir sobre él. La comunicación inalámbrica es el medio por el cual los nodos se conectan, utilizando diferentes estándares y protocolos, por ejemplo WiFi, Bluetooth, ZigBee, LoRa, etc. Dentro de una misma Red podemos utilizar una combinación de tecnologías inalámbricas y distintas topologías para la distribución de nodos. En este trabajo se hará un análisis de hardware específico en el que se combinan dos tecnologías diferentes de comunicaciones inalámbricas: WiFi y LoRa. Cada una de esas tecnologías se aprovechará por sus características de alcance y ancho de banda. Los puntos de análisis serán la distancia y confiabilidad en el envío de paquetes.

Keywords: Red de Sensores, LoRa, WiFi, WSN.

1 Introducción

El crecimiento exponencial en el desarrollo de hardware del último tiempo ha generado nuevos microcontroladores de pequeño tamaño, con mayor capacidad de procesamiento y menor consumo energético [3]. Estos dispositivos, combinados con el avance sobre nuevas tecnologías de comunicación inalámbrica nos dan como resultado la capacidad de desplegar diferentes tipos de nodos que conforman Redes de Sensores aplicadas en diferentes ámbitos, para obtener datos y poder realizar operaciones. Entre las características de las redes de sensores podemos nombrar:

- Baja tasa de transferencia
- Limitación en la alimentación o bajo consumo de los nodos sensores
- Alimentación por baterías
- Escalabilidad y/o gran cantidad de nodos

En la estructura básica de una Red de Sensores se pueden definir 3 partes: los nodos

sensores, la red inalámbrica de comunicación y nodos centralizadores o gateways [1].

Los nodos sensores son pequeñas unidades conformadas por un microcontrolador, uno o varios sensores, un módulo de energía, un módulo de comunicación y en algunos casos actuadores. Estos nodos tienen como función primaria la recolección y transmisión de datos para un control centralizado. Entre los sensores que conforman este tipo de nodos podemos nombrar sensores térmicos, de humedad, de movimiento, sísmicos, de diferentes gases, acústicos, de distancia, de luz, de flujo de líquido, etc. También pueden incluir actuadores que influyen directamente en el ambiente, como ser, algún tipo de alarma sonora, una luz, un motor, o simplemente un pulso para activar un mecanismo, como por ejemplo abrir una puerta o cerrar una válvula [1].

Las redes de comunicación que se utilizan están dentro de las denominadas WPAN (Wireless Personal Area network) o las LPWAN (Low Power Wide Area Network), redes inalámbricas basadas en estándares como 802.15.4 y 802.11 [13] [14], que definen los dos primeros niveles del modelo OSI, Capa Fisica y Capa de Enlace de datos. Las WPAN, son redes que comúnmente no cubren largas distancias y conectan diferentes dispositivos inalámbricos que no requieren de una alta tasa de transferencia, por debajo de 1Mbps, como ejemplo se puede nombrar Bluetooth. Las LPWANs, son redes orientadas a ocupar largas distancias con un bajo consumo, operan en la banda libre ISM (Industrial, Scientific and Medical) por debajo de 1GHz y poseen esquemas de modulación, como modulación en banda estrecha o modulación *Spread Spectrum*, que posibilitan enlaces de varios kilómetros en ciudad, e incluso decenas de kilómetros en ambientes rurales, con un bajo costo energético.

Los nodos centralizadores son los encargados de recibir los datos que generan los nodos sensores, y servir como puntos de acceso a otra red o como conector a un servidor en donde se procesarán los datos obtenidos para tomar las decisiones correspondientes. Generalmente estos dispositivos tienen una unidad de procesamiento de mayor capacidad, y en la mayoría de los casos con dos módulos de comunicación, uno para la interacción con los nodos sensores y otro para transmitir los datos recolectados al dispositivo de recolección central.

Estas Redes tienen diferentes ámbitos de aplicación debido a que dan la posibilidad de desarrollar los nodos de manera personalizada para su utilización en cada área específica, y adaptar la topología de red a las necesidades del problema. Entre los ejemplos más comunes de aplicación de las Redes de Sensores Inalámbricos están, desarrollos militares, monitoreo ambiental, agricultura, seguridad, domótica, ciudades inteligentes [1] [2]. Quizás justamente por esta razón, no existe una topología determinada común a todas las WSN, pueden no tener un patrón específico para la distribución de nodos o ser combinaciones de topologías conocidas, como estrella, malla y árbol. La estructura de la topología va a ser ad hoc cumpliendo con los requerimientos del desarrollo y las limitaciones que puede tener el ambiente físico en donde se colocan los nodos.

Por ultimo uno de los conceptos nuevos que está fuertemente emparentado con las Redes de Sensores, es el Internet de las Cosas (IoT), ya que comparten características

comunes, como la comunicación inalámbrica, el bajo consumo energético y la recolección de datos en ambientes físicos, pero hay que aclarar que no son lo mismo, a pesar de sus similitudes. Entre las diferencias de estos dos conceptos podemos nombrar, que los componentes de una red de sensores no tienen la necesidad de estar conectados a internet o que el IoT está orientado a la conexión de miles de dispositivos a lo largo del mundo, en cambio los nodos de una WSN generalmente están confinados a un espacio/aplicación específica y limitada.

2 Protocolos

Si bien como se aclaró antes existen diversos protocolos involucrados o que se pueden utilizar en redes de sensores, se describen a continuación los que hemos utilizado en nuestros experimentos.

WiFi: Es una tecnología de interconexión inalámbrica, que hace referencia a "Wireless Fidelity" (fidelidad inalámbrica). Wi-Fi Alliance es una asociación que se encarga de certificar que los dispositivos de WiFi cumplen con el conjunto de IEEE de estándares inalámbricos 802.11 [13]. Este estándar define los niveles inferiores de la capa del Modelo OSI, la capa física y la capa de enlace de datos, para conexiones inalámbricas que utilizan ondas electromagnéticas. Dentro de e ste estándar existen distintas variaciones en las cuales se definen características específicas para cada caso, modificando principalmente la frecuencia en la cual se trabaja y la velocidad de transferencia de datos. 802.11 b/g/n es el estándar de mayor utilización en los dispositivos, trabaja en la frecuencia de 2.4GHz y alcanza velocidades de hasta 300Mbps. Su topología de red consiste en la creación de puntos de acceso en donde se conectan los diferentes dispositivos que quieren pertenecer a una red. Respecto a la seguridad, la conexión a los puntos de acceso puede ser o no protegida con algún tipo de contraseña.

LoRa (acrónico de Long Range): es una tecnología de modulación de radio desarrollada por la empresa Semtech, que define y es propietaria de la capa física. Las demas capas (denominadas LoRaWAN) son desarrolladas por una organización sin fines de lucro denominada Lora Alliance de forma abierta. Semtech se encarga de comercializar los dispositivos [10] [11].

La estructura de la red LoRa tiene 3 partes: Servidores de Red que centralizan y procesan la información, estaciones base o concentradores que hacen de punto de acceso entre el servidor, y la tercera parte de la infraestructura, que son los nodos finales. Estos nodos finales se encargan de recolectar la información del ambiente y transmitirla. Los nodos finales y el concentrador forman una topología estrella de un solo salto. La Fig. 1 muestra esquemáticamente una red LoRa [15]. Esta red permite seleccionar 6 factores ortogonales de ensanche de espectro (Spreading Factor- SF) numerados del 7 al 12, para los cuales define una relación de potencia y tasa de transferencia. Como resultado, a mayor SF mayor alcance y menor velocidad en la comunicación. Las velocidades varían entre 0.3kbps y 50kbps. La banda de frecuencia en donde trabaja se encuentra dentro de la banda libre ISM, por debajo de 1GHz, específicamente funciona en, 868MHz, 433MHz y 915MHz, las dos primeras en Europa y la tercera en América. LoRaWAN, es el

protocolo de red que utiliza LoRa para las capas inferiores. Agrega características como: Encriptación AES 128, 3 clases de nodos, administración de dispositivos, redes públicas y privadas.

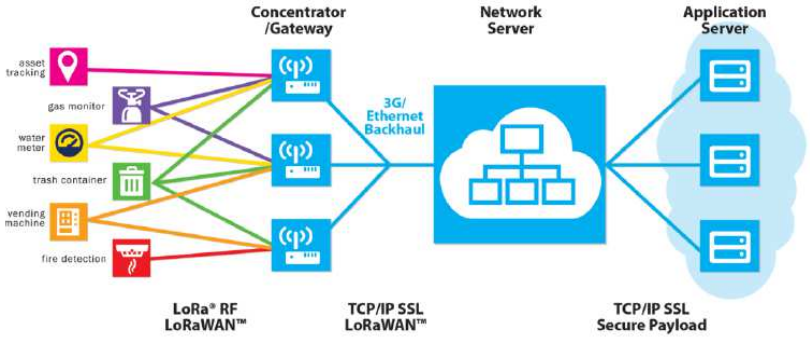


Fig. 1 – Topología de Red LoRaWAN

3 Hardware Utilizado

De manera similar a lo que sucede con las tecnologías de comunicaciones inalámbricas, los microcontroladores y sensores utilizados en las redes de sensores son muy heterogéneos. Describiremos a continuación los utilizados en la experimentación para este trabajo.

NodeMCU: es un firmware y una placa de desarrollo open source, basada en el chip de la empresa Espressif ESP8266-12E. La placa agrega al microcontrolador el chip para la conexión UART a través de un puerto micro USB, que sirve como alimentación y punto de comunicación [9]. ESP8266 es un microchip WiFi de bajo consumo que contiene como CPU al procesador Tensilica LX106 de 32-bit RISC, a 80MHz; con 64 KB de memoria de instrucciones y 96KB de datos [4] [5]. Este integrado varias características de hardware que lo hacen apropiado para una variedad muy amplia de aplicaciones:

- 1 ADC de 10-bit
- 17 GPIO pins
- Interfaces: 3 SPI, 1 I2C, 1 I2S y 2 interfaces UART
- PWN
- Protocolo WiFi 802.11 b/g/n

Las dos características más importantes, sin embargo, no son necesariamente de hardware: bajo costo y amplio soporte de programación. En particular, la programación utilizando el IDE y las bibliotecas de Arduino han generado una gran cantidad de ejemplos disponibles gracias al uso masivo. En la Fig. 2 se puede apreciar, por ejemplo, que los pines GPIO son identificados como los del Arduino UNO (D0, D1, etc.).

Heltec WiFi LoRa 32: La Fig. 3 muestra la placa de desarrollo construida por la empresa Heltec. Esta placa incluye dos chips principales, el ESP32 y el transceptor LoRa

SX1278 (con conector I-PEX, para antena externa), sumados a un convertidor CP210x MicroUSB a UART, una pantalla 0,96" OLED y un conector Molex PicoBlade de 2 pines para batería LiPo/Li-Ion [12]. El módulo WiFi puede transmitir a una tasa de hasta 150Mbps. El módulo Bluetooth, tiene funcionamiento normal y también BLE (Bluetooth Loe Energy).



Fig. 2 – Placa de desarrollo NodeMCU



Fig. 3 – Placa de Desarrollo de Heltec

EL chip ESP32 es una solución integrada de WiFi-Bluetooth con radiofrecuencia de 2.4GHz. Basado en el microprocesador Tensilica Xtensa LX6 de 32 bit, a 160MHz, con 448 KB de memoria ROM y 520 SRAM [6] [7]. Algunas de sus características más importantes de hardware son:

- 34 GPIOs programables
- 1 SAR ADC de 12bit, 2 DAC de 8 bit
- Interfaces: 4 SPI, 2 I2S, 2 I2C
- PWM
- Protocolo 802.11 b/g/n

EL SX1278 es un chip transceptor de radio frecuencia LoRa desarrollado por Semtech, dentro de la placa de desarrollo de Heltec, tiene dos versiones de frecuencia: 433MHz y otro que soporta las frecuencias 868MHz y 915MHz [8]. Este módulo se conecta vía SPI al ESP32 (conexión que integrada en la propia placa).

4 Arquitectura de la Red Heterogénea

La Fig. 4 muestra esquemáticamente la red inalámbrica heterogénea construida para los experimentos. Los nodos finales se conectan al nodo master utilizando WiFi y en distancias relativamente cortas. El nodo master se conecta a distancias mayores con el nodo gateway utilizando Lora. El nodo gateway puede estar conectado al exterior del

sistema utilizando una variedad muy amplia de tecnologías. El rango de estas tecnologías varía desde la conexión directa a un servidor donde se procesan los datos y se controla el sistema hasta enviar vía internet los datos a una nube (*cloud*) donde se concentren datos a procesar de diferentes sistemas.

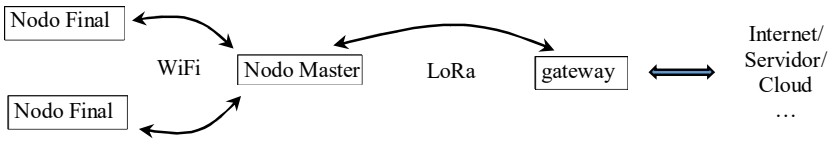


Fig. 4 – Red Inalámbrica Heterogénea

WiFi es la tecnología de comunicación que utilizan los nodos finales, basados en la placa de desarrollo NodeMCU. Este tipo de nodo se conecta a la red wifi generada por el master a través de una dirección ip fija preconfigurada. Iterativamente genera un paquete (de manera análoga a lo que sería el muestreo de un sensor), lo envía y espera la respuesta del master para verificar la integridad de la comunicación. Específicamente en la implementación actual, se utiliza el protocolo de transporte UDP que tiene muy poca sobrecarga y actividad de administración para el nodo master (no representa un problema, sin embargo, en los nodos finales). Además, el uso de UDP tiene la ventaja de ahorro de consumo de energía respecto a TCP, dado que con UDP no se tienen los mensajes propios de establecimiento de sesión/conexión necesarios para usar TCP.

El nodo master se basa en la placa de Heltec presentada anteriormente y tiene tres funcionalidades básicas:

- Generar un Punto de Acceso WIFI para los nodos finales recolectores
- Reconocer a cada nodo final que se conecta a su red
- Retransmitir la información a través de la comunicación LoRa

En ejecución, el nodo master iterativamente recibe en un puerto específico los mensajes de los nodos finales, los responde con un ACK y retransmite lo recibido hacia el nodo gateway. Básicamente un nodo master concentra los datos de los nodos finales y los retransmite al nodo gateway.

El nodo gateway también utiliza una placa Heltec, que con el SX1272 se maneja los mensajes LoRa y con el integrado ESP32 se conecta vía WiFi a la red en donde se ejecuta un servidor para el manejo/procesamiento de los datos. La funcionalidad del gateway es similar a la de los nodos master, aunque utilizando LoRa: concentra los paquetes LoRa de cada uno de los nodos master de la red, y los reenvía hacia el servidor para su procesamiento.

5 Experimentación

Todos los experimentos se llevaron a cabo en el mismo contexto, que se muestra en la Fig. 5: el nodo gateway se instaló en la terraza de la Facultad de Informática de la UNLP, a unos 10 metros de altura. La posición del nodo master se fue variando para hacer las distintas mediciones, siempre dentro de este ambiente “semi urbano”, intentado lograr

una línea de vista clara hacia el gateway para la conexión LoRa. Cabe aclarar que en este punto hay varias antenas trabajando a diferentes frecuencias: de telefonía móvil, de conexiones punto a punto WiFi y hasta un gateway LoRaWAN. Las distancias del nodo master al nodo gateway se muestran también en la Fig. 5, donde a partir de las cuales determinamos de manera específica las características de conectividad de LoRA a 100m, 200m, 350m y 500m. A 350m y a 500m ya la línea de visión no es tan despejada como en los casos de 100m y 200m.



Fig. 5 – Ambiente de Experimentación

Inicialmente se realizaron experimentos para analizar el funcionamiento de WiFi de manera tal que luego sea posible enfocarse en los experimentos específicos de la red LoRa a diferentes distancias. De estos experimentos iniciales se obtuvo como resultado que la comunicación WiFi entre el nodo final y el nodo master debe mantenerse por debajo de los 50 metros (siempre utilizando el protocolo de transporte UDP). Superando los 50m no se puede garantizar una reconexión a la red en caso de reinicio del módulo (escenario que se puede dar por ejemplo al utilizar los modos sleep para ahorro de energía de los módulos). La distancia máxima de comunicación lograda fue de 100m, pero solo si se realizó la conexión a la red en una distancia menor y luego se mantiene.

Se plantearon dos clases de experimentos en función de la cantidad de nodos finales que se utilizan, para verificar algunos aspectos importantes de análisis de redes de sensores heterogéneas además de la propia distancia detallada anteriormente:

- Un único nodo final, generando un paquete de 10 bytes cada 1 segundo, simulando el censado de un dato del ambiente. Este paquete es enviado por un socket UDP al nodo master quien lo encapsula en LoRa y lo reenvía al gateway de manera inmediata. Como se detalló antes, la distancia entre el nodo master y el nodo final se mantuvo constante en 50 metros y siempre con una línea de vista directa entre los dos puntos de la conexión.

- Varios nodos finales con un máximo de 4, generando un paquete de 10 bytes cada 1 segundo cada uno. El nodo master en este caso se encarga de recibir cada mensaje y armar un paquete LoRa compuesto a partir de los (hasta 4) mensajes UDP para luego enviarlo al gateway. En este caso las distancias entre los nodos finales y el nodo master no fueron constantes, pero siempre se respetó no sobrepasar el límite de 50 metros de radio y manteniendo una línea de visión directa.

6 Resultados

Para la presentación de los resultados obtenidos en la experimentación se incluyen diferentes aspectos de la comunicación, en las diferentes distancias entre el nodo gateway y el nodo master. Específicamente, en términos de tiempo de llegada entre paquetes se muestra el promedio, el máximo y el mínimo (todos en segundos). Para el análisis de la pérdida de paquetes se muestra la cantidad total de mensajes enviados y la cantidad de paquetes perdidos.

La Tabla 1 muestra los resultados obtenidos en el experimento con 1 nodo final. Las variaciones de tiempo pueden adjudicarse a que los diferentes métodos para el manejo de tiempo que tienen los nodos no son exactos, en especial el tiempo de retardo que genera el nodo final para la periodicidad del envío de paquetes. Quizás los resultados que pueden ser menos esperables se refieren a la pérdida de paquetes para 200m y 350m. Lo que se puede estimar de antemano es que a mayor distancia haya mayor pérdida de paquetes, pero no lo inverso, es decir que haya más pérdida de paquetes en 200m que en 350m. Es posible que la diferencia de ángulo respecto del nodo nodo gateway genere que la señal tenga que atravesar un área de mayor interferencia (cercano al nodo gateway hay al menos una antena direccional en esa dirección). Sin embargo, para confirmar esta hipótesis deberíamos llevar a cabo más experimentos al menos a 200m, para que el ángulo respecto del nodo gateway sea el mismo que el de 350m. Se tomaron algunas muestras a 500m con 1 nodo final y los resultados fueron relativamente esperables en cuanto a que el promedio, máximo y mínimo se mantuvieron relativamente similares a los de la Tabla 1: 1,002s, 1,02s y 0,98 respectivamente, y en 100 mensajes se perdieron 21. Se debe tener en cuenta que una frecuencia de 1 segundo entre paquetes es relativamente alta para la tecnología LoRa.

Distancia	100m	200m	350m	
Promedio	1,001	1,011	1,007	
Max	1,02	1,09	1,02	
Min	0,97	0,92	0,98	
Cant. Paquetes	200	200	200	
Cant. paq. perdidos	0	9	2	

Tabla 1 – Resultados con Un Nodo Final

La Tabla 2 muestra los resultados obtenidos en el experimento con 2 nodos finales. En este caso se debe recordar que la frecuencia es de 2 segundos y cada mensaje contiene 2 muestras. Una vez más se podrían considerar que es anómalo que se pierdan más paquetes a una distancia de 200m (17 de 100) que a una distancia de 350m (2 de 100). La razón puede ser la misma que en el caso anterior, dado que la posición del nodo master es la misma que para el experimento con 1 nodo final. Queda, por lo tanto, la validación al menos experimental de la hipótesis comentada anteriormente para el experimento con 1 único nodo final.

Distancia	100m	200m	350m
Promedio	2,008	2,021	2,005
Max	2,18	2,73	2,17
Min	1,98	1,99	1,98
Cant. paquetes	100	100	100
Cant. paq. perdidos	0	17	2

Tabla 2 – Resultados con Dos Nodos Finales

Como se comentó anteriormente, las frecuencias con las que se hicieron los experimentos fueron relativamente muy altas para la comunicación LoRa. En parte, para abreviar el tiempo de los experimentos propiamente dichos, y en parte para verificar lo que sucede en la práctica para frecuencias altas de comunicaciones con esta tecnología. Muchas aplicaciones con las características detalladas de alimentación por batería no necesariamente tienen requerimientos de muy alta frecuencia para los mensajes.

7 Conclusiones y trabajo futuro

Este trabajo muestra el funcionamiento de dos tecnologías de comunicación inalámbrica aplicadas dentro de una misma red de Sensores. Se realizaron distintos análisis del funcionamiento de estas tecnologías dentro del hardware utilizado. Como mínimo, se ha construido la red y se ha verificado la factibilidad de su funcionamiento. Asimismo, se tiene una mínima caracterización de tiempo y pérdida de paquetes.

Queda pendiente en este trabajo la medición del consumo energético de cada tipo de nodo con algún equipamiento externo (como por ejemplo, el sensor INA219 de Adafruit). En función de esas mediciones se podría, por ejemplo, utilizar modos bajo consumo, o disminuir la periodicidad de envío de datos. Dado que la placa Heltec ESP32 LoRa, cuenta también con un módulo Bluetooth de bajo consumo, también queda como prueba a futuro, integrar a la red de sensores módulos de esta tecnología (por ejemplo HM-10), para evaluar las distancias y velocidades de transmisión. Esto también lleva a la interrogante de cómo afectaría el funcionamiento conjunto de WiFi y Bluetooth dentro la misma placa trabajando sobre la misma frecuencia.

Por último, puede ser de interés profundizar en las pruebas de la tecnología LoRa, esto implica la utilización de módulos más complejos, que soporten LoRaWAN, combinados

con gateways que permitan funcionar como concentradores de la red. La utilización de nuevos módulos permitiría ampliar las distancias de funcionamiento de la red y desarrollos de mayor escala, por ejemplo aplicaciones orientadas Internet de las cosas y las ciudades inteligentes.

Bibliografía

1. Sohraby, K., Minoli, D., Znati T.: Wireless sensor networks: technology, protocols, and applications. John Wiley & Sons (2007).
2. Raghavendra, C. S., Sivalingam, K. M., Znati, T.: Wireless sensor networks. Springer (2006).
3. Akyildiz, I., Vuran, M.: Wireless sensor networks(Vol. 4). John Wiley & Sons (2010).
4. ESP8266EX Datasheet Version 5.8, Espressif Systems. Año 2018. Disponible en https://www.espressif.com/sites/default/files/documentation/0a-esp8266ex_datasheet_en.pdf
5. ESP8266 Technical Reference Version 1.3, Espressif Systems. Año 2017. Disponible en https://www.espressif.com/sites/default/files/documentation/esp8266-technical_reference_en.pdf
6. ESP32 Datasheet Version 2.3, Espressif Systems. Año 2018. https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf
7. ESP32 Technical Reference Manual Version 3.4, Espressif Systems. Año 2018. https://www.espressif.com/sites/default/files/documentation/esp32_technical_reference_manual_en.pdf
8. SX1272/73 Datasheet Rev 3.1, Semtech. Año 2017. https://www.semtech.com/uploads/documents/sx1272.pdf
9. Documentación NodeMCU, www.nodemcu.com
10. Documentación LoRa, www.semtech.com
11. Documentación LoRaWAN, www.lora-alliance.org
12. Documentación WIFI LoRa 32, http://www.heltec.cn
13. IEEE Standard for Information technology-Telecommunications and information exchange between systems Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, in IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012), doi: 10.1109/IEEESTD.2016.7786995, IEEE (2016)
14. IEEE Standard for Low-Rate Wireless Networks, in IEEE Std 802.15.4-2015 (Revision of IEEE Std 802.15.4-2011), doi: 10.1109/IEEESTD.2016.7460875, IEEE, (2016).
15. Topología LoRa, https://www.semtech.com/technology/lora/what-is-lora

Workshop Innovación en Sistemas de Software (WISS)

Coordinadores

Pablo Fillottrani (UNS) - Marcelo Estayno (UNLZ) - Alicia Mon (ITBA) - Dante Zanarini (UNR)

Temas

Dispositivos móviles (celulares, smartphones, PocketPCs, TabletPC, etc)
Software para la Web Semántica
Nuevas herramientas y métodos para ingeniería de software
Computación ubicua
Interfaces emergentes
Ingeniería de ontologías
Nuevas tendencias en ingeniería de software
Computación colaborativa (e-commerce, e-government, e-business, etc)

Propuesta de Proceso para Selección de Arquitectura de un Modelo Predictivo Inteligente

Cinthia Vegega[1,2], Luciano Straccia[1,2], Pablo Pytel[1,2] & María Florencia Pollo-Cattaneo[1,2]

[1] Grupo de Estudio en Metodologías de Ingeniería en Software (GEMIS). Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Argentina.

[2] Programa de Maestría en Ingeniería en Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina

cinthiavg@yahoo.com.ar; {ppytel, flo.pollo}@gmail.com

Resumen. Diariamente, los gerentes deben tomar decisiones sobre cómo asignar sus recursos. Estas decisiones deben basarse en predicciones sobre el tiempo, el esfuerzo y/o los riesgos de llevar a cabo sus actividades. Actualmente esta situación se encuentra acentuada por el entorno complejo que rodea a las organizaciones lo cual provoca que deban actuar más allá de los sistemas tradicionales de gestión incorporando nuevos mecanismos, como los provistos por la Inteligencia Artificial. En este caso, se estaría desarrollando un Modelo Predictivo Inteligente. Se ha observado una dificultad asociada a la construcción de estos modelos al tratar de determinar el tipo de arquitectura a aplicar. Por lo tanto, el presente trabajo tiene como objetivo proponer un proceso para llevar a cabo la selección de la arquitectura más apropiada de acuerdo a las características asociadas al proyecto y su contexto, a los datos disponibles y a los resultados esperados que se generen.

Palabras Claves: Modelo Predictivo. Red Neuronal Artificial. Red Bayesiana. Toma de Decisiones.Organización.

1. Introducción

La toma de decisiones es un proceso esencial en la vida de las organizaciones. Aunque cada miembro dentro de una organización toma decisiones, este proceso es particularmente importante para los roles gerenciales. Es por esto, que a los gerentes se los conoce como 'tomadores de decisiones' en sus actividades de planear, organizar, dirigir y controlar [1]. Diariamente deben decidir cómo asignar sus valiosos recursos basados en predicciones [2] sobre el tiempo, el esfuerzo y/o los riesgos que implica llevar a cabo sus actividades. Esta situación se encuentra acentuada debido al entorno altamente complejo y de difícil predicción del siglo XXI [3], lo cual genera que las organizaciones deban actuar más allá de los sistemas tradicionales de gestión e incorporar nuevos mecanismos para la “creación y potenciación del conocimiento organizativo", tales como los provistos por la Inteligencia Artificial [4, 5].

Un caso de estos nuevos mecanismos se puede encontrar en el ámbito de los Modelos Predictivos. A pesar de que tradicionalmente se han aplicado Técnicas Estadísticas y Modelos Paramétricos para generar predicciones [6], en las dos últimas décadas se

han incorporado diversos métodos asociados al Aprendizaje Automático [2, 7]. Estos métodos son considerados normalmente como 'cajas negras' o modelos basados en datos [8, 9, 10], debido a que aplican procedimientos (totalmente) automáticos para generar modelos no-lineales que describen la relación entre diferentes atributos de un conjunto de datos. De esta manera, es posible construir modelos que permitan encontrar la relación entre situaciones pasadas y futuras usando los datos históricos disponibles. En este sentido, se pueden destacar las Redes Neuronales Artificiales o RNA [11, 12] y, las Redes Bayesianas o RB [13, 14] como las principales arquitecturas de Sistemas Inteligentes a ser utilizadas para este tipo de problema [15-18]. Estos Modelos Predictivos, basados en Sistemas Inteligentes, presentan características muy útiles, tales como son la capacidad de generalización, robustez y auto-organización [3, 19]. Sin embargo, una dificultad asociada a la construcción de los Modelos Predictivos Inteligentes tiene que ver con determinar el tipo de arquitectura a aplicar. De acuerdo al Teorema de 'No Free Lunch' [20, 21], no existe ningún algoritmo de aprendizaje automático que pueda ser aplicado para cualquier problema, por lo que se debería siempre llevar a cabo una selección, teniendo en cuenta los objetivos planteados, las características del dominio y de los datos disponibles [22]. Por consiguiente, todo ésto debe ser definido al comienzo del proyecto.

En este contexto, el presente trabajo tiene como objetivo proponer un proceso para llevar a cabo la selección de la arquitectura que permita evaluar la adecuación de cada una de ellas y así determinar cuál es la más apropiada. Aunque en esta propuesta se ha decidido acotar el alcance a dos tipos de Sistemas Inteligentes, las RNA y las RB, ninguna de ellas es la panacea [17, 18]. Cada una tiene sus puntos fuertes y débiles que deberán ser contrastadas contra los elementos del problema para determinar cuál es la mejor solución a aplicar. Para ello, primero se detallan las condiciones a considerar para evaluar cada arquitectura, este caso RNA y RB (sección 2). Luego se presenta el proceso propuesto (sección 3), realizando dos pruebas de concepto para su validación (sección 4). Finalmente se indican las conclusiones obtenidas y futuras líneas de trabajo (sección 5).

2. Condiciones a considerar al Evaluar una Arquitectura

Para poder proponer un modelo que permita la evaluación de una arquitectura, primero es necesario identificar las principales condiciones que una tecnología debe cumplir en un proyecto para ser considerada como adecuada para la construcción de un Modelo Predictivo. A partir de la investigación documental realizada en [2, 6, 17, 22-33], se determinan los conceptos de interés para evaluar las arquitecturas a fin de construir el Modelo Predictivo y así seleccionar la que resulte más apropiada de acuerdo a las particularidades del proyecto. A continuación, se presentan dichas condiciones para cada una de las arquitecturas consideradas:

- Se considera apropiado ***utilizar una RNA o una RB*** para la implementación del Modelo Predictivo si se cumplen las siguientes condiciones:
  - Los datos disponibles son representativos del problema que se quiere resolver.
  - Los datos disponibles son suficientemente complejos con una relación no lineal entre sus atributos, por lo que las técnicas estadísticas tradicionales no permitirían encontrar una buena solución al problema.

  - El problema a ser resuelto mediante el Modelo Predictivo no va a cambiar en el corto plazo por lo que se considera como estable.
- Además de las condiciones anteriores, se considera apropiado ***utilizar una RNA*** si se cumplen las siguientes condiciones:
  - Los datos disponibles incluyen gran cantidad de ejemplos para realizar el entrenamiento y validación de la red, y además poseen todos sus valores en forma numérica (continuos o discretos).
  - La precisión de los resultados es muy importante y se puede considerar como crítica para asegurar el éxito del modelo.
  - No se necesita conocer cómo se obtuvieron los resultados generados, en otras palabras, no es necesario que la red explique o justifique los resultados obtenidos.
  - No se requiere que la red represente las relaciones entre los datos para poder realizar ajustes manuales, por lo que no necesario contar con la asistencia de expertos del dominio para participar en el desarrollo.
- En cambio, es apropiado ***utilizar una RB*** si a las condiciones iniciales se agregan las siguientes:
  - Los datos disponibles incluyen una cantidad suficiente de ejemplos para realizar el entrenamiento y validación de la red y. poseen sus valores como valores alfanuméricos, numéricos discretos, o numéricos continuos para los cuales es posible definir rangos.
  - En el caso que los datos disponibles son poco representativos o no incluyen una cantidad suficiente de ejemplos, se dispone de expertos del dominio que pueden aportar su conocimiento para participar en el desarrollo (especialmente en el modelado de la red).
  - La precisión de los resultados no es tan importante porque se valoran también otras cuestiones tales como: poder comparar las predicciones para diferentes posibilidades o escenarios; conocer cómo se obtuvieron los resultados generados, es decir que, la red explique o justifique los resultados; poder realizar ajustes en la red en base al conocimiento disponible sobre el problema y las relaciones entre los atributos de los datos.

3. Proceso Propuesto

Para definir el proceso propuesto se consideran las características del proyecto, las cuales se han organizado en tres grupos: las asociadas al problema detectado y su contexto, las asociadas a los datos disponibles para el entrenamiento y, las correspondientes a los resultados esperados que genere el modelo. Por lo tanto, el proceso necesita que se responda a un conjunto de preguntas cuyas respuestas permitan caracterizar al proyecto. Sin embargo, al comienzo de un proyecto no resulta sencillo contestar estas preguntas con un adecuado grado de certeza (proporcionando respuestas del tipo 'sí' / 'no' o, con un valor numérico). En virtud de estas consideraciones, el proceso propuesto se basa en el principio de los Sistemas Expertos Difusos [34], permitiendo de esta manera, manejar un rango de cinco valores lingüísticos (entre '*nada*' y '*todo*') y, de esta manera dar respuesta a cada una de las preguntas a considerar. Al hacer uso de un procedimiento sencillo es posible transformar estos valores lingüísticos, indicados por el ingeniero, en intervalos difusos que luego serán utilizados para obtener una valoración sobre la adecuación de cada arquitectura. A continuación, se describen los cinco pasos que se deben realizar (los cuales han sido implementados en la planilla disponible en [35]):

Paso 1: Determinar el valor correspondiente para las características del proyecto
Para caracterizar un proyecto de construcción de un Modelo Predictivo, y evaluar la arquitectura más adecuada, se utilizan las características definidas en la Tabla 1, las cuales están basadas en las condiciones indicadas en la sección 2. El ingeniero debe responder las preguntas asociadas a cada característica a partir del resultado de las entrevistas realizadas en la organización. Los valores lingüísticos permitidos para las respuestas son: '*nada*', '*poco*', '*regular*', '*mucho*' y '*todo*', donde cuanto más verdadera parezca una característica, mayor valor se le debe asignar; y, cuanto más falsa parezca, menor valor. Nótese que por cada característica, el método tiene asignado un valor de umbral mínimo que el proyecto debe igualar o superar. En caso que el valor asignado a la característica no supere el umbral, se puede considerar que ninguna de las arquitecturas es apropiada para ser aplicada y, por lo tanto, no es necesario continuar con los pasos siguientes.

Tabla 1. Características del Proyecto a ser Evaluadas por el Proceso Propuesto.

Categoría	ID	Pregunta Asociada a la Característica	Umbral	Peso RNA	Peso RB
Datos Disponibles	D1	¿Cuánta confianza se tiene que los datos son representativos?	regular	10	10
	D2	¿En qué medida se pueden considerar a los datos como complejos y con una relación no lineal entre sus atributos?	poco	9	6
	D3	¿Cuánta cantidad de ejemplos incluyen los datos?	poco	15	7
	D4	¿Cuál es la proporción de datos con valores numéricos continuos (con respecto a valores no numéricos o numéricos discretos)?	nada	10	-7
Resultados Esperados	R1	¿Cuán crítica se considera la precisión de la predicción?	nada	12	-3
	R2	¿En qué medida se desea conocer comparar las predicciones para diferentes posibilidades o escenarios?	nada	-4	12
	R3	¿Con qué grado de importancia se desea que se puedan explicar cómo se obtuvieron los resultados generados?	nada	-13	15
Dominio del Problema	P1	¿Qué tan estable es el problema a resolver?	regular	10	10
	P2	¿Cuánta disponibilidad tienen los expertos del dominio para participar en el desarrollo?	nada	-6	14
	P3	¿Qué tan deseable es poder ajustar manualmente la red a partir de los conocimientos sobre los datos?	nada	-11	16

Donde, por cada característica, se definen los siguientes atributos:

- *Categoría*, que es utilizado únicamente para poder agrupar las características de acuerdo a qué o quién se refiere.
- *ID*, que indica el código para identificar unívocamente a la característica.
- *Pregunta Asociada a la Característica*, que describe la condición que debe responder el ingeniero con el valor lingüístico correspondiente.
- *Umbral*, que define el valor que la característica debe igualar o superar.
- *Peso RNA*, que indica la importancia relativa a cada característica en la evaluación de la adecuación de las *Redes Neuronales Artificiales*.
- *Peso RB*, que indica la importancia relativa a cada característica en la evaluación de la adecuación de las *Redes Bayesianas*.

Asimismo, se puede observar en la Tabla 1 que existen diferentes valores de pesos por cada característica y tipo de arquitectura ('*Peso RNA*' y '*Peso RB*'). Dichos valores han sido asignados mediante la evaluación de las condiciones antes indicadas junto con la experiencia de ingenieros en la construcción de Sistemas Inteligentes. Como se puede notar, si los valores de los pesos son mayores a cero, significa que la característica es favorable para la arquitectura, por lo que cuanto mayor valor tenga

dicha característica más debería aumentar la adecuación de dicha arquitectura. Sin embargo, si el peso es menor a cero, significa que la característica es desfavorable, por lo que su aumento disminuirá su adecuación. Por último, cabe aclarar que esta diferencia entre los valores de los pesos son soportados por los siguientes pasos del proceso (siempre que la suma de los valores absolutos por cada arquitectura sea igual a 100).

Paso 2: Convertir los valores en intervalos difusos

Una vez que para cada característica de la Tabla 1 se han asignado los valores lingüísticos correspondientes, estos valores se deben traducir a los intervalos difusos correspondientes, dependiendo del valor del peso de la característica. En este sentido, a cada valor lingüístico se le define un intervalo difuso expresado por cuatro valores numéricos (entre cero y diez), que representan los puntos de ruptura (o puntos angulares) de su función de pertenencia correspondiente. Dichos intervalos, junto con la representación gráfica de la función de pertenencia, se indican en la Figura 1.

Cuando el valor de los pesos para las características es mayor o igual a cero, se asigna directamente el intervalo difuso al valor lingüístico asignado por el ingeniero. Pero, cuando el valor del peso de la característica es negativo, la asignación no es tan directa. En este caso, se debe asignar el intervalo difuso del valor opuesto al asignado por el ingeniero. Entonces, si el valor asignado es '*nada*' se debe utilizar el intervalo correspondiente a '*todo*', si el valor asignado es '*poco*' se utiliza el intervalo de '*mucho*', si es '*mucho*' se utiliza el de '*poco*', y si es '*todo*' se usa el de '*nada*'. Nótese que para el valor '*regular*' no existe un valor opuesto, por lo que se usa su intervalo difuso directamente (es decir, el opuesto de '*regular*' es '*regular*').

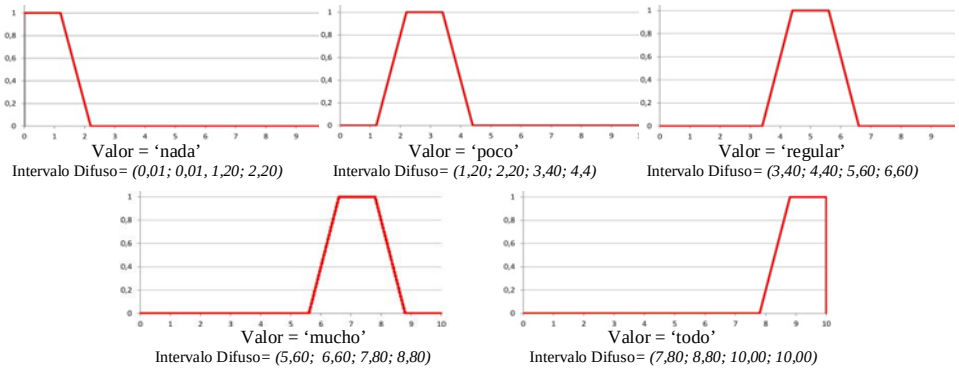


Fig. 1. Representación de los Intervalos Difusos para los Valores Lingüísticos.

Paso 3: Calcular la valoración de cada arquitectura

Una vez obtenidos los intervalos difusos por cada característica y arquitectura en el paso anterior, los mismos son utilizados para calcular la valoración de cada arquitectura teniendo en cuenta el valor del peso correspondiente (los cuales han sido indicados en la Tabla 1). Esto se realiza aplicando la fórmula #1 de la Tabla 2. Esta fórmula está compuesta por la combinación de la media armónica y la media aritmética del conjunto de intervalos. De esta forma, se busca reducir la influencia de valores bajos en el cálculo de la dimensión. Como resultado se obtiene un intervalo que representa la adecuación de cada arquitectura (A_R). Dado que el resultado de dicha fórmula es

otro intervalo difuso, para convertirlo en un único valor numérico (V_T) se utiliza la media aritmética de los valores del intervalo, como se indica en la fórmula #2 de la Tabla 2.

Tabla 2. Fórmulas Utilizadas para Calcular la Valoración de cada Arquitectura.

#	Fórmula	
1	$A_R = \left(\frac{1}{2} \cdot \frac{\sum_{i=1}^{n_R} \lvert P_{R_i} \rvert}{\sum_{i=1}^{n_R} \left(\frac{\lvert P_{R_i} \rvert}{C_{R_i}} \right)} \right) + \left(\frac{1}{2} \cdot \frac{\sum_{i=1}^{n_R} (\lvert P_{R_i} \rvert \cdot C_{R_i})}{\sum_{i=1}^{n_R} \lvert P_{R_i} \rvert} \right)$	Donde: A_R: representa el intervalo difuso de la adecuación para la arquitectura R (usando como nomenclatura 'RNA' para Redes Neuronales Artificiales y 'RB' para Redes Bayesianas). $\lvert P_{Ri} \rvert$: representa el valor absoluto del peso de la característica i perteneciente a la arquitectura R. C_{Ri}: representa el intervalo difuso asignado a la característica i perteneciente a la arquitectura R. n_R: representa la cantidad de características asociadas a la arquitectura R.
2	$V_R = \frac{\sum_{j=1}^{4} (I_R\{j\})}{4}$	Donde: V_R representa el valor numérico calculado para la arquitectura R. $I_R\{\}$: representa el valor correspondiente a la posición j del intervalo difuso calculado para la arquitectura R.

<u>Paso 4:</u> Interpretar los resultados obtenidos

Una vez obtenidos los valores correspondientes a cada arquitectura en el paso 3, éstos deben ser analizados e interpretados. Para ello, se recomienda graficar los intervalos correspondientes a cada arquitectura, utilizando la función de pertenencia correspondiente a su intervalo difuso (A_R). Se considera que, una dimensión está aceptada si el gráfico obtenido supera al intervalo del valor '*regular*' (indicado en la Figura 1). En forma análoga, esto se puede determinar si se utiliza el valor numérico asociado a la arquitectura: una arquitectura será aceptable si su valor V_R es mayor a 5.

<u>Paso 5:</u> Seleccionar la mejor arquitectura

Finalmente, teniendo en cuenta la interpretación de los resultados obtenidos (paso 4), se procede a seleccionar la mejor tecnología para el proyecto aplicando el criterio definido en la tabla de decisión de la Tabla 3.

Tabla 3. Tabla de Decisión para Seleccionar la Mejor Arquitectura.

<table>
<tr><th colspan="2" rowspan="2">Selección de la Mejor Arquitectura</th><th colspan="3">Valor numérico asociado a Redes Neuronales Artificiales (V_{RNA})</th></tr>
<tr><th colspan="2">aceptable ($V_{RNA} > 5$)</th><th>no aceptable ($V_{RNA} \leq 5$)</th></tr>
<tr><td rowspan="3">Valor numérico asociado a Redes Bayesianas (V_{RB})</td><td rowspan="2">aceptable ($V_{RB} > 5$)</td><td>$\lvert V_{RNA} - V_{RB} \rvert < 0,25$</td><td colspan="2">$\lvert V_{RNA} - V_{RB} \rvert \geq 0,25$</td><td rowspan="2">selecciona RB</td></tr>
<tr><td></td><td>$V_{RNA} > V_{RB}$</td><td>$V_{RNA} < V_{RB}$</td></tr>
<tr><td>selecciona AMBAS</td><td>selecciona RNA</td><td>selecciona RB</td></tr>
<tr><td>no aceptable ($V_{RB} \leq 5$)</td><td colspan="3">selecciona RNA</td><td>selecciona NINGUNA</td></tr>
</table>

Como se puede notar en la Tabla 3, se puede decidir que se seleccione:

- *a una de las arquitecturas, Redes Neuronales Artificiales* (*RNA*) ó *Redes Bayesianas (RB*), si esa arquitectura tiene un valor de adecuación aceptable y es muy superior al de la otra arquitectura (con una diferencia absoluta mayor o igual a 0,25);
- *ambas arquitecturas*, si ambos valores de adecuación son aceptables con una diferencia absoluta menor a 0,25 por lo que se consideran como equivalentes y cualquiera de las dos se podría utilizar para implementar el modelo; o

- *ninguna de las arquitecturas*, si ninguno de los valores de adecuación es aceptable, lo cual significa que se debería utilizar otro tipo de arquitectura.

4. Pruebas de Concepto

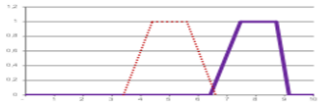
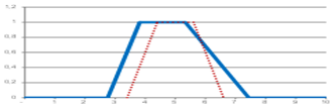
En esta sección, se presentan dos pruebas de concepto para ilustrar el funcionamiento del proceso propuesto. Se utiliza una prueba de concepto en la que se selecciona a una Red Neuronal Artificial (RNA) como mejor alternativa y, otra en la que se selecciona a una Red Bayesiana (RB). Todos los cálculos necesarios fueron realizados mediante la planilla creada ad-hoc [35] que implementa los pasos definidos anteriormente para cada una de las pruebas de concepto (los cuales están disponibles en [36]).

Como primera prueba de concepto se evalúa el caso presentado en [30], el cual tiene como objetivo realizar el pronóstico de las ventas de una empresa dedicada a la comercialización de productos alimenticios. Para llevar a cabo la evaluación, se aplica el proceso propuesto en la sección 3. A partir de la información indicada en la publicación se responden las preguntas requeridas con el valor lingüístico correspondiente (paso 1), tal como se muestran en la Tabla 4. Luego, los valores lingüísticos son convertidos en intervalos difusos (paso 2) y se calcula el intervalo de cada arquitectura (paso 3), la cual se representa en forma gráfica, como se muestra en la Tabla 5. A partir de esos intervalos, se calcula el valor numérico de cada arquitectura y se interpreta (paso 4). En este caso, se obtiene 7,94 para Redes Neuronales Artificiales, lo cual es superior a 5, por lo que se considera como aceptable. En cambio, para las Redes Bayesianas corresponde 4,83 que, aunque cercano, es inferior a 5 por lo que no es aceptable. Por consiguiente, se recomienda aplicar a las Redes Neuronales Artificiales como la arquitectura más apropiada, lo cual es consistente con las conclusiones de la publicación donde se indica que dicha tecnología es la que mejores resultados ha generado.

Tabla 4. Asignación de las Características de la Primera Prueba de Concepto.

Categoría	ID	Respuesta	Valor Asig.
Datos Disponibles	D1	Los datos se corresponden a históricos de la empresa para los años 2010 a 2013, por lo que se consideran totalmente representativos del problema a predecir.	todo
	D2	Según se indica, todos los datos disponibles para realizar el pronóstico de ventas presentan un comportamiento no lineal.	todo
	D3	Aunque la cantidad de ejemplos disponible no es muy grande (uno por cada mes de los cuatro años considerados, es decir, 48 registros en total) se consideran suficientes para el problema a resolver.	regular
	D4	Dado que se utiliza el monto total de ventas para cada mes, la mayoría los atributos tienen valores continuos.	mucho
Resultados Esperados	R1	Dado que el objetivo es comparar la precisión de una RNA para pronosticar la cantidad de ventas, se considera como totalmente crítica.	todo
	R2	No se ha solicitado poder realizar una comparación de diferentes situaciones o escenarios.	nada
	R3	No se ha solicitado que se expliquen los resultados generados por el modelo.	nada
Dominio del Problema	P1	El problema se considera muy estable para poder ser resuelto.	mucho
	P2	No se cuenta con expertos disponibles para solicitar información sobre los datos o el problema.	nada
	P3	No se ha solicitado ajustar manualmente la red a partir de los conocimientos sobre los datos.	nada

Tabla 5. Cálculo de Intervalos para las Arquitecturas de la Primera Prueba de Concepto.

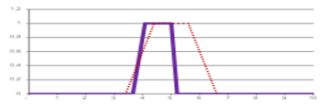
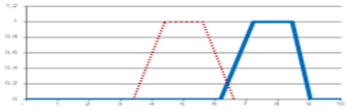
Arq.	Intervalo de la Arquitectura (A_R)	Representación de la Función de Pertenencia de A_R
A_{RNA}	(6,41; 7,47; 8,71; 9,15) Intervalo sobrepasa al intervalo de '*regular*' y por una pequeña diferencia al de '*mucho*'.	
A_{RB}	(2,75; 3,82; 5,33; 7,43) Intervalo no sobrepasa al intervalo de '*regular*' pero supera por una pequeña diferencia al de '*poco*'.	

Para llevar a cabo el análisis de la segunda prueba de concepto se aplica el proceso para evaluar el caso presentado en [25]. En dicho trabajo, se combinan los conocimientos de expertos del dominio con los datos de cuestionarios realizados a pacientes para la construcción de una Red Bayesiana que permita asistir la toma de decisiones médicas. Nuevamente, a partir de la información indicada en la publicación se respondieron las preguntas requeridas como se muestra en la Tabla 6. Con estos valores, se vuelve a aplicar el resto de los pasos del proceso obteniéndose los resultados que se muestran en la Tabla 7.

Tabla 6. Asignación de las Características de la Segunda Prueba de Concepto.

Categoría	ID	Respuesta	Valor Asig.
Datos Disponibles	D1	Dado que los datos fueron obtenidos de cuestionarios a pacientes, no se consideran muy confiables pero sí lo suficientes para ser utilizados como base.	regular
	D2	Por su naturaleza, los datos a ser utilizados son muy complejos con una relación muy variable.	todo
	D3	Se cuenta con muchos datos obtenidos prevenientes varios cuestionarios (la cantidad de variables es mayor a 900 atributos)	todo
	D4	Las respuestas a las preguntas de los cuestionarios son valores alfanuméricos (sí/no o uno de una lista de valores), por lo que no se cuenta con valores continuos.	nada
Resultados Esperados	R1	Aunque se considera importante tener una buena precisión en los resultados generados, la misma no se considera crítica.	regular
	R2	Se considera de interés poder comparar resultados para diferentes escenarios.	mucho
	R3	También es muy deseable que se justifiquen los resultados generados a partir de los resultados parciales generados por el modelo.	mucho
Dominio del Problema	P1	Dado que las preguntas tienen que ver con estudios realizados desde hace mucho tiempo, el problema se considera muy estable.	mucho
	P2	Se cuenta con varios expertos cooperativos para asistir en el desarrollo.	todo
	P3	Dado el objetivo de la publicación, se va a utilizar los conocimientos de los expertos disponibles para ajustar la red generada.	todo

Tabla 7. Cálculo de Intervalos para las Arquitecturas de la Segunda Prueba de Concepto.

Arq.	Intervalo de la Arquitectura (A_R)	Representación de la Función de Pertenencia de A_R
A_{RNA}	(3,65; 4,08; 5,00; 5,21) Intervalo no sobrepasa al intervalo de *'regular'* pero supera por una pequeña diferencia al de '*poco*'.	
A_{RB}	(6,18; 7,22; 8,45; 9,03) Intervalo sobrepasa al intervalo de '*regular*' y por una pequeña diferencia al de '*mucho*'.	

Finalmente, se calculan e interpretan los valores numéricos de cada arquitectura. En este caso, se obtiene 4,49 para Redes Neuronales Artificiales y 7,72 para Redes Bayesianas, por lo que la primera se considera como no aceptable pero la segunda sí lo es. Por consiguiente, y con una diferencia significativa, la tecnología recomendada como la más apropiada para este proyecto son las Redes Bayesianas, lo cual es también consiste con las conclusiones de la publicación analizada.

5. Conclusiones

Este trabajo presenta la propuesta de un proceso para la selección de la arquitectura más adecuada con el fin de construir un Modelo Predictivo Inteligente, dada que la dificultad asociada a la construcción de este tipo de modelos tiene que ver con determinar el tipo de arquitectura a aplicar, de acuerdo a las características del proyecto.

El proceso propuesto permite evaluar la adecuación de cada característica del proyecto, aplicando información difusa y de esta manera, determinar cuál se considera la más recomendable a aplicar. Para facilitar su utilización, también se han implementado todos sus pasos en la planilla [35], la cual ha sido utilizada para validar exitosamente el proceso con dos pruebas de concepto, una para RNA y otra para RB.

Queda como futura línea de trabajo extender el proceso para que pueda evaluar la adecuación de otras tecnologías. Para ello, solo se necesitan definir los nuevos valores de los pesos correspondientes. Si fuera necesario también incorporar nuevas características, ello se puede realizar sin tener que modificar las fórmulas ni los pasos del proceso. Sólo se debe agregar la característica asignando un peso de cero para las arquitecturas actuales (de forma que no se modifique el comportamiento actual) y asignando los pesos correspondientes para la nueva arquitectura.

Referencias

1. Robbins, S. & Coulter, M. (2010). *Administración*. *Décima Edición*. Prentice Hall. ISBN: 978-607-442-388-4.
2. Mair, C., Kadoda, G., Lefley, M., Phalp, K., Schofield, C., Shepperd, M. & Webster, S. (2000*). An investigation of machine learning based prediction systems*. Journal of Systems and Software, 53(1), 23-29.
3. García Martínez, R., Servente, M. & Pasquini, D. (2003). *Sistemas Inteligentes*. Editorial Nueva Librería. Buenos Aires. ISBN 987-1104-05-7.
4. Nilsson, N. J. (2014). *Principles of Artificial Intelligence*. Morgan Kaufmann.
5. Russell, S. J., Norvig, P., Davis, E., Russell, S. J. & Russell, S. J. (2010). *Artificial Intelligence: a modern approach* (Vol. 2). Englewood Cliffs: Prentice Hall.
6. Shepperd, M. & Kadoda, G. (2001). *Comparing software prediction techniques using simulation*. IEEE Transactions on Software Engineering, 27(11), 1014-1022.
7. Bontempi, G., Taieb, S. B. & Le Borgne, Y. A. (2013). *Machine learning strategies for time series forecasting*. In Business Intelligence (pp. 62-77). Springer Berlin Heidelberg.
8. Alpaydin, E. (2014) *Introduction to machine learning*. MIT press.
9. Benítez, J. M., Castro, J. L. & Requena, I. (1997). *Are artificial neural networks black boxes?*. IEEE Transactions on neural networks, 8(5), 1156-1164.
10. Mitchell, T.M. (1997) *Machine Learning*. McGraw-Hill, New York.
11. Wang, S. C. (2003). *Artificial Neural Network*. In Interdisciplinary Computing in Java Programming (pp. 81-100). Springer US.

12. Wang, L. & Fu, K. (2009). *Artificial Neural Networks*. Wiley Encyclopedia of Computer Science and Engineering. 181–188.
13. Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. The MIT Press.
14. Premchaiswadi, W. (2012) *Bayesian Networks*. Ed. In-Tech.
15. Chatfield, C. (2016). *The analysis of time series: an introduction*. CRC press.
16. De Gooijer, J. G. & Hyndman, R. J. (2006). *25 years of time series forecasting*. International journal of forecasting, 22(3), 443-473.
17. Zhang, G. & Hu, M.Y. (1998). *Neural network forecasting of the British pound/US dollar exchange rate*. Omega, International Journal of Management Science, 26 (4) 495–506.
18. Zhang, G., Patuwo, B. E. & Hu, M. Y. (1998). *Forecasting with artificial neural networks*: *The state of the art*. International journal of forecasting, 14(1), 35-62.
19. Cohen, P. R. & Feigenbaum, E. A (2014*). The handbook of Artificial Intelligence*. Vol. 3. Butterworth-Heinemann.
20. Wolpert, D. H. (1996). *The lack of a priori distinctions between learning algorithms*. Neural computation, 8(7), 1341-1390.
21. Wolpert, D. H. & Macready, W. G. (1997). *No free lunch theorems for optimization*. IEEE transactions on evolutionary computation, 1(1), 67-82.
22. Domingos, P. (2012). *A few useful things to know about machine learning.* Communications of the ACM, 55(10), 78-87.
23. Adya, M. & Collopy, F. (1998). *How effective are neural networks at forecasting and prediction? A review and evaluation*. J. Forecasting, 17, 481-495.
24. Bontempi, G., Taieb, S. B. & Le Borgne, Y. A. (2013). *Machine learning strategies for time series forecasting*. In Business Intelligence (pp. 62-77). Springer Berlin Heidelberg.
25. Constantinou, A. C., Fenton, N., Marsh, W. & Radlinski, L. (2016). *From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support.* Artificial intelligence in medicine, 67, 75-93.
26. Chen, S. H. & Pollino, C. A. (2012). *Good practice in Bayesian network modelling. Environmental* Modelling & Software, 37, 134-145.
27. Masegosa, A. R. & Moral, S. (2013). *An interactive approach for Bayesian network learning using domain/expert knowledge*. International Journal of Approximate Reasoning, 54(8), 1168-1181.
28. Mendes, E. (2011). *Knowledge representation using Bayesian networks - A case study in Web effort estimation*. In Information and Communication Technologies (WICT), 2011 World Congress on (pp. 612-617). IEEE.
29. Pitchforth, J. & Mengersen, K. (2013). *A proposed validation framework for expert elicited Bayesian Networks*. Expert Systems with Applications, 40(1), 162-167.
30. Ruelas-Santoyo, E. A. & Laguna-González, J. A. (2014) *Comparación de predicción basada en redes neuronales contra métodos estadísticos en pronósticos de ventas*. Ingeniería Industrial. Actualidad y Nuevas Tendencias, IV (Enero-Junio). ISSN 1856-8327.
31. Trujillano, J., March, J. & Sorribas, A. (2004*). Aproximación metodológica al uso de redes neuronales artificiales para la predicción de resultados en medicina*. Med Clin (Barc), 122(s1).
32. Woodberry, O., Nicholson, A. E., Korb, K. B. & Pollino, C. (2004). *Parameterising bayesian networks*. In Australasian Joint Conference on Artificial Intelligence (pp. 1101-1107). Springer Berlin Heidelberg.
33. Zhou, Y., Fenton, N. & Neil, M. (2014). *Bayesian network approach to multinomial parameter learning using data and expert judgments*. International Journal of Approximate Reasoning, 55(5), 1252-1268.
34. Jang, J.S.R. 1997. *Fuzzy inference systems*. Upper Saddle River, NJ: Prentice-Hall.
35. Vegega, C., Pytel, P. & Pollo-Cattaneo, M.F. (2018a). *Implementación del Método Evaluador de las Tecnologías*. Disponible en: https://tinyurl.com/MetEvTecnol
36. Vegega, C., Pytel, P. & Pollo-Cattaneo, M.F. (2018b). *Aplicación del Método Evaluador de las Tecnologías para las Pruebas de Concepto*. Disponible en: https://tinyurl.com/pruebasConcMetEvTecnol

An Open Source Solution for Money Bill Recognition for the Visually Impaired User using Smartphones

Lucila Lang[1], Nicolas Gazcón[1,2,3], Martín Larrea[1,2,3]
[1]Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (DCIC-UNS)
[2]Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET)
[3]Laboratorio de I+D en Visualización y Computación Gráfica, (UNS-CIC Prov. de Buenos Aires)
lucilalang07@hotmail.com,nfg@cs.uns.edu.ar mll@cs.uns.edu.ar

Resumen According to estimates made by the World Health Organization there are approximately 285 million people with visual disabilities. In Argentina, it is estimated that there are about 6 thousand blind people per million inhabitants. These people must face daily different situations that generate physical and psychological tension conditioning their quality of life. In particular in Argentina, one of these situations is to recognize and correctly differentiate the denomination of Argentine money bills. The current system used is heavily affected by the usage of the money bills. In this paper, we present the design and development of a novel application for mobile phones that automatically recognize money bills and credit cards. The application was design with the assistance of a visually impaired user. *abstract* environment.

Keywords: visually impaired, image recognition, software innovation

1. Introduction

The Human-Computer Interaction field of Computer Science seeks to design systems that involve computers and support users so that they can perform tasks providing appropriate interactions to their needs and capabilities. This enables the systems to be accepted and used effectively. The term Interaction refers to the relationship between the human and the machine through an interface, which acts as an intermediary between two systems of different nature. More specifically, the interface is defined as the set of components used by users to communicate with computers ([1]). It is a translation system since since both user and computer "speak" different languages.

In order to obtain an effective communication between the user and the computer, the interfaces must be designed with the particular needs of the user to whom the system is directed. Using a valid and effective communication link between both parties is mandatory. Otherwise, interaction would not be possible. Today, with the growth in sales of smartphones ([5]), it is increasingly common

to use mobile applications that facilitate the development of people's daily lives. Technological advances have made possible the appearance of very sophisticated and easy to use software systems, based on the design of interfaces that are very intuitive for the user. To achieve this, modern interfaces are mostly graphic and based on different visual components such as icons, images, videos, and animations. Although this is very advantageous for a sighted person, what happens if that sense is not available or does not work at 100 %? In general, these designs are not conceived to meet the needs of people who are blind or visually impaired, this situation prevents information from being transmitted properly.

According to estimates made by the World Health Organization (WHO) in the world there are approximately 285 million people with visual disabilities, of which 39 million are blind, and 82 % of people who suffer from blindness are 50 years old or older ([3]). With regard to Latin America, it should be noted that of the 26.6 million people living in 2010 with visual impairment, 3.2 million were blind ([2]), and in our country it is estimated that there are about 6 thousand blind per million inhabitants ([4]). These people must face daily different situations that generate physical and psychological tension conditioning their quality of life. Tasks that for most people are simple and routine -like walking on the street, household chores, holding a conversation- can be a very arduous obstacle that is imperceptible to those who do not have this disability.

In particular, one of the situations that a person with visual disability must face when residing in Argentina is to recognize and correctly differentiate the denomination of Argentine money bills. This is because of the identification system for the blind that is currently used consists of small squares in setoff printed in chalcography. The quantity of those squares varies according to the denominations, and which wears very quickly with the use.

In this context, one of the challenges facing the software development community is to ensure that new technologies are adapted and available to all users, ensuring that people with visual impairment can benefit from them. In this work, we present the design and development of a mobile application intended for visually impaired users. The goal of this application is to automatically recognize money bill as well as different type of cards with an easy to use set of interactions.

The remainder of the article is organized as follows. Next section, "Previous Work", presents the research done and the solutions currently available for money recognition by visually impaired users. Then, in section "Proposal & Implementation" our solution is presented and details are given on how it was developed. The set of interaction available is also described. In section "Validation" we present the test cases done to probe the effectiveness of the application. Finally, in the last section, we draw some conclusions and outline future work.

2. Previous Work

Users with decrease or lack of their visual ability compensate this with information obtained by means of other senses. In particular, these people use the

sense of touch and hearing as the main source of information. Over time, there have been various methods for blind people or people with significant visual impairment to have access to information. The main one has been the braille method, a system based on points that allow, by combining them, to have as many symbols as those of the alphabet, making it possible to read through touch ([6]). However the braille system has two major disadvantages, the high cost of its material production and the volume of printed documents as they tend to become cumbersome.

Due to the listed drawbacks and the need for interaction with the user, they have had to look for other methods so that people with visual disabilities could have access to the information. In this field, tiflotechnology faces the objective of developing new technologies or adapting and providing accessibility to existing ones for their use and exploitation by people with visual disabilities ([7]).

There are several strategies on how to help a visually impaired person to identify money bills. Not all of them are based on software development, but in order to keep this paper under the maximum number of pages, we will limit our discussion to software-based solutions. There are two trends in the field of money recognition research, scanner-based systems and camera-based systems.

Scanner-based Systems are suitable for money counting machines and ATMs. Examples of these types of systems can be found in Trupti and Bawane ([8]). The recognition is made considering the size of the bill as one of the three most important characteristics, as well as color and texture. In addition, in Yaseri et al., ([9]), Ahmed and Mirfa ([10]) and Fatemeh Daraee and Saeed Mozaffari ([11]) can be found different techniques used for recognition through the use of scanners. Because our work is center on smartphones we will focus on Camera-based Systems.

Camera-based Systems allow the recognition of the money by means of capturing the bill by a smartphone camera. It is not necessary to take a picture of the ticket in its entirety, so they are very useful for people with visual disabilities. There are very few applications on this topic. Some are available on software stores such as Google Play or App Store but do not have an exact reference for their scientific contributions.

IDEAL Currency Identifier [12] is an app developed by IDEAL Group, Inc. for Android 2.2 or higher. It is an application that allows the recognition of US dollars (USD). For this, it periodically captures the frames through the camera of the device and processes them quickly to notify the user of the value of the recognized bill. It uses vibration patterns that indicate when the image was recognized and then reports it by voice. One of the advantages it presents is the automatic focus of the camera, which allows adjusting the brightness according to the environment and thus efficiently recognize bills in different lighting conditions. This application uses an image processing technique called template matching, which consists of in making a comparison by regions between stored templates and captured frames. Normally it uses two templates of 100 pixels wide and 100 pixels high, both for the front and for the back of each bill. As a result of this, it can provide information that indicates whether the scanned

part corresponds to the back or to the front of the ticket. Unfortunately, it has some limitations when bills are broken or worn, and when lighting conditions are not appropriate. In the latter case, it does not provide the option to turn on the device's flash and may give erroneous results.

LookTel Money Reader ([13]) was developed by IPPLEX and is available for iOS 8.0 or higher. It is a mobile application that works on iOS to help people identify and count money. It can notify the user, by voice, the amount of money identified. It can also show in the center of the screen with an appropriate font size, the value of the recognized ticket. It has a large database that supports twenty different currencies, including US Dollar (USD), Euro, Dirham of the United Arab Emirates, Japanese Yen, Kuwaiti Dinar. It also supports different languages such as English, Spanish, French, Japanese. It has few limitations regarding the location of the device when capturing an image and it works in real time. One of its best features is that users can start a session and the application can count the amount of money recognized during the session.

ViaOptaDaily ([14]) was developed by Novartis Pharmaceuticals Corporation for Android 4.0.3 and iOS 7.0. ViaOptaDaily is an application that provides an integrated environment to help people with disabilities. One of its feature is the money recognizer, the application can recognize Euro, US Dollar, and Saudi Riyal. Although this application provides several useful services and is compatible with different languages, such as English, French, Arabic, and Spanish, the interaction with the user is quite complicated, because the interface contains many icons and there are no voice commands that facilitate the process.

Floosy ([58]) was developed by I Can See and is available for Android 2.3 and higher. It was created to recognize the Egyptian currency. Unlike of other applications, it uses the color of the image instead of template matching to identify the ticket. For this, it captures frames periodically using the camera of the device. It extracts the characteristics of the image, compares them with their pre-calculated characteristics, and on this basis determines the corresponding value. Because it is based on the color of the image, lighting is a key problem. It behaves differently under different lighting conditions and may fail or give erroneous values when the lighting conditions are not adequate. The application allows using the flashlight and the manual focus of the camera. It must be kept in mind that this affects usability in blind or visually impaired users since it makes the process very complex.

2.1. Solutions Available for Argentine Currency

The Argentine National Institute of Industrial Technology (INTI) has developed a prototype of an open source application that allows recognizing Argentine peso bills from a tablet or cell phone, but that is still in the testing phase so it is not available and does not have an estimated release date. ([15]). There is a project called PROCER ([16]) that is still in the development phase, consisting of a device capable of converting printed text to speech for people who have some kind of reading difficulty, which can be used for the recognition of money bills. Although it is not yet available in the market, it must be kept in mind that

for a person to be able to use it, they will have to acquire the device, which not only generates an additional expense but also requires to take it with them at all times. This could generate unnecessary discomfort taking into account that most people today have a smartphone that could perform this same task.

Blind-Droid Wallet was developed by Oleg Taystruk for Android 2.3 or higher. It is a free application that allows you to recognize money bills from different countries, in particular, Argentina. It is designed especially for people with visual disabilities, so its interface does not represent a difficulty. The money is scanned by the camera and the application communicates by voice the value of it. One of the advantages that it presents is that it does not require an internet connection, so once it is installed it can be used when it is needed. However, it is not enough to install the application on the device that is going to be used since it is only preconfigured for the recognition of US dollars. In order to detect Argentine pesos, the user must download and install an additional module that is available called Blind-Droid Wallet - ARS. After doing some tests, we found out that Blind-Droid Wallet could not be configured on some devices, eg. on Samsung Galaxy A3. In the devices where it was successfully configured, 6 bills of the 16 bills available in Argentina could be correctly recognized, which represents 37.5 % of the tickets. The lack of recognition of some bill is due to the fact that the database was not updated to the new bills that have entered into circulation in our country in recent years, such as the $200 (issued on 10/26/2016) and $500 (issued on 06/29/2016). Even though the last update was made on October 7, 2015, the application does not recognize tickets that have been issued prior to that date, such as the $50 bill (issued on 03/02/2015) and the $5 bill (issued on 10/01/2015). Finally, the application is not open source, so it is not possible to update the database to the new bills in circulation.

WhatIsThis ([17]) is an Android application (2.1 or higher) created by students from Universidad Nacional de La Matanza. The application allows users with visual impairment to use their Android phones to read printed texts and recognize bills. Although the application was published on Google Play, it is not currently available for downloading, so it was not possible to install it for evaluation.

3. Proposal & Implementation

This software arises from a real need to satisfy a requirement raised by a blind person. This person, despite having managed to acquire a degree of significant autonomy in their daily lives, still has to deal with money bills that are not prepared for the visually. Based on interviews with the user and a subsequent investigation, it was determined that the situation posed was not only a particular need but was part of a common problem of people with visual impairment residing in Argentina ([18]). Based on the aforementioned, an investigation was carried out, which was developed in the previous section, where the different alternatives and software products accessible to people with visual disabilities were analyzed. Based on this, and taking into account that the Android ope-

rating system heads the ranking of the most used operating systems, it was proposed to design and develop a mobile application for the Android platform with which blind or with a high degree of visual disability they can distinguish the nomination of the different Argentine bills. Our goals were to:

- Increase the autonomy of people with visual disabilities, facilitating the development of daily life and encouraging their integration into society, through of the use of an application that provides independence in money management in cash, without requiring the help of a sighted person.
- Provide a simple and accessible interface that allows improving usability and user experience in the use of smartphones. For this, the application should have a very simple interface that will not present any type of difficulty to the user that wants to use it and it will not get in the way of the task they are doing.
- Provide availability at all times, without the need to require the use of data or Internet connection, so that it can be used anywhere and at any moment.
- Develop an open source and accessible application so that it can be updated to recognize new bills and freely adapted to new needs.
- Provide a free solution to the problems raised, available for download for anyone.

Given the proposed goal, we decided to implement "What's in my wallet?", a prototype system that provides assistance to people with visual disabilities for the recognition of money. We also add the recognition of credit cards and other cards that are commonly used locally.

Based on the tests carried out with the application listed in the "Previous Work" section, it was concluded that most of the applications make a abusive use of voice synthesizers, repeatedly notifying the user information that already knows. In addition, many interfaces, although they have been designed for people with visual impairment, carries graphic elements that hinder its usability, since they require, for example, that the user interacts in a specific place on the screen. Another aspect to consider was that many applications provide instructions to the user every time he executes them, without taking into account that after several uses the person has already learned to use it.

Our software capture frames periodically using the device's camera. When a bill or card enters the camera viewport, the application quickly processes the captured image and determines the denomination, communicating it by voice and in real time. The application provides a very simple and easy to use interface. When the user uses it for the first time, he/she listens to the instructions and then the interaction will be done only by means of tapping anywhere on the screen. All the information that is transmitted to the user is by sound means.

A key feature and one of the main contributions of our work is how the bills and cards are recognized. For this matter, we used an Augmented Reality (AR) framework. Our application does not have any AR feature but a very important aspect of an AR application is how they detect markers. An AR marker is a distinctive pattern that can be recognized by using image processing techniques and it' s used as a frame of reference to place a virtual object. This means

that AR frameworks include very high-end efficient image recognition systems. In these frameworks, the AR markers are preloaded and the framework can quickly determine if it found a marking and which it is. This can be used to our advantage; in our application, the markers are the bills and cards (Figure 1), which are preloaded into AR framework. When "What's in my wallet?" is running the AR framework constantly tries to detect a marker, and when it successfully does it tells which marker is it. We take this response and transform it into sound. The flashlight of the phone, if available, is used automatically by the application. If the application detects that the light conditions are not appropriate then the flashlight is turn on with the involvement of the user.



Figura 1. Characteristic points found by the Vuforia algorithm to identify the bill of $50.

The definition, design and implementation of the application interactions with the user were done under a User Centred Design strategy ([19]) and with consideration of the work done by Batusek & Kopecek in [20]. Based on them, the following decisions were made. When the user runs the application, he can choose if he wants to listen to the instructions or directly going towards the desired functionality. The user's interaction with the system, through touches, can be done in anywhere on the screen. In each of the screens, the system will only provide information referring to the state of the application, in order to avoid repeated and annoying voice notifications to the user. The main functionality of the application, which consists of scanning and recognition of Argentine money bills and some other objects, will be done automatically without the need for the user to interact through a touch on the screen. The application will have the automatic focus of the camera set so that it does not require the user intervention. The application will handle the flash automatically so that the user does not worry about the luminosity of the environment. The system will communicate to the user only the relevant information regarding the state of the application. The result of the recognition of an object will be provided briefly and accurately. The application was implemented for Android using Android Studio. The source project is a 97MB rar file (available here `https://goo.gl/2qUZP7`) and the appli-

cation is a 7.97MB file (available `https://goo.gl/36VTM5`). There is also a user manual available, but only in spanish for the moment (`https://goo.gl/hBaeAq`) "What's in my wallet?" is available in Google Play (Figure 2) for free. It has been downloaded more that a thousand times with a review score of 4.7.

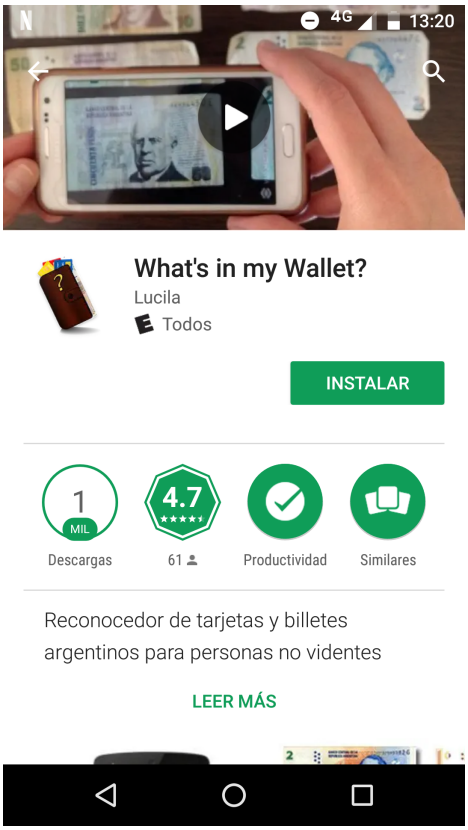


Figura 2. "What's in my wallet?" is available in Google Play for free. It has been downloaded more that a thousand times with a review score of 4.7.

4. Validation

A statistical analysis was performed that allowed evaluating the behavior of the prototype in different conditions. In the first part, a group of bills in circulation under different degrees of illumination was evaluated. The hit rate for the implemented system gave a result of 97.5 % for a high or medium environmental lighting, and 97.91 % without ambient lighting. Through this result, it is possible to affirm that scanning is successful in most cases. In a second part, the bills were

exposed to different conditions that altered the capture of the image, obtaining satisfactory results in most cases. This allowed corroborating that the application still works when the conditions in which the scanning is performed are not optimal. In addition, it should be noted that no false positives were obtained in the analysis, which is very important in an application of this style that is used in financial transactions. Several usability evaluations were conducted and the user managed to successfully complete all the proposed tasks. However, some of them required extra help. For this reason, it is recommended that the first time the application be used, it should be done in the company of a sighted person who explains the instructions to the visually impaired user.

5. Conclusions & Future Work

In the present work, we presented the design and development of a mobile application specifically designed for users with visual disabilities. This work arose from a real need raised by a blind person and that is part of a common problem of people with visual impairment residing in our country. As explained earlier in the article this problem consists in the difficulty which represents for a person with visual impairment the recognition of money bills in Argentina. During the development of this work new ideas have emerged, some of which have been incorporated as additional functionality, and others that have been left open and are expected to be implemented later. The following are the lines of work that are proposed to be followed in the future: Expand the database of the application so that it recognizes the currency of various countries. Also add the currency conversion option, through which a person could scan a bill from any country to obtain the equivalent value in the desired currency and vice versa. In this way the number and diversity of users who would use the application would be expanded, since, if a person with visual disability travel abroad, or if a tourist with vision problems visit our country, they could use the application to recognize the value of money and perform business transactions autonomously. We would like to allow the user to customize the database by adding images so that later they are recognized. It would be very important to test the application with people who have different degrees and types of visual impairment and based on this adapt and improve the interface so that it can be used by a greater number of users. Finally, it can be stated that as an outcome of this particular research work a totally positive experience was obtained. The conducted research work is an example of how software systems can be used to overcome daily life difficulties of people with visual disabilities.

Acknowledgment

This work was partially supported by the following research projects: PGI 24/N037 and PGI 24/ZN29 from the Secretaría General de Ciencia y Tecnología, Universidad Nacional del Sur, Argentina.

Referencias

1. Rogers, Y., Sharp, H., Preece, J. (2011). Interaction design: beyond human-computer interaction. John Wiley Sons.
2. Leasher, J. L., Lansingh, V., Flaxman, S. R., Jonas, J. B., Keeffe, J., Naidoo, K., Wong, T. Y. (2014). Prevalence and causes of vision loss in Latin America and the Caribbean: 1990–2010. British Journal of Ophthalmology, bjophthalmol-2013.
3. Blindness and visual impairment. Facs by the WHO. http://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment
4. 50000 blind persons in Argentina. https://www.lanacion.com.ar/439243-hay-50000-ciegos-en-el-pais-por-falta-de-cirugia-de-cataratas
5. Top 25 countries with most mobile phone usehttps://www.infobae.com/2015/05/21/1730171-la-argentina-los-25-paises-mayor-cantidad-usuarios-smartphones/
6. Mochales, S. (1998). Blind on the internet: Visual impairment and access to information. https://www.sidar.org/publica/press/recortes/25ciegos.htm
7. Zúñiga, S. P. A., Martínez, V. G., Sandoval, M. J. I. (2014). Tiflotecnología y educación a distancia: propuesta para apoyar la inclusión de estudiantes universitarios con discapacidad visual en asignaturas en línea. Apertura, 6(1), 32-45.
8. Pathrabe, A. M. T., Bawane, B. D. N. (2010). Paper currency recognition system using characteristics extraction and Negatively Correlated NN Ensemble. International Journal of Latest Trends in Computing (E-ISSN: 2045-5364), 1(2), 121-124.
9. Yaseri, A., Anisheh, S. M. (2013). A novel paper currency recognition using Fourier mellin transform, hidden Markov model and support vector machine. International Journal of Computer Applications, 61(7).
10. Ali, Ahmed Manzoor, Mirfa. (2013). Recognition System for Pakistani Paper Currency. Research Journal of Applied Sciences, Engineering and Technology. 6.
11. Daraee, F., Mozaffari, S. (2010, October). Eroded money notes recognition using wavelet transform. In Machine Vision and Image Processing (MVIP), 2010 6th Iranian (pp. 1-5). IEEE.
12. Wirelessrerc.org. (2017). 2013 Research Brief (Number 01): Technology Use by People with Hearing and Speech Loss for Communicating with Emergency Response Services | Wireless Inclusive RERC| Georgia Institute of Technology | Atlanta, GA. http://www.wirelessrerc.org/node/160
13. LookTel Products. LookTel Money Reader for iPhone, iPod Touch and Mac. http://www.looktel.com/moneyreader
14. Viaopta-apps.com. ViaOpta Daily. https://viaopta-apps.com/ViaOptaDaily.html
15. A money bill reader for people with visual disabilities in full development. https://www.inti.gob.ar/noticiero/2015/noticiero465.htm
16. PROCER Inclusive Technology http://procertecnologias.com/
17. They create the first App for the blind that recognizes bills and other objects. http://www.redusers.com/noticias/crean-la-primera-app-para-ciegos-que-reconoce-billetes-argentinos-y-otros-objetos/
18. Blind request changes in the making of tickets http://www.lavoz.com.ar/ciudadanos/ciegos-solicitan-cambios-confeccion-billetes
19. COlEmaN, R., ClaRkSON, J. O. H. N., Cassim, J. (2016). Design for inclusivity: A practical guide to accessible, innovative and user-centred design. CRC Press.
20. Batusek, R., Kopecek, I. (1999). User Interfaces for Visually Impaired People. In Proceedings of the 5th ERCIM Workshop on'User Interfaces for All (No. 21, p. 6).

SIP (Sistema de Identificación de Personal), una solución informática al servicio de la seguridad de nuestros adultos mayores

Augusto José Markic (augusmarkic@gmail.com), Leandro Germán Herrera (germanhe@gmail.com), Rodolfo Bertone (rodolfo.bertone@ucalp.edu.ar)[1]

[1] Facultad de Ciencias Exactas e Ingeniería, Universidad Católica de La Plata

Resumen. Los adultos mayores constituyen uno de los grupos más vulnerables en cuanto a la posibilidad de verse afectados en hechos de inseguridad en sus hogares. Una de las modalidades utilizadas por delincuentes, es la de hacerse pasar por representantes o empleados de empresas prestadoras de servicios (luz, gas, teléfono, etc.) y, por ejemplo, con la excusa de tener que realizar un arreglo interno en el domicilio, ingresan al mismo y cometen un acto delictivo [1][2].

Ante esta problemática, surge como una posible solución el Sistema de Identificación de Personal (SIP). Este sistema tiene por finalidad permitir a la persona que se encuentre en el domicilio constatar de forma simple, rápida y a bajo costo, mediante una solución informática, la veracidad del envío de personal por parte de las empresas prestadoras de servicios, con el fin saber de qué forma actuar con la persona que está, "del otro lado de la puerta".

Keywords: Seguridad, suplantación de identidad, minicomputadora, SIP.

1 Introducción

Según el Instituto Nacional de Estadística y Censos "INDEC", al año 2010 en Argentina, la población mayor a 65 años equivale al 10,4% del total [3]. Este porcentaje es significativamente mayor en las grandes ciudades, y por ejemplo asciende al 16,4% en la Ciudad autónoma de Buenos Aires.

Además, según el boletín de la Universidad Católica Argentina "Percepciones Psicosociales: Capacidades Psicológicas Y Aspectos De La Inseguridad", "…*El problema de la inseguridad se puede analizar desde dos puntos de vista, el de la cantidad o número de delitos concretos registrados o desde la percepción o sentimiento de inseguridad que experimentan las personas*…" [4].

Conforme a esto, se reporta que, durante el año 2008 en Argentina, el 18% de la población mayor de 60 años sufrió un hecho de delincuencia, mientras que, en ese mismo año el 71% del total de la población mayor de 60 años se manifestó con miedo al delito.

El boletín, también presenta las consecuencias de la inseguridad cuando dice: "… El haber sufrido un hecho de delincuencia influye de manera decisiva en la vida social de las personas constituyendo un problema grave para la vida colectiva, en tanto que estimula el aislamiento y la desconfianza en el lazo social…".

El asalto por suplantación de identidad en los domicilios es uno de los hechos más habituales en lo que se refiere a denuncias policiales ante robos o hurtos. Ante estos escenarios se ha considerado relevante el desarrollo de soluciones tecnológicas que traten de evitar los delitos que utilizan esta modalidad.

1.1 Robo o Suplantación de Identidad en Internet.

La suplantación de identidad puede definirse como el acto deliberado de una persona para hacerse pasar por otra ante un tercero. Es un hecho conocido que en internet esta actividad actualmente resulta muy común; lamentablemente la forma más frecuente de suplantación es la apertura o registro de un perfil que en muchos de los casos se da sólo utilizando la foto, sin que en él se viertan datos personales, y donde la única solución posible para el suplantado es comunicarse con el portal web o red social, para que sean sus administradores quienes eliminen el perfil falso, porque no se considera un delito si no hay de por medio datos sensibles.

Existe otro rubro de la suplantación de identidad que va más allá de robar el nombre y foto para abrir una cuenta en cualquier red social, y es aquel acto que tiene como finalidad el fraude; es decir el "Phishing". Esta modalidad es considerada como una estafa, mediante la cual se obtienen datos, claves, números de cuentas bancarias, tarjetas de crédito u otros datos para ser usados de forma fraudulenta.

El caso en estudio no está orientado en a los problemas ante descriptos, sino en proveer una herramienta que permita detectar delincuentes que se hacen pasar por el servicio técnico de una compañía cualquiera (telefónica, luz, cable, etc.). La modalidad no es nueva; pero se sofisticó. Ahora las bandas que se dedican a robar casas, disfrazados de operarios telefónicos, cortan las líneas, interceptan llamados y se aparecen en domicilios que efectivamente pidieron el servicio técnico, con lo cual el grado aumenta la confianza de las personas, sobre todo mayores que los atienden y resultan asaltadas.

Algunas empresas de servicios recomendaron dos acciones por parte de los usuarios para verificar la identidad de las personas que se presentan a hacer el servicio técnico. Una de ella es pedir la credencial de identificación del empleado. Pero, más allá de la credencial, que podría llegar a ser adulterada, siempre se les puede preguntar el código de reclamo, este código sólo lo conoce el cliente que hizo la gestión y lo debe llevar el técnico al momento de la visita al domicilio. Esta posibilidad suele ser desconocida por los clientes mayores.

En el caso estudiado, se presenta una nueva alternativa de solución para brindarle al cliente una mayor seguridad respecto de la identificación del servicio técnico, a partir del uso de la tecnología que resulta ser de bajo costo para empresa prestadora del servicio.

2. SIP (Sistema de Identificación Personal)

A partir de la problemática presentada, se diseño e implementó una solución tecnológica/informática que aborda el problema. Como parte del trabajo de un grupo de alumnos de la Ingeniería en Sistemas de la Universidad Católica de La Plata se diseñó e implementó el proyecto SIP que brinda una solución original, económica y sencilla al problema de identificación del personal. El mismo consiste en el desarrollo de un prototipo funcional el cual ofrece una solución al problema de inseguridad.

SIP se diseño como un sistema que involucra a 3 usuarios diferentes:

- Empresa Prestadora de servicios: sin importar el tipo de servicio prestado, la empresa recibe el pedido o reclamos por parte del cliente y debe satisfacer la inquietud en un día y hora que normalmente se pacta con el cliente.
- Empleado de la empresa: en este caso, el empleado en cuestión es el encargado de realizar el servicio técnico solicitado. Es un empleado de la Empresa Prestadora de Servicios y debe concurrir al domicilio del cliente a fin de atender el reclamo
- Cliente de la empresa: el último componente del circuito, es quien, ante un problema en su domicilio, se comunica con la empresa prestadora del servicio, y quien recibe al empleado que debe realizar la reparación.

SIP prevé entonces tres roles diferentes para interactuar con el sistema. Esta diferencia se observa en varios aspectos: tipo de equipamiento utilizado, interfaces desarrolladas y nivel de conocimiento y familiaridad de uso de herramientas tecnológicas.

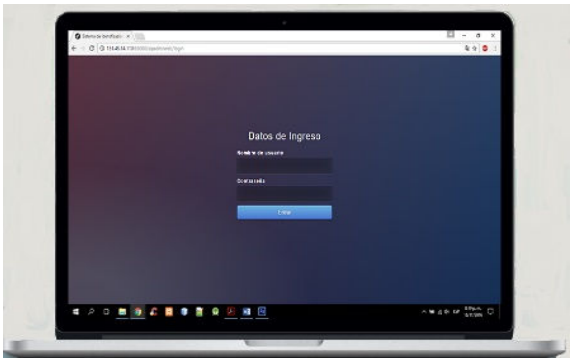


Figura 1. Página web para la empresa prestadora de servicios.

2.1. Desarrollos propuestos y realizados

Para la empresa prestadora de servicios, se diseño e implementó un sitio web para la administración del sistema. La figura 1 presenta el ejemplo de la interfaz de conexión al sistema. La misma fue desarrollado con el framework php “symfony”, el cual se ejecuta sobre un servidor apache y utiliza una base de datos MYSQL. El sitio

está diseñado para que lo utilice un usuario con una mínima capacitación. Para ello se puso especial énfasis en la usabilidad, tratando de adoptar las mejores prácticas en ese sentido.

Para el caso del empleado de la empresa, la solución implementada utiliza una aplicación móvil. En el caso del prototipo desarrollado, esta pensado para ejecutarse sobre dispositivos con SO Android, ya que en la actualidad es el sistema operativo más difundido en la región. De esta forma facilita la implementación y adaptación de la misma para su utilización. La figura 2 presenta una fotografía de la aplicación SIP ejecutándose en un teléfono celular.



Figura 2. Aplicación móvil para el empleado de la empresa**.**

Esta aplicación interactúa con el servidor de SIP a través del uso de web services (también de desarrollo a medida para la solución), los cuales fueron codificados utilizando el lenguaje php y el protocolo de acceso SOAP. El uso de servicios web, permitiría a futuro añadir aplicaciones para otros sistemas operativos y/o nuevas tecnologías.



Figura 3. Minicomputadora, instalada en el domicilio del cliente.

Por último, para el hardware a instalar en el domicilio del cliente , se seleccionó una Raspberry Pi 3 . Este dispositivo funciona como un minicomputador, y posee una

pantalla táctil, un sistema operativo basado en Linux, una aplicación precargada desarrollada en JAVA (la cual se encarga de interactuar con el servidor SIP utilizando web services) y un sistema de alimentación ininterrumpida (UPS). Los objetivos planteados con este complemento fueron: que sea fácil de utilizar por parte del cliente, que sea simple de conectar, que no dependiera exclusivamente de tener un servicio de electricidad, que fuera económico y que el tamaño del mismo permitiera una instalación simple en algún lugar de fácil acceso, preferentemente cercano a la puerta de ingreso del domicilio. La figura 3. presenta una fotografía del prototipo diseñado.

3. Operación de SIP. Circuito de validación de una visita

El circuito de validación de una visita por parte del servicio técnico se define como una secuencia de pasos o actividades que se describen a continuación. La Figura 4 presenta de manera esquemática el proceso.

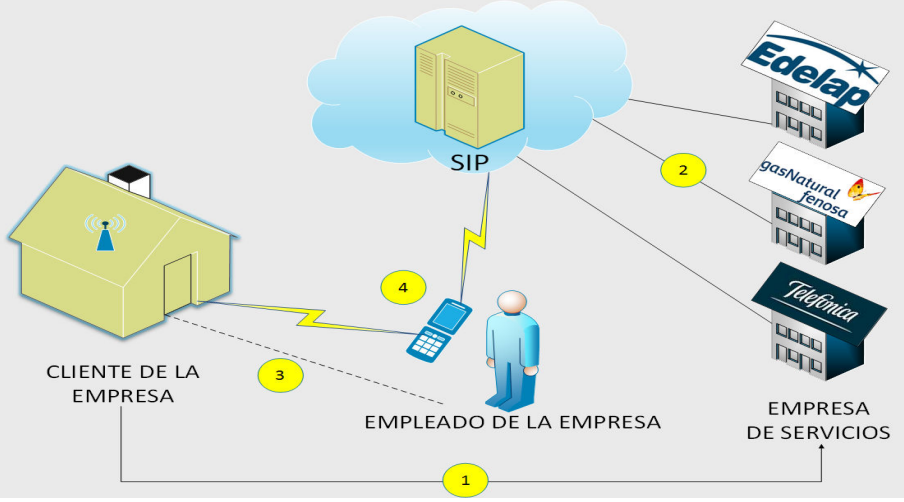


Figura 4. Interacción entre los diferentes componentes de la solución.

- El cliente realiza un reclamo por la falla de servicio. Esta operación se realiza mediante los canales habilitados y difundidos oportunamente por la empresa prestataria. Aquí SIP no actúa.
- La empresa genera y pauta con el cliente, utilizando la aplicación WEB SIP, una visita al domicilio del cliente afectado. Para ello asigna un empleado, especificando: fecha y rango horario de la misma.
- El empleado asignado por la empresa se dirige al domicilio del cliente que realizó el reclamo y notifica a SIP (mediante la aplicación de su celular), que ya se encuentra en la puerta del domicilio.
- El empleado de la empresa a través de la aplicación de su celular, oprime un botón para brindarle conexión a internet a la minicomputadora del cliente de

manera inalámbrica (de esta forma no es necesario que el cliente disponga de internet en su domicilio) y luego anuncia su visita (toca el timbre)

- El cliente ante el "llamado a la puerta" donde la persona se anuncia como personal de cierta empresa, oprime un botón con la leyenda "Verificar Visita" en la pantalla de su minicomputadora. La aplicación automáticamente consulta a SIP si la visita que está en la puerta de su domicilio es válida. En ese caso, se presenta en la pantalla la foto del empleado y sus datos personales. En caso de no ser válida la misma, se recomendará al cliente ignorar la visita.

4. Infraestructura de seguridad de SIP

La seguridad integral del sistema es parte vital en el desarrollo de esta aplicación. Si las aplicaciones son vulnerables, la solución que se pretende brindar al cliente tendrá debilidades que harán dudar de su utilización.

Debido a que Internet es una red pública, compartida y no controlada, y considerando también que la minicomputadora instalada en el domicilio del cliente puede realizar la conexión a través del dispositivo del empleado que se pretende autentificar, se consideraron las siguientes medidas de seguridad:

- Página Web (en la empresa): Acceso a la página de administración mediante autentificación de usuario y contraseña, certificado SSL propio otorgado para tal fin (factor adicional de seguridad).

- Aplicación Móvil (disponible para el empleado de servicio técnico):
  - Validación de dispositivo mediante IMEI (código de identidad internacional de equipo móvil),
  - Autentificación mediante usuario y contraseña,
  - Comunicación hacia los servidores sobre TLS a través de web services, con certificados X.509 propios (firmado por SIP), que cumple a su vez la función de validar al dispositivo y al empleado ya que está emitido a su nombre.

- Aplicación de la Minicomputadora (cliente):
  - Validación de dispositivo, mediante Dirección MAC de la interfaz utilizada para realizar la conexión, esta aplicación la recupera y la envía en cada llamado.
  - Comunicación hacia los servidores a través de web services utilizando TLS, con certificados X.509 propios (firmado por SIP), que cumple a su vez, la función de autenticar al dispositivo del domicilio.

5. Resultados Obtenidos

A partir de las especificaciones definidas, se desarrolló íntegramente un prototipo funcional en el que cada parte fue desarrollada y probada independientemente hasta que se logró su adecuado funcionamiento. La figura 5 muestra el prototipo funcional de la UPS desarrollada como fuente de energía alternativa para la microcomputadora (Raspberry) a instalar en el domicilio del cliente.

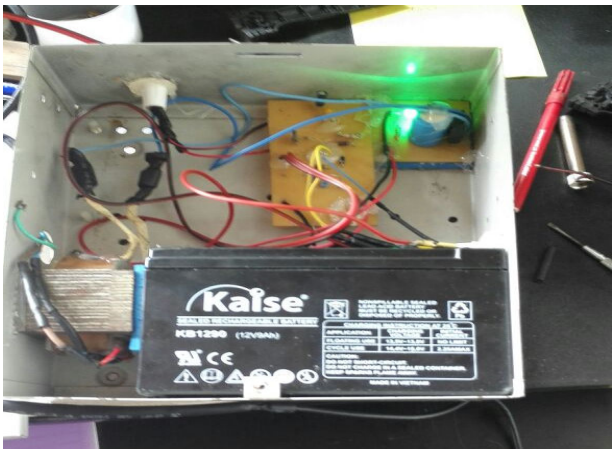


Figura 5. UPS del prototipo funcional**.**

La figura 6 presenta una fotografía de la microcomputadora instala en el domicilio del cliente luego de validar al operario que concurre a realizar el servicio técnico. Posteriormente la figura 7 presenta diferentes interfases del sistema.



Figura 6. Validación de una visita aceptada en el prototipo funcional**.**



Figura 7. Interfaz de la aplicación de la minicomputadora**.**

Por ultimo, la figura 8 describe tres ejemplos de interfaz de la aplicación móvil desarrollada.

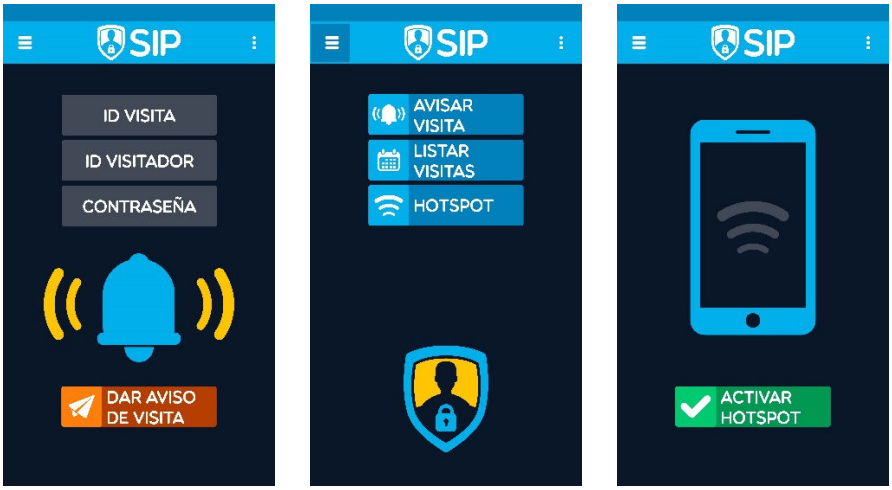


Figura 8. Interfaz de la aplicación mobile.

6. Conclusiones y Trabajos Futuros

De acuerdo al trabajo realizado, es posible destacar características que constituyen una serie de beneficios que se pueden obtener a partir del prototipo generado:

- Facilidad de uso: considerando que el cliente al cual está dirigido principalmente es un adulto mayor, que podría no estar habituado al uso de la tecnología.
- No suma gastos considerables. El costo de la microcomputadora que se instala en el domicilio del cliente es bajo.
- No depende de otros servicios (por ejemplo: electricidad, internet, telefonía).
- Posee un gran potencial, es altamente escalable en funcionalidades.
- La solución de software puede operar bajo distintas plataformas, en especial para la aplicación móvil.
- La cantidad de datos necesaria para realizar la validación de una visita, es lo suficientemente baja como para obtener tiempos de respuestas menores a 5 segundos en una ciudad con cobertura de telefonía celular normal.

Como trabajos futuros pueden plantearse los siguientes desafíos:

- Los adultos mayores tienen mayores posibilidades de desarrollar algún tipo de discapacidad motora, visual, auditiva. La solución puede mejorarse para contemplar las adaptaciones y la incorporación de funcionalidades necesarias.
- Si el cliente de la empresa posee una conexión a internet en su domicilio, podrían añadirse diferentes funcionalidades que hagan uso de esta conexión (botón antipánico, video vigilancia, sistemas de scoring sobre la calidad de la visita obtenida, etc.).

7. Bibliografía y Referencias:

1. http://tn.com.ar/policiales/violento-robo-a-una-anciana-en-villa-devoto-le-hicieron-el-cuento-del-tio-y-le-robaron-los-ahorros_625168
2. http://www.eldia.com/nota/2017-2-28-3-47-33-pidio-un-medico-a-su-prepaga-y-a-la-casa-fue-un-ladron
3. http://www.indec.gov.ar//ftp//nuevaweb/cuadros/7/sesd_11c52.xls
4. http://www.uca.edu.ar/uca/common/grupo68/files/Boletin_3_PERCEPCIONES_PSICOSOCIALES_CAPACIDADES_PSICOLGICAS_Y_ASPECTOS_DE_LA_INSEGURIDAD.pdf
5. Redes de Computadoras 4ta edición - Andrew S. Tanenbaum
6. Fundamentos de la Dirección de Proyectos - G. Guido Lavalle, J. Gadze, R. Wehbe
7. Raspberry Pi Wireless Networks - Agus Kurniawan
8. Thinking in Java (4th Edition) - Bruce Eckel
9. Learning Android - Marko Gargenta

Interoperabilidad e Integración entre Sistemas de Información en Salud

Marcela Demichelis, Martin Correa, Rodolfo Bertone

Facultad de Ciencias Exactas e Ingeniería
Universidad Católica de La Plata
{marcela.demichelis, martinsebastian.correa} @ucalpvirtual.edu.ar
rodolfo.bertone@ucalp.edu.ar

Resumen. En el ámbito asociado a salud, se observa que existe un alto grado de intercambio de información entre los diferentes actores involucrados. Los datos se encuentran altamente fragmentados y distribuidos en múltiples sistemas no integrados, imposibilitando la comunicación y el intercambio de información entre ellos.
Este trabajo se analiza, investiga, describe y documenta los distintos niveles de interoperabilidad de datos de salud (IO), mencionando sus beneficios y barreras de implementación.
Un sistema es interoperable gracias a su capacidad de transferir información de un paciente de un sistema a otro. Esta acción es imposible de lograr sin el uso de estándares que permitan interoperar efectivamente. El estudio realizado se investiga el estado de implementación, clasificación y mecanismos para su desarrollo en Argentina, tanto para el sector público y privado.
Con el aporte de Las Tecnologías de la Información y la Comunicación (TIC) se podrá mejorar la calidad de vida y bienestar de las personas, la equidad en la atención de los pacientes, optimizar la relación coste-beneficio, contribuir a la seguridad, incrementar la eficacia y la eficiencia de los prestadores o de la organización. Para aprovechar su potencial que ofrecen, es fundamental acordar las normas técnicas, semánticas y sintácticas de interoperabilidad.

Palabras Clave: intercambio de datos, sistema de salud.

1.Introducción

En un mundo interconectado la meta debería ser que todos se puedan comunicar con todo el mundo sin ningún impedimento y que los datos, el activo más valioso, se mantenga en un formato que permita acceder a ellos en el tiempo, e intercambiarlos sin problemas con los diferentes receptores en la red.

El verdadero éxito de internet se centra en el hecho de que se basa en estándares abiertos y se puede acceder a la red con independencia del sistema operativo, o de la tecnología utilizada. Sin importar el Sistema Operativo o hasta la arquitectura de computadora que se utilice, se puede acceder a Internet sin inconvenientes. Esto se

logra solamente utilizando estándares abiertos de comunicaciones como TCP/IP,HTML, POP, SMTP, etc.

Con independencia de la tecnología que se decida a usar, hay que tener en cuenta que la opción tecnológica puede llevar a situaciones complejas, si no se toman ciertas precauciones. Puede llegar a ocurrir qué sin políticas adecuadas, los propios datos generados por sistemas anteriores sean inaccesibles para otras personas, o incluso para la misma organización que lo genera luego de transcurrido cierto tiempo.

En muchas ocasiones se confunde el concepto de Interoperabilidad de datos (IO) con el de neutralidad tecnológica. Esta confusión puede deberse a que la mayor parte de las estrategias necesarias para conseguir la neutralidad tecnológica (ejemplo el uso de estándares abiertos) son idénticas a las que se deben usar para conseguir la IO.

Una de las mejores definiciones para el concepto de neutralidad tecnológica es la condición en la que una acción, definición, ley, estándar, etc., no se formula de tal manera que sesgue todas las decisiones subsecuentes a favor de una tecnología en particular, entre aquellas capaces de resolver el problema [1].

2.IO en el ámbito de la salud

En el ámbito de salud, es crítico fortalecer las cinco fuentes de datos clave: 1) encuestas de salud, 2) registros de nacimiento y muerte, 3) censos, 4) sistemas de información y vigilancia, 5) sistemas administrativos y su disponibilidad. Se debe dar un uso estratégico e integrado a las herramientas para la práctica del cuidado de salud.

Para aprovechar el potencial que ofrecen las TIC, es fundamental acordar las normas técnicas, semánticas y sintácticas de IO de los sistemas. Los sistemas de información de salud deben cumplir con las normas en todos los niveles, incluidos los sistemas utilizados para captar los datos de los pacientes en el lugar de atención.

Se debe lograr un consenso con la terminología en común utilizada y los conjuntos mínimos de datos para que la información pueda recopilarse sistemáticamente y compartirse fácilmente, sin ser malinterpretada.

Además, las políticas nacionales deben de procurar brindar una adecuada protección de los datos, privacidad y contar con el consentimiento del paciente para el intercambio de datos en salud.

En el ámbito de la salud, IO posee un alcance más específico. Es definida como la capacidad de diferentes sistemas de información en salud (sistemas hospitalarios, departamentales, registros clínicos electrónicos, etc.) para intercambiar datos y utilizar la información que ha sido intercambiada dentro y a través de los límites de la organización. La finalizad principal es mejorar la prestación efectiva de los cuidados de salud a individuos y comunidades.

La IO aumenta la seguridad del paciente al permitir el acceso y disponibilidad a los datos clínicos, accediendo a sus datos en tiempo real. Permite atender a pacientes desde cualquier punto, mejorando la calidad y continuidad asistencial con independencia del centro donde acuda.

Es imperativo, entonces, que los diferentes sistemas de salud puedan intercambiar información y transferirla de un sistema a otro a través de interfaces específicas, adaptadas o personalizadas. Se debe garantizar la identificación del paciente de

manera unívoca, para intercambiar la información de forma confiable, garantizando además la seguridad técnica, jurídica y de procedimiento. Todo ello mejora la calidad, gestión y seguimiento del paciente desde cualquier punto de la red de salud [2].

2.1. Niveles de Interoperabilidad

Para clasificar los niveles de interoperabilidad en salud se utiliza la clasificación propuesta por la European Telecommunication Standards Institute (ETSI).La figura 1 presenta la clasificación que de define a continuación [3].

1. Técnica: se asocia generalmente con componentes de hardware y/o software, sistemas y plataformas que permiten la comunicación que tendrá lugar de máquina a máquina. Este tipo de interoperabilidad a menudo define protocolos de comunicación y la infraestructura necesaria de los mismos para operar.
2. Sintáctica: relacionada habitualmente con los formatos de datos. Los mensajes transferidos por protocolos de comunicación necesitan una sintaxis y codificación bien definida. Sin embargo, muchos protocolos transportan datos o contenidos, y esto se puede representar mediante sintaxis de transferencia de alto nivel, tales como HTML, XML o ASN 12 (intercambio de información entre sistemas que no garantiza la correcta interpretación y uso, se resuelven temas técnicos y de estándares en las áreas de protocolos y formatos de intercambio).
3. Semántica: nivel de IO que se asocia con el significado de los contenidos y se refiere a la interpretación humana del contenido. IO en este nivel significa que hay un entendimiento común entre personas sobre el significado del contenido (información) que se intercambia (se garantiza la correcta interpretación y uso de la información intercambiada, para lo cual se necesitan definiciones formales de cada entidad, atributo, relación, restricción y término intercambiado).
4. Organizacional: es la capacidad de las organizaciones para comunicar y transferir efectivamente los datos (información), sin importar los sistemas utilizados, o eventualmente diferentes infraestructuras, a través de diferentes regiones geográficas y culturas. La IO organizacional depende del éxito de los niveles técnico, sintáctico y semántico.

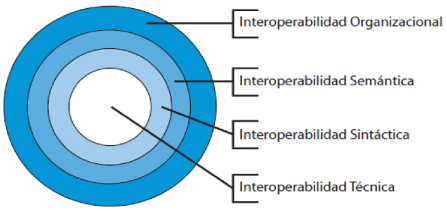


Figura 1. Niveles de Interoperabilidad

2.2 Estándares en salud

Un estándar es un documento aprobado por consenso por un organismo reconocido, que proporciona pautas, reglas y/o características para un uso común, con el objeto de

obtener un óptimo nivel de resultado en un contexto dado. En los sistemas de información en salud, el primer paso para que un sistema sea interoperable es su capacidad de transferir información de un paciente de un sistema a otro. En general, esta transferencia se realiza a través de una interfaz adaptada y personalizada [4].

2.3 TIC en salud

Los principales beneficios de la implementación de las TIC en el sector sanitario son cuatro:

1. Incremento de la calidad en la atención al paciente y de la eficiencia: Una de las carencias más importantes del sector es la fragmentación de la atención sanitaria y las dificultades de transmitir eficazmente la información. Las TIC pueden ayudar a mejorar la seguridad del paciente mediante el acceso directo al historial médico, la consulta online de los tratamientos, registrando la evolución de los enfermos y previendo posibles errores médicos. El estudio refleja que estas herramientas son percibidas entre profesionales y usuarios como sustancialmente positivas para mejorarla seguridad del paciente, por lo que los países promueven su implantación.
2. Reducción de los costes operativos de los servicios médicos: Las TIC ayudan a disminuir este tipo de costes con la reducción del tiempo requerido para procesar datos o manejar documentación. El sistema de archivo y transmisión de imágenes se presenta como indispensable para el desarrollo de la historia clínica electrónica y la telemedicina, ya que reduce los tiempos de los análisis y los resultados.
3. Disminución de los costes administrativos: La facturación presenta grandes posibilidades de ahorro gracias al uso de las TIC.
4. Posibilidad de llevar a cabo modelos de sanidad completamente nuevos: Las TIC han sido definidas como tecnologías con un alto potencial transformador ya que presentan nuevas formas de ejercer la medicina y desarrollar la sanidad [5].

3 Estándares para Interoperabilidad

3.1 HL7

HL7 es una iniciativa que comenzó en 1987, en base a la necesidad de normalizar las interfaces entre los múltiples sistemas heterogéneos de información, y rápidamente se convirtió en el estándar de facto para el intercambio electrónico de datos clínicos y administrativos en los servicios de salud de los Estados Unidos. La estructura internacional de la organización, el procedimiento de votación balanceado, y las políticas abiertas de asociación, aseguran que todos los requerimientos sean tenidos en cuenta uniformemente y equitativamente con calidad y consistencia [6].

HL7 es una organización que desarrolla estándares para minimizar las incompatibilidades entre sistemas de información en salud, permitiendo la interacción

y el intercambio productivo de datos entre aplicaciones heterogéneas, independientemente de su plataforma tecnológica o de su lenguaje de desarrollo. Diferentes sectores participan en forma colaborativa en la discusión y en el desarrollo de estándares por consenso, en un entorno abierto. Estos sectores pueden ser prestadores de servicios de salud, desarrolladores de software, consultores, usuarios finales, organizaciones gubernamentales y no gubernamentales, etc.

La figura 2 presenta el denominado Documento Clínico ARC (DCA), uno de los estándares HL7 más importantes. Provee un lenguaje de marcado, para la creación de documentos clínicos electrónicos, donde la presentación se realiza mediante lenguaje HTML, soportando lenguaje XML. [4].

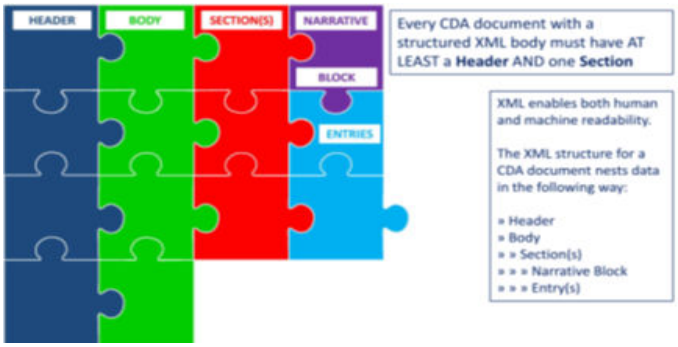


Figura 2. Documento Clínico (DCA)

3.2 Estándar FHIR

Es un marco de estándares de última generación creado por HL7. FHIR combina las mejores características de las líneas de producto HL7 versión 2, versión 3 y CDA.

La figura 3 presenta la estructura FHIR. Los artefactos de IO de FHIR se generan a partir de un conjunto de componentes modulares, cuyo contenido ha sido diseñado para el intercambio de datos. Estos artefactos, denominados recursos, son representaciones de conceptos empleados en el contexto de la salud (paciente, medicación, observación, etc.), y funcionan como bloques de construcción que permiten componer estructuras de mensajes y/o documentos.

Todos los recursos comparten las siguientes características: Una manera común de definirlos y representarlos, un conjunto común de metadatos y un elemento para facilitar la legibilidad humana. [7].

3.3 Apoyo de SNOMED

SNOMED Internacional es una organización compuesta por 31 países miembros (Argentina es el último en incorporarse durante 2018), que ha conformado un estándar global para la representación de información en salud.

La terminología definida es para generar estándares en sistemas de historia clínica electrónica, de prescripción electrónica, laboratorios, encuestas sanitarias, registros de enfermedades y otros, tiene el objetivo de capturar de manera precisa y unívoca referencias a conceptos clínicos. Incluye más de 330.000 conceptos clínicos, divididos en jerarquías temáticas con síntomas, diagnósticos, antecedentes familiares, procedimientos quirúrgicos, estudios diagnósticos, etc.[8]

Este nivel de detalle permite lograr una atención médica más segura y un sistema de salud que cuenta con información confiable y actualizada para la vigilancia y la toma de decisiones. Su objetivo es facilitar el registro exacto de la información médica y otros temas relacionados con la salud y la IO semántica de los registros clínicos.

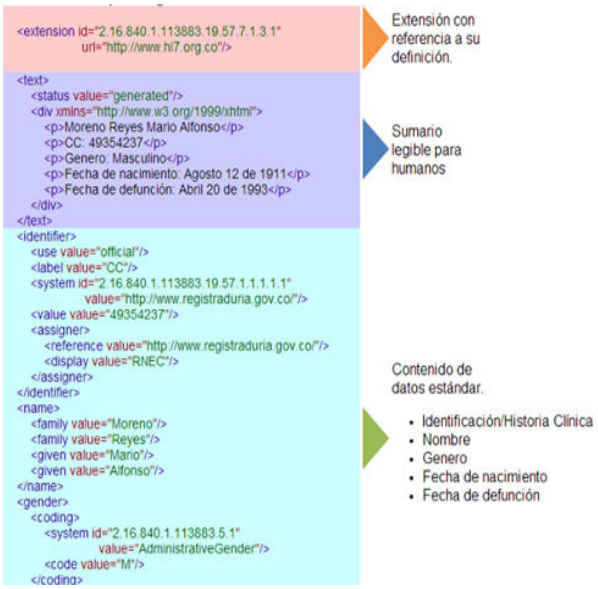


Figura3: Estructura FHEIR

4 Acuerdo Público-Privado (APP)

Actualmente en Argentina no existe regulación para el intercambio estandarizado de información en salud. Esta situación genera algunas situaciones controversiales como:

- Cobertura desigual tanto en el sector público como en el sector privado. [9].
- Fragmentación e ineficiencia en el uso de los recursos.
- Desigualdad en el acceso al sistema de salud de la población por su nivel de ingresos o por la ubicación geográfica del paciente.
- Variabilidad en la calidad y equipamientos desiguales, sin planificación ni concepto de redes sanitarias; ausencia o deficiencia de mantenimiento edilicio y de aparatología.
- Falta de planificación de recursos humanos.

- Falta de regulación de tecnologías.
- No se puede compartir la información de una forma rápida y confiable.
- No se puede controlar la validación, integridad, seguridad y disponibilidad de la información.
- El trabajo de carga de información se multiplica y se vuelve ineficiente.
- Los recursos humanos no se vinculan para optimizar esfuerzos, sino que por el contrario trabajan de forma desintegrada y aislada, repitiendo en muchos casos las mismas tareas.

Bajo la concepción que la salud es un derecho humano fundamental, y la equidad es un aspecto primordial para hacerlo efectivo, la Cobertura Universal de Salud (CUS) implica que todas las personas y las comunidades tengan acceso, sin discriminación alguna, a servicios integrales de salud, adecuados, oportunos y de calidad, así como a medicamentos seguros, eficaces y asequibles, sin que los usuarios se expongan a dificultades financieras, en particular los grupos en situación de vulnerabilidad.

El Estado necesita contar con sistemas interoperables, y que los desarrollos que se hacen en el marco público sean de código abierto, para garantizar no quedar ligados en ningún emprendimiento privado que después generen una dependencia que como Estado no se puede poseer [10].

4.1 Marco Argentino de IO en Salud (MAIS)

El Marco Argentino de IO en Salud (MAIS) es un proyecto para estandarizar el intercambio de documentación clínica, el detalle de la facturación y los débitos entre prestadores y financiadores. Utilizando el estándar HL7, CDA para especificar la estructura y semántica de documentos clínicos.

Persigue el objetivo de intercambiar documentos electrónicos con información en salud entre instituciones con Sistemas de Información diferentes, sin la necesidad de una plataforma de IO central.

La información fluye de los prestadores a los financiadores, como adjunto de la facturación o en forma independiente.

La metodología de trabajo de este Proyecto MAIS consiste en reuniones periódicas, donde se proponen y consensuan las tecnologías a utilizar y las adaptaciones a realizar, vocabularios e identificadores comunes para facilitar la IO. También se decide los tipos de documentos para interoperar, que datos son obligatorios y que datos son optativos [11].

4.2 Historia Clínica Electrónica (HCE)

La HCE es un documento virtual donde se dejan asentados las características de cada paciente y sus respectivos diagnósticos y medicaciones. La ventaja de un formato electrónico es que permite el desarrollo de una red de cuidados integrales y progresivos porque toda la información que se vuelca en la HCE está centralizada e integrada en un mismo lugar.

Su implementación con el soporte de SNOMED rompe la barrera con el inconveniente del idioma, permite el acceso preciso y dirigido a la información relevante, (con disminución de costosas duplicaciones y errores), mejora la calidad de la atención de los pacientes, mejora la calidad de atención, y en conciencia el bienestar de la sociedad. [12].

5 Conclusiones

5. 1 Relación TIC e IO

La tecnología digital permite que un servicio o contenido se puede alcanzar en cualquier momento y desde cualquier sitio. Esto no se puede conseguir si los sistemas (en particular, para el caso de estudio los sistemas de salud) no interoperan, mostrando capacidad de trabajo conjunto. Para que esta interoperabilidad se alcance con éxito, son varios los factores implicados: uso de herramientas, tecnologías y estándares que colaboren a tal fin.

Un buen programa de interoperabilidad es capaz de aprovechar con éxito las inversiones existentes en la Administración Pública y facilitar el medio para que los sistemas existentes puedan participar en un más amplio entorno de sistemas de información, generando eficiencias internas y el despliegue de un mejor acceso online.

Con interoperabilidad, los sistemas de las empresas y Administraciones Públicas son capaces de alcanzar un intercambio de datos mejorado, así como un acceso más integrado a sistemas y aplicaciones. Esta acción, puede lograr una reducción muy en los costos de la salud [13].

5.2 Enfoque de modelado HL7

El enfoque de modelo que presenta la normal HL7 tiene varias ventajas. En primer término, resulta en fácil de implementar a través de un sistema de software. Como segundo aspecto, es estándar de mensajería es el más difundido en la comunicad de la salud. Utiliza elementos de representación de mensajería XML. El esquema de comunicación es simple, basado en eventos y mensajes, lo que simplifica compartir e intercambiar datos. [14].

5.3 SNOMED ayuda a la Interoperabilidad semántica

Para aprovechar el potencial que ofrecen las TIC es fundamental acordar las normas técnicas, semánticas y sintácticas de interoperabilidad para los sistemas.

La IO semántica es la interpretación humana de símbolos, palabras o expresiones, por lo tanto, las interpretaciones diferentes pueden ser infinitas para una misma frase. Por este motivo, se debe acordar una terminología en común y los conjuntos mínimos de datos para que la misma pueda compartirse y no ser malinterpretada. SNOMED

presenta una solución al describir una terminología que brinda apoyo a la interoperabilidad semántica. De esta forma, constituye un instrumento consistente que permite introducir información clínica en los sistemas de forma estandarizada, asociados a códigos únicos, facilitando la representación de datos clínicos entre sistemas heterogéneos. [13].

5.4 Acuerdo Privado – Publico

Actualmente Argentina presenta un sistema de información relacionado con la salud totalmente fragmentado. Está distribuido en múltiples silos de datos no integrados, con grandes volúmenes de información (en logística, vigilancia, registros y reportes) imposibilitando su adecuada disponibilidad. Esto hace muy dificultoso o hasta imposible compartir datos de forma rápida y confiable y de comunicarnos con otros sistemas. Para mejorar esta situación es necesario la integración de la información, apoyándose en las ventajas que provee la IO, en estándares que ayudaran a ganar en eficiencia y seguridad, con el aporte de TIC, y con acuerdos de centros públicos y privados impulsado por el Ministerio de Salud de la Nación. [14].

El sector público trabaja para consolidar un sistema de atención que haga posible satisfacer las necesidades de salud de la población. Para que el mismo sea igualitario es necesario el acceso a la información esté disponible siempre. La forma que esto es posible es a través de la HCE, de esa forma los antecedentes médicos del paciente están disponibles con independencia del establecimiento donde reciba atención.

La estrategia del estado es disminuir brechas de calidad de atención de la salud, implementado sistemas de información que identifican las necesidades de la población, permiten el seguimiento longitudinal e integral del paciente, y proveen herramientas innovadoras a los profesionales de la salud y a los pacientes. Basando en sus principios básicos, lograr la cobertura universal en salud, poder integrar información, implementar tecnología, integrar subsistemas, empoderar al paciente, asegurar privacidad, y desarrollar RRHH. Empoderamiento del paciente incluye al paciente como el dueño de la información clínica, definiendo la arquitectura de un registro clínico personal decidiendo que tipo de datos puede contener y como se puede compartir. Con el apoyo de la IO se podrá contar con herramientas para el registro de información por parte del paciente, que se pueda realizar mediante la importación desde dispositivos, la decisión de compartirlas, y activar recordatorios y alertas.

Es conocido por la mayoría de los habitantes de Argentina, la gran cantidad de dinero que se gasta, o mal gasta, en salud. La tecnología provee los medios para poder integrar y compartir la información. De esta forma se podría brindar una solución al modelo de salud que permita mejorar las prestaciones con la utilización de menos recursos económicos que los actuales. [15].

Referencias

1. https://seguinfo.wordpress.com/2009/05/15/interoperabilidad-i-las-definiciones/
2. https://www.agesic.gub.uy/innovaportal/file/1339/1/renoir_selena_y_antonio.pdf

3. http://iris.paho.org/xmlui/bitstream/handle/123456789/28188/9789275318812_spa.pdf?sequence=1&isAllowed=y
4. http://www.minsa.gob.pe/renhice/documentos/Pre2Jor/01%20Diego%20L%C3%B3pez%20-%20Principios%20de%20Interoperabilidad%20&%20Est%C3%A1ndares%20Proyectos%20Internacionales.pdf
5. http://iris.paho.org/xmlui/handle/123456789/28188
6. http://www.hl7.org.ar/
7. http://hl7es.blogspot.com/2013/12/fhir-el-nuevo-miembro-de-la-familia-hl7.html
8. https://www.argentina.gob.ar/salud/snomed
9. https://www.argentina.gob.ar/salud
10. https://www.hospitalitaliano.org.ar/#!/home/principal
11. https://www.mais.org.ar/
12. http://hsshealth.com/historia-clinica-electronica-unificada/
13. https://reunir.unir.net/bitstream/handle/123456789/3077/Cristina_delaFuente_Gomez.pdf?sequence=1
14. http://artemisa.unicauca.edu.co/~peter/docs/ArticuloSOA.pdf
15. https://revistas.upb.edu.co/index.php/puente/article/viewFile/7104/6495

Framework para la Generación de Templates en Sistema de Catálogos de Realidad Aumentada

Nahuel Mangiarua, Jorge Ierache, Martin Becerra, Hernán Maurice, Santiago Igarza, Osvaldo Spositto

Universidad Nacional de La Matanza, Departamento de Ingeniería e Investigaciones Tecnológicas.

Grupo de Realidad Aumentada Aplicada
Florencio Varela 1903, La Matanza, Buenos Aires, Argentina
{jierache,nmangiarua}@unlam.edu.ar

Abstract. Se presenta el desarrollo de un sistema de aumentación de metacontenidos sobre las bases del sistema de catálogo virtuales aumentados [1], [2] a fin de mejorar la usabilidad del sistema original para los usuarios no expertos. Se propone pues la incorporación del concepto de template de aumentación de la realidad para los catálogos de realidad aumentada. Los mismos permiten predefinir la cantidad y tipos de contenidos de realidad aumentada, junto con sus transformaciones geométricas y el orden de visualización relativo al resto de los elementos. Así, al aplicarlo a un catálogo aumentado, ayudará a mantener un formato uniforme entre todos sus marcadores, como también simplificará la carga del material a aumentar por parte de los usuarios no expertos al incorporar terminologías propias del tema de explotación de dicho catálogo.

Keywords: Realidad Aumentada, Catálogo Virtual Aumentado, herramienta de generación de contenidos aumentados

1 Introducción

Dentro del marco del Sistema de Catálogos Virtuales Aumentados [1], [2] se plantea el objetivo de lograr la simplificación de la labor del usuario al momento de subir y publicar su propio contenido aumentado. Dicho objetivo nace sobre dificultades observadas durante el trabajo de extensión del sistema para su aplicación concreta al evento Bienal de Arte Integral realizado en la Universidad Nacional de La Matanza. Durante el mismo, se encontró que, si bien las interfaces web implementadas eran relativamente simples, el usuario no escapaba a la necesidad de comprender los conceptos básicos de la Realidad Aumentada, además, la generación del Marcador y el Contenido Aumentado incluidos en la creación de un Catálogo Virtual Aumentado [1] de tamaño considerable se volvía una tarea tediosa y propensa a pequeños errores de carga. Se propone entonces el desarrollo de un sistema de aumentación de metacontenidos sobre las bases del Sistema de Catálogos Virtuales Aumentados a fin de solucionar la problemática planteada, con la incorporación del concepto de template de aumentación de la realidad para los catálogos de realidad aumentada. Estos

templates permiten a un usuario con un nivel de conocimiento específico básico de la Realidad Aumentada (RA), definir la estructura de los elementos para un marcador. Al aplicarlo a un catálogo aumentado, el template permite mantener un formato uniforme entre todos los marcadores, como también simplifica la carga del material a aumentar por parte de usuarios sin necesidad de conocimiento específico de la RA al incorporar terminologías propias del tema de explotación de dicho catálogo. La Fig.1 muestra un diagrama conceptual del sistema de templates para catálogos virtuales aumentados.

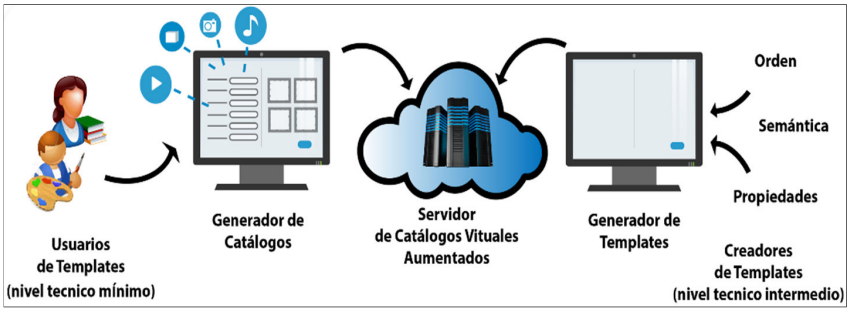


Fig. 1. Diagrama Conceptual del sistema de templates.

En cuanto a la RA, esta tecnología generalmente es comparada con la Realidad Virtual, sin embargo, no se refieren a lo mismo. Esta última introduce al usuario en un ambiente completamente virtual sustituyendo a la realidad física, mientras que RA complementa (aumenta) a la realidad física, manteniendo al usuario en contacto con la misma, a la vez que interactúa con objetos o información virtual superpuestas sobre ella [3]. La Realidad Aumentada es un conjunto de dispositivos que añaden información virtual a la información física ya existente, es decir, añadir una parte sintética virtual a lo real. Esta es la principal diferencia con la realidad virtual, puesto que no sustituye la realidad física, sino que sobreimprime los datos informáticos al mundo real. [4] La combinación de la información obtenida desde distintas fuentes, más aquella que podemos obtener del contexto gracias al uso de los sensores y componentes que están presentes en las nuevas generaciones de dispositivos móviles, proporcionan la piedra angular para la construcción de aplicaciones que hagan uso de la Realidad Aumentada [5].

2 Templates de Aumentación de la Realidad

Mediante su interface web se presenta en la Fig.2, un template para el Sistema de Catálogos Virtuales Aumentados permite definir la cantidad y tipos de contenidos de realidad aumentada (texto, imagen, audio, objetos 3D, etc.), junto con sus transformaciones geométricas (posición, rotación, escala), un nombre y el orden de visualización relativo al resto de los contenidos para un marcador. Adicionalmente, permiten definir un término que aplicará sobre el concepto mismo de marcador,

otorgándole un sentido semántico propio del área temática para el cual se crea el template.

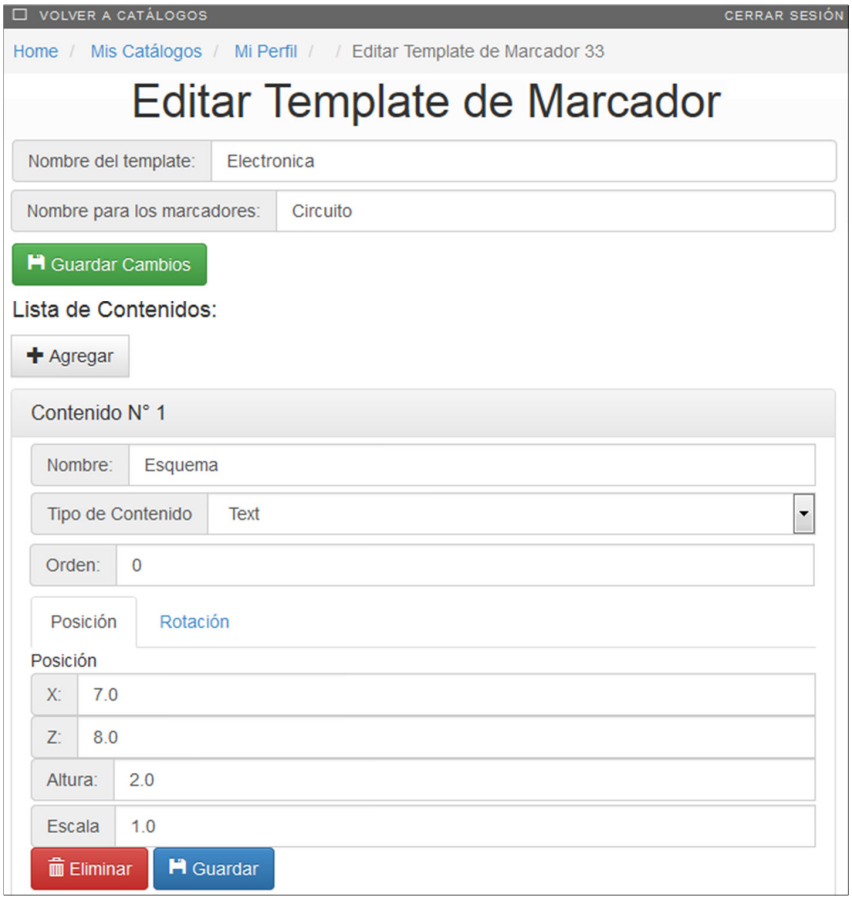


Fig. 2. Interfase web de edición de template, donde se agrega un contenido de tipo imagen llamado Esquema con sus respectivos valores de orden y posición.

Se presenta en la en la Fig.3, el momento de la creación de un Catálogo Virtual Aumentado, donde se puede seleccionar un template existente mediante una lista desplegable. Una vez aplicado, el template modifica el flujo de trabajo Fig.4 y la apariencia del editor de catálogo y marcadores. Intercambiando el término “marcador” por el definido en la template y proporcionando un nombre a los contenidos, se vuelve explícita la temática del catálogo y se facilita su correcta interpretación por parte del usuario creador de contenidos de RA. Al mismo tiempo, la creación de un elemento (marcador) nuevo conlleva la inmediata creación y configuración de los contenidos

definidos en el template, no sólo ahorrando tiempo, sino también abstrayendo al usuario creador de contenidos de RA de la configuración espacial necesaria para la correcta visualización del contenido en el ambiente de RA. En la figura 4 se detalla la carga de procesos con y sin el empleo del template del sistema, a continuación, en la figura 4a se presenta el flujo que se realiza empleando el sistema de template ilustrado para un caso en particular denominado bienal de arte. En la figura 4b se presenta el flujo de procesos que se debe realizar sin el empleo de template, detallando en rojo los pasos adicionales sin el uso del mismo.

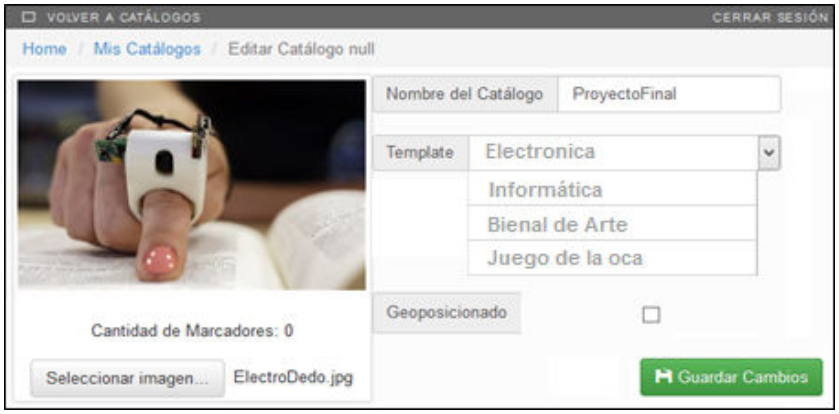


Fig. 3. Selección de template durante la creación de un catálogo.

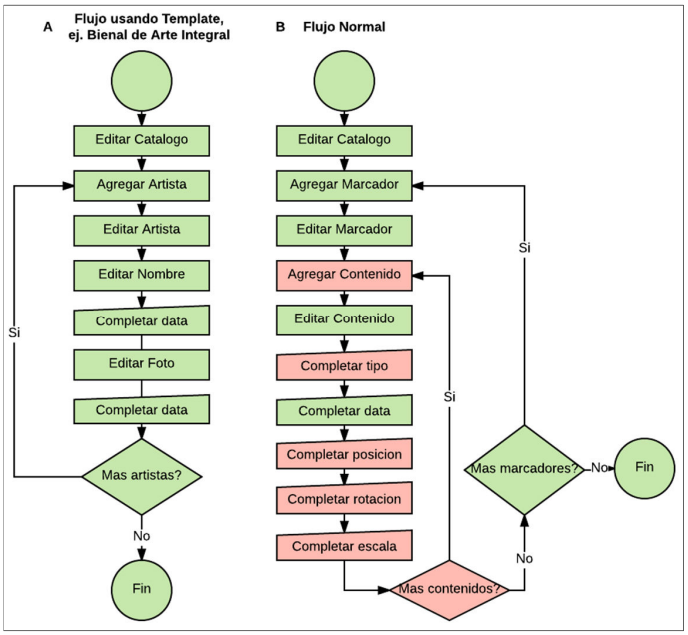


Fig. 4. Fig 4a y 4b. Flujo de procesos con y sin el empleo del template respectivamente

La implementación del Framework para la Generación de Templates, incorpora en el Sistema de Catálogos de Realidad Aumentada componentes pertenecientes a la creación, contenido y tipos de templates. Se puede apreciar en la Fig. 5, el servidor contiene el web editor, que nos permitirá la creación del contenido aumentado, el cual además puede hacer uso del componente "templates" proporcionando acceso a las distintas plantillas almacenadas para dicho usuario o para el grupo de usuarios. Ambos componentes mantienen comunicación con API REST para tener acceso a la base de datos y el envío de datos desde el servidor hasta la aplicación mobile. Esta otra aplicación toma la información proveniente del servidor de forma serializada y comprimida, gracias a los componentes de compresión y serialización la aplicación mobile puede tener acceso a la información de origen. Esta información contiene a los catálogos, marcadores y contenidos aumentados creados desde la aplicación web, los cuales son usados por el Core para establecer qué marcadores y qué contenidos serán puestos en el reconocimiento de marcadores y serán finalmente mostrados al usuario.

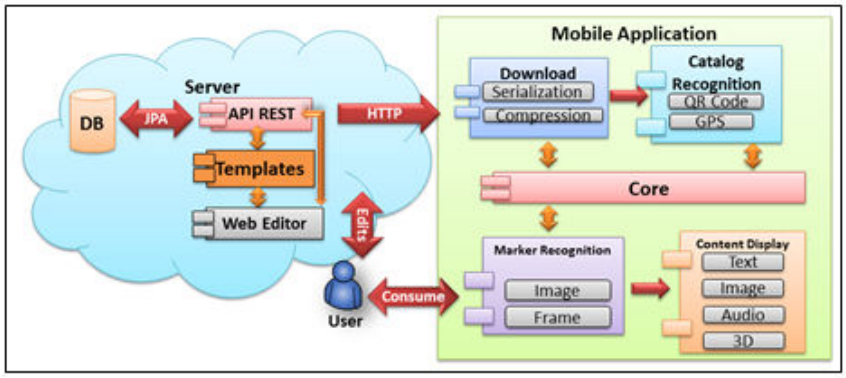


Fig. 5. Diagrama de arquitectura del sistema de catálogos aumentados

Este nuevo componente de template, asociado a la aplicación web contenida en el servidor, incorpora la clase ContentTemplate, encargada de contener todos los datos necesarios de cada template, se crea una instancia por cada contenido del template y posee las responsabilidades de almacenar el tipo de contenido (texto, imagen, mesh, audio, video) asociado desde la clase ContentType, el orden en el cual será mostrado ese contenido en el editor web, su transform (posiciones espaciales, dimensiones y rotación del contenido) y su MarkerTemplate asociado, que será el template padre de cada uno de los contenidos. Por esta razón la clase MarkerTemplate tiene las responsabilidades de obtener o brindar un array de contenidos pertenecientes a ese template y es un componente que agrega a la clase principal Catalog. A su vez, tenemos las clases propias del catálogo, la clase Content se encarga de la creación de cada contenido para un catálogo, se instancia una vez por cada contenido creado, almacena el tipo de contenido, su transform y contiene al Marker al cual pertenece. La clase Marker se encarga de contener un marcador nuevo que será utilizado para visualizar cada uno de los contenidos que se asociaron a él, por lo que contiene un array de contenidos aumentado, una imagen de marcador predefinida que es la primera imagen que se muestra en el editor al crearlo, y una imagen de base que es la imagen que sube el usuario y aparecerá en el centro del marcador, además está asociado al Catálogo al cual pertenece. Se tiene también grupos de usuarios que son creados con la clase AccessGroup, el cual agrega a la clase principal Catalog y contiene un array de los usuarios del catálogo que son creados desde la clase User, se instancia una vez por cada usuario y puede estar contenido o no en un grupo de usuario. Esta nueva clase posee las responsabilidades de obtener o almacenar la contraseña del usuario, nombre, lista de grupos de usuarios al cual pertenece y una lista de los catálogos a los cuales tiene acceso. Por último, tenemos a la clase principal Catalog, que posee responsabilidades de crear, eliminar, obtener o brindar marcadores de cada catálogo, asignar un usuario dueño de ese catálogo, dar acceso a un grupo de usuarios al mismo y establecer, si es requerido por el usuario, un template predefinido a ese catálogo. Ver Fig 6.

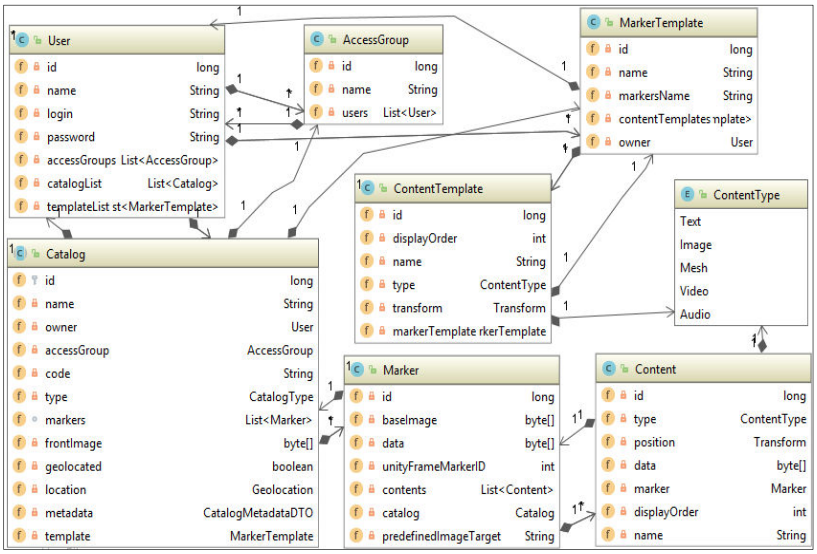


Fig. 6. Diagrama de clases incluyendo la implementación de templates

3 Presentación de un Caso de empleo real del sistema: "Bienal de Arte Integral 2016"

El framework de templates para el sistema de Catálogos Virtuales Aumentados [6] fue puesto a prueba durante el trabajo de extensión para el evento Bienal de Arte Integral 2016, de la Universidad Nacional de La Matanza, facilitando la descarga de la aplicación [7]. Este proyecto requería la creación de un catálogo virtual para la aumentación de obras de arte plástica en exhibición, y de su correspondiente catálogo impreso para un total aproximado de 30 marcadores con múltiples contenidos de diversos tipos. Mediante la creación del template "Bienal de Arte Integral", se definió en colaboración con la curadora del evento, la cantidad, tipos, nombres y transformaciones geométricas de los contenidos que debían ser presentados en realidad aumentada para cada obra, fijándose el término "artista" para reemplazar el concepto de marcador. Para la creación de este template se siguen los siguientes pasos que se detallan a continuación. Desde el panel principal de catálogos se puede apreciar un catálogo ya creado "BAI 2016" y botones para agregar un nuevo template o catálogo. Se accede entonces a la creación de un nuevo template (Fig. 7a). En el siguiente paso se coloca un nombre de template, en este caso "Bienal de Arte Integral Template" y se determina un nombre para los marcadores, que gracias a esto luego serán denominados "Artista". (Fig. 7b). Una vez guardado los cambios aparecerá un botón para agregar cada uno de los contenidos pertenecientes al template (Fig. 7c) lo cual abrirá el editor de contenido de templates. En este ejemplo en particular (Fig. 7d) se muestra cómo se hace la carga del contenido de tipo texto llamado "Título de la obra", campo que será

mostrado en segundo orden en el editor de contenidos del catálogo cuando se use este template, también se pueden asignar la posición, dimensiones y rotación que tendrán en la aplicación móvil, este paso se repite por cada contenido que se quiera agregar al template.

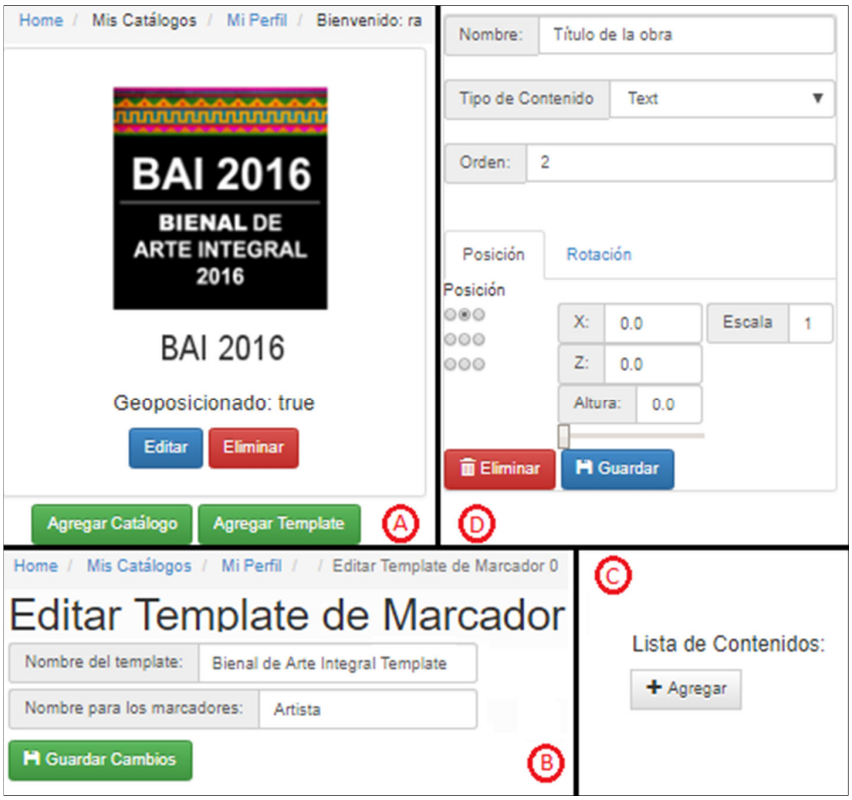


Fig. 7. Interfaces por pasos del editor de templates

Al aplicar dicho template sobre un nuevo catálogo virtual aumentado, se logró entonces que el personal organizador del evento cargará y editará los contenidos concretos de cada artista en el editor de catálogos tras una breve introducción verbal, sin la intervención de ningún miembro del grupo de investigación. Como resultado de este trabajo se puede ver un catálogo que contiene tres artistas con sus respectivos contenidos y marcadores (Fig. 8). Esto actuó como pruebas de producción del sistema.



Fig. 8. Catálogo creado a partir del template "Bienal de Arte Integral".



Fig. 9. Aplicación móvil de Bienal de Arte Integral 2016.

Las pruebas llevadas a cabo durante la Bienal de Arte Integral 2016 (ver Fig. 9) fueron satisfactorias, se contó con varias personas involucradas en el ingreso de contenidos, obteniendo un feedback sobre la utilización del sistema. Los desarrollos a futuros están destinados a expandir las posibilidades del framework de generación de templates incorporando grupos de usuarios donde los mismos podrán compartir sus templates para que otras áreas específicas e incluso ayuda a que nuevos usuarios se adapten más rápido a un proyecto ya creado, con el fin de realizar un mantenimiento al mismo.

4 Conclusiones y Futuras Líneas de Trabajo

El Framework de generación de templates integró exitosamente el sistema de catálogos aumentados contribuyendo a facilitar la tarea de creación de contenido aumentado específico para los usuarios. Su utilización mejora el tiempo que demanda la creación de contenidos en usuarios o grupos de usuarios puedan hacer uso del mismo o personalizarlo a sus necesidades. De la experiencia obtenida con el empleo del sistema se consideran como futuras líneas de trabajo la incorporación de tipos de contenidos dinámicos e interactivos. Los primeros permitiendo mostrar datos actualizables en tiempo real, tales como últimos comentarios de redes sociales, valores de sensores, etc, mientras que los últimos permitirán el ingreso o selección de valores de contenidos aumentados por parte del usuario del módulo web, proveyendo un canal de feedback para quien publica un catálogo aumentado.

5 Agradecimientos

Nuestro agradecimiento al Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza, el presente trabajo se financia en el marco del proyecto PROINCE C-185 C,202.

6 Referencias

1. Jorge Ierache, Nahuel Adiel Mangiarua, Sebastián Ariel Bevacqua, Nicolás Nazareno Verdicchio, Martín Ezequiel Becerra, Diego Rubén Sanz, Matías Ezequiel Sena, Fernando Martín Ortiz, Nicolás Daniel Duarte, Santiago Igarza (2014). "Sistema de Catálogo para la Asistencia a la Creación, Publicación, Gestión y Explotación de Contenidos Multimedia y Aplicaciones de Realidad Aumentada". XX Congreso Argentino de Ciencias de la Computación (Buenos Aires, 2014), Red de Universidades con Carreras de Informática (RedUNCI). http://hdl.handle.net/10915/42339
2. Ierache, J., Mangiarua, N., Bevacqua, S., Verdicchio, N., Becerra, M., Sanz, D., Sena, M., Ortiz, F., Duarte, N., Igarza, S. (2015). "Development of a Catalogs System for Augmented Reality Applications". World Academy of Science, Engineering and Technology, International Science Index 97, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 9(1), 1 - 7. http://waset.org/Publications/development-of-a-catalogs-system-for-augmented-reality-applications/10000077
3. Manresa Yee, M. Abásolo, R Más Sansó and M Vénere:. Realidad virtual y realidad aumentada. Interfaces avanzadas. (2011)
4. Jorge Gaviria Hincapie1, Guillermo Alonso Castaño, Byron PortillaRosero, Jose León Sierra Ospina: SLD203 Realidad Aumentada En el Tratamiento de las Enfermedades Mentales y las Adicciones. XV edición de la Convención y Feria Internacional Informática (2013)
5. Pablo J. Iuliano, Claudia A. Queiruga, Francisco J. Diaz: UNLP Aumentada: Desafíos y Retos. In: Biennial Congress of Argentina (ARGENCON), 2014 IEEE pp. 43-18. (2014)
6. N. Mangiarua, C. Montalvo, F. Petrolo, D. Sanz, N. Verdicchio, E. Lobatto,A.Rosenthal, M. Becerra, S. Igarza, J. Ierache. Framework para la Generación de Templates en Sistemas de Catálogos de Realidad Aumentada. WICC 2017 ISBN (en prensa).
7. BAI2016: https://play.google.com/store/apps/details?id=com.GrupoRA.BAI2016

Desarrollo multiplataforma de Aplicaciones Móviles combinadas con el uso de Beacons

Franco M. Borrelli[1], Pedro Brost[1], Cecilia Challiol[1,2], Diego H. Orellano[1], Facundo I. Mendibru[1], Matías A. Santoleri[1] and Federico M. Alconada Verzini[1]

[1] UNLP, Facultad de Informática, LIFIA. Calle 50 y 120, La Plata, Argentina.
[2] CONICET, Argentina
{fborrelli, pbrost, ceciliac, dorellano, fmendiburu, msantoleri, falconada}@lifia.info.unlp.edu.ar

Resumen. En este trabajo se presenta la exploración de dos APIs de *beacons* existentes, una para *PhoneGap* y otra para *React Native*. Para esta exploración, se desarrollaron dos aplicaciones móviles, una en *PhoneGap* y otra en *React Native*; las cuales usan estas APIs para detectar la proximidad a los *beacons*, al entrar en la proximidad de un *beacon*, se brinda en pantalla la información del mismo. Se detallan distintas características de estas aplicaciones desarrolladas, haciendo hincapié en el análisis comparativo de estos desarrollos respecto a lo que provee cada API. Además, el trabajo presenta un espacio de discusión para analizar las lecciones aprendidas en la exploración realizada, en pos de contribuir a la temática relacionada al desarrollo de tipo de aplicaciones.

Palabras Claves: Mecanismos de Sensado de Posicionamiento Indoor; Desarrollo Multiplataforma de Aplicaciones Móviles; Beacons; PhoneGap; React Native.

1 Introducción

Los avances tecnológicos de los últimos años han permitido que emerjan diferentes mecanismos de sensado de posicionamiento para espacios indoor [1], [2]. Cada uno de estos tiene diferentes niveles de precisión [2]. Es decir, brindan la posición actual del usuario con mayor o menor nivel de exactitud respecto de la posición real del mismo. En algunos casos, estos mecanismos se pueden combinan para estimar con mayor precisión la posición actual del usuario [1]. Actualmente, uno de los mecanismos de sensado que se perfila como confiables y eficientes, acorde a [1], son los *beacons*; los cuales emiten datos por radiofrecuencia utilizando tecnología BLE (*Bluetooth Low Energy*) [1]. Estos se caracterizan por tener bajo costo y además su uso conlleva bajo consumo de energía.

Las aplicaciones móviles que usan mecanismos de sensado de posicionamiento para brindar servicios o información acorde a la posición actual del usuario, se denominan *sensibles a la posición*, y son un subconjunto de las aplicaciones sensibles al contexto [3], [4]. Este tipo de aplicaciones todavía no cuenta con una solución unificada, es decir, su abordaje sigue siendo un tema abierto de investigación [3]. Más

aún, en [4] se plantea la importancia de crear aplicaciones sensibles al contexto realmente usables, ya que muchas de las aplicaciones existentes no están focalizadas en las necesidades reales de los usuarios. La mayoría de los framewoks o modelos de contexto consideran la adquisición del valor sensado como algo que "*llega de alguna manera*" [3], pero no profundizan en cómo dicho valor se sensa. Por otro lado, los mecanismos de sensado de posicionamiento son analizados desde la perspectiva de cómo funcionan (como se detalla en [1] y [2]), sin hacer hincapié en qué servicios se brindan con ese valor sensado [1]. En [2] se presenta un ranking de cuáles son los mecanismos de sensado de posicionamiento más apropiados para distintos tipos de aplicaciones. Por ejemplo, los *beacons* se clasifican como el mejor mecanismo para aplicaciones de navegación indoor y tracking de usuarios. Sin embargo, en [2] no se detalla cómo se deberían desarrollar estas aplicaciones acorde a cada tipo de sensado.

El desarrollo de aplicaciones móviles, como se menciona en [5] y [6], también viene evolucionando en los últimos años debido a la constante demanda del mercado. Para cubrir la mayor cantidad de dispositivos móviles, y llegar así a más usuarios, han surgido los desarrollos multiplataforma; los cuales pueden ser, por ejemplo, híbridos (como *PhoneGap* [7]) o interpretados (como *React Native* [8]). Cada uno de estos tiene un flujo de ejecución distinto generando así ventajas y desventajas [6]. Los desarrollos multiplataforma permiten acceder a características propias de los dispositivos móviles como en un desarrollo nativo, por ejemplo, acceder a los sensores de los dispositivos. En [5] se plantea que el ciclo de vida de las aplicaciones creadas con desarrollos multiplataforma es un tema poco explorado, en particular, cómo abordar el mantenimiento de las mismas. Esto podría complejizarse aún más, si este tipo de aplicaciones se combina con mecanismos de sensado de posicionamiento.

El uso de *beacons* se viene explorando en diferentes dominios tales como publicidad [9], sistemas de salud [10], guías móviles dentro de museos [11], etc. Esta exploración está en una etapa inicial, principalmente se viene analizando para cada dominio cómo se comporta este mecanismo de sensado o qué servicios se pueden brindar a partir del mismo. Este análisis se realiza generalmente mediante aplicaciones móviles ad-hoc prototípicas (por ejemplo, [9], [10] y [11]). Sin embargo, hay poco explorado respecto a la comparación de diferentes tipos de desarrollos multiplataforma para aplicaciones móviles sensibles a la posición que usan *beacons*, esto es la principal motivación del presente trabajo.

El objetivo principal del trabajo es presentar la exploración relacionada al uso de dos APIs de *beacons* existentes, una para *PhoneGap* (*híbrido*) [12] y otra para *React Native* (*interpretado*) [13]. Para esta exploración, se desarrollaron dos aplicaciones móviles, una en *PhoneGap* y otra en *React Native*; estas aplicaciones usan las APIs mencionadas para detectar la proximidad a los *beacons*. De esta forma, al entrar en la proximidad de un *beacon*, se brinda en pantalla la información del mismo. Se realiza un análisis comparativo de estos desarrollos respecto a lo que provee cada API. Además, el trabajo presenta un espacio de discusión para analizar las lecciones aprendidas en la exploración realizada, en pos de contribuir a la temática de contar con una solución unificada para abordar el desarrollo de las aplicaciones móviles sensibles al contexto.

El trabajo se estructura de la siguiente manera. En la Sección 2 se describe el estado del arte acorde al objetivo del trabajo. En la Sección 3 se presenta la exploración realizada sobre el uso de dos APIs de *beacons* existentes. Algunas

características relacionadas a la temática son discutidas en la Sección 4. Las conclusiones y los trabajos futuros son presentados en la Sección 5.

2 Estado del Arte

En el último año han surgidos distintos tipos de *beacons* poco documentados en la bibliografía académica, por esta razón estos se detallan ampliamente. El desarrollo de aplicaciones móviles se describirá brevemente, ya que esta temática está más explorada y analizada en la bibliografía existente.

2.1 Beacons

Un *beacon* es un dispositivo que emite datos por radiofrecuencia utilizando tecnología BLE (*Bluetooth Low Energy*) [1]. Existen distintos tipos de *beacons*, los tres más conocidos se describen a continuación:

- *Beacons de Proximidad*, fueron los primeros en surgir y son al día de hoy los que más se utilizan. Estos plantean "*etiquetar*" ciertas áreas o *regiones* específicas, las cuales pueden resultar de interés para reaccionar ante la proximidad a las mismas. Estas áreas son circulares y el *beacon* se ubica en el centro de estas. Una aplicación móvil que usa estos *beacons* podría detectar cuando el usuario ingresa a unas de estas áreas, y brindarle alguna información o servicio. Por ejemplo, un área de interés podría ser un local de un shopping [9]; cuando el usuario ingresa a esta área (porque está en la proximidad de un *beacon*), se le brindan ofertas de ese local. En la Figura 1.a se puede apreciar a un usuario dentro de la región de un *beacon*.
- *Beacons de Posicionamiento* (*Location UWB Beacons*), han sido lanzados por la empresa *Estimote* [14]; permiten obtener la posición del usuario en un formato (x,y) en relación a una zona de sensado. Estos *beacons* requieren primero establecer la zona de sensado, para lo cual se podría usar una aplicación brindada por *Estimote*, la cual crea un mapa de la zona automáticamente, para luego poder posicionar al usuario sobre el mismo. En una zona al menos debe haber cuatro *beacons* de este estilo. Cabe mencionar que estos *beacons* también pueden funcionar como de *proximidad* en el caso de no usar el concepto de zona de sensado. En [10] se utilizan estos *beacons* para saber la posición de una enfermera en un centro asistencial para adultos mayores. En la Figura 1.b se puede observar un usuario dentro de la zona de sensado delimitada por cuatro *beacons*.
- *Beacons Stickers (o Nearables)*, estos también han sido creados por la empresa *Estimote* [14]; sirven para posicionar objetos que pueden moverse en una zona de sensado, es decir, se combinan con los *beacons de posicionamiento* mencionados anteriormente. Un aplicación móvil podría reaccionar al detectar el movimiento o cercanía de este tipo de *beacon*, además de poder determinar en qué posición (x,y) se encuentran los mismos. En [10] se utilizan estos *beacons* para saber la posición de una paciente con capacidad de movilidad reducida en un centro asistencial para adultos mayores. En la Figura 1.c se puede observar un *sticker* dentro de la zona de sensado delimitada por cuatro *beacons de posicionamiento*.

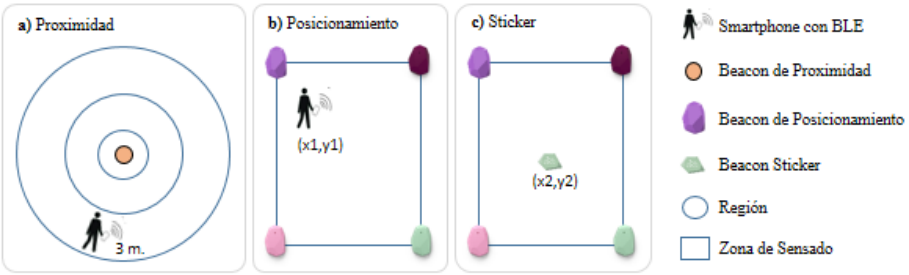


Fig. 1. Ejemplos de tipos de *beacons*. 1.a) *beacon de proximidad*. 1.b) *beacons de posicionamiento*. 1.c) *beacon sticker delimitado por beacons de posicionamiento*.

Cabe mencionar que los *beacons de proximidad* pueden combinarse bajo alguna heurística para funcionar en forma conjunta como los *beacons de posicionamiento*, es decir, brindar una posición (x,y). Esto requiere desarrollar el algoritmo de análisis de señales, algo que es trasparente cuando se usan *beacons de posicionamiento*. En [15] se explora el uso de seis *beacons de proximidad* delimitando una área cuadrada para analizar la precisión de la posición en términos de (x,y); los autores identifican un margen de error de entre treinta y setenta centímetros en las pruebas realizadas. Por otro lado, en [11] se exploran cómo se comportan nueve *beacons de proximidad* formando un área circular, una línea o un cuadrado. Algo similar ocurre con la exploración realizada en [16] donde se prueban distintas distribuciones de los *beacons*. Los autores mencionan en [11] y [16] que en las áreas donde hay muchos *beacons* para determinar la posición se logra mayor precisión, pero esto requiere colocar en pocos metros muchos *beacons*, elevando así el costo implicado.

2.1.1 Protocolos asociados a los distintos tipos de beacons

Cada uno de los tres tipos de *beacons* mencionados anteriormente (*proximidad*, *posicionamiento* y *sticker*) tiene protocolos para transmitir datos. Para los *beacons de proximidad* existen distintos protocolos que pueden ser usados por las aplicaciones móviles que hacen uso de estos *beacons*. Por ejemplo, *iBeacon* [17] es un protocolo desarrollado por *Apple* (en el año 2013). Si bien este protocolo brinda soporte para los dispositivos móviles con iOS, existe una versión open source compatible, denominada *AltBeacon* [18], que da soporte a otras plataformas como *Android*. Otros protocolos son *Estimote Monitoring* [19] implementado por *Estimote* y *Eddystone* [20] desarrollado por *Google*. Todos estos protocolos de proximidad definen mínimamente dos eventos: entrada y salida a una región. Estos eventos son abordados por los diferentes protocolos bajo distintos nombres: *Monitoring* tanto en *iBeacon*, *AltBeacon* y *Estimote*, mientras que *Eddystone* lo llama *UID* (y *EID* a su versión segura, que garantiza que solo las aplicaciones y servicios autorizados pueden comunicarse con los *beacons*).

En la Tabla 1 se puede apreciar el tiempo de demora (*delay*) que tiene cada uno de los protocolos mencionados para detectar la entrada y salida a las regiones. Se puede observar en dicha tabla que todos los protocolos permiten funcionamiento en

background. En la misma también se analizan otras características, las cuales se describen a continuación. La *Configuración de Rango de Acción* se refiere a la capacidad de poder definir distintos rangos de proximidad sobre los cuales actuar. En relación a esto se puede apreciar que solo el protocolo *Estimote Monitoring* brinda soporte a esta característica. La *Cantidad Máxima de Regiones*, define el número máximo de regiones detectables simultáneamente; en el caso de *iBeacon*/*AltBeacon* esto se limita a veinte regiones simultáneas, mientras que en *Estimote Monitoring* es ilimitado. El *Efecto Rebote* es ese "*parpadeo*" que se puede visualizar en una aplicación al estar cerca de un determinado *beacon*; esto sucede porque se acciona falsamente un evento de salida y luego de unos pocos segundos se acciona nuevamente un evento de entrada. Este tipo de comportamiento puede afectar la experiencia del usuario que usa la aplicación. En la Tabla 1 se puede apreciar que solo el protocolo *Estimote Monitoring* brinda soporte para evitar este efecto.

Tabla 1. Comparación entre los distintos protocolos de proximidad.

	iBeacon/AltBeacon	Estimote Monitoring	Eddystone UID/EID
Tiempo de demora del evento de entrada	< 3 segundos	< 3 segundos	< 3 segundos
Tiempo de demora de evento de salida	20-30 segundos	< 5 segundos	20-30 segundos
Funcionamiento en background	si	si	si
Configuración de Rango de Acción	no	si	no
Cantidad Máxima de Regiones	20	ilimitado	-
Efecto Rebote ("*parpadeo*")	si	no	si

Además, el protocolo *iBeacon*/*AltBeacon* provee la funcionalidad de *ranging*, la cual para una determinada región retorna una lista con los datos básicos de los *beacons* detectados, así como también información adicional como la proximidad estimada entre el dispositivo y cada uno de los *beacons*. Esta información podría ser útil para poder realizar heurísticas para que estos *beacons* se comporten como de *posicionamiento* (como se realiza, por ejemplo, en [11], [15] y [16]). *Eddystone* por su parte provee además el protocolo *Eddystone URL*, el cual resulta útil para transmitir direcciones URL. Supongamos que un cine tiene *beacons* cerca de los banners de presentación de las películas en cartelera; cuando el usuario está cerca de uno de los *beacons*, le llega una notificación al celular con el link para ver el tráiler de la película asociada al *beacon*.

Para los *beacons de posicionamiento*, *Estimote* provee el protocolo denominado *Estimote Location Protocol* [21], el cual permite recibir información respecto a la posición (x,y) de un dispositivo móvil o sticker, facilitando que la misma sea mostrada en un mapa. Este protocolo está pensado para agilizar la construcción de aplicaciones móviles basadas en posicionamiento indoor. Para este tipo de *beacons* se

pueden recibir, mediante otro protocolo, los datos telemétricos, que incluyen, por ejemplo, temperatura, luminosidad, etc.

En el caso de los *beacons Stickers*, *Estimote* provee el protocolo denominado *Nearable* [22], el cual permite enviar información contextual en un solo paquete, como por ejemplo, temperatura, nivel de batería, movimiento sobre un (x,y,z), orientación en el espacio, etc.

2.2 Desarrollo de Aplicaciones Móviles

En esta sección se hará hincapié solo en aquellas características que pueden ayudar al lector a comprender mejor la exploración que se presenta en este trabajo. Dada las características de las aplicaciones móviles sensibles a la posición [3], es de interés cubrir la mayor cantidad de dispositivos móviles, para llegar así a más usuarios. Acorde a esto, el desarrollo multiplataforma ([5], [6]) de este tipo de aplicaciones agiliza tiempos y cubre mayor cantidad de dispositivos. En [6] se clasifican estos desarrollos como: híbrido (por ejemplo, *PhoneGap*), interpretado (como es *React Native*), complicación cruzada, basado en componentes y desarrollo dirigido por modelos. Acorde a [6], estos dos últimos son los menos explorados en la bibliografía. Cabe mencionar que los enfoques *híbridos* muestran su funcionalidad sobre un *WebView*, mientras que los *interpretados* compilan a nativo esta funcionalidad. En [6] se presenta y detalla cómo es el flujo de ejecución para el acceso a características internas del dispositivo móvil, tanto en *PhoneGap* como en *React Native*, analizando ventajas y desventajas de cada uno.

Considerando la popularidad que tienen *PhoneGap* y *React Native*, estos fueron seleccionados para hacer la exploración presentada en este trabajo, sumado a que existen APIs disponibles ([12] y [13] respectivamente) para acceder a los datos de los *beacons*. En particular, estas APIs utilizan el protocolo de proximidad de *iBeacon/AltBeacon* que se describió en la Sección 2.1.1.

Un tema abierto es cómo abordar el mantenimiento de las aplicaciones móviles generadas a partir de desarrollos multiplataforma [5]. Esta problemática impacta en forma directa cuando este tipo de aplicaciones tiene que usar además sensores, por ejemplo, *beacons*; ya que se le suma la complejidad del mantenimiento de los mismos. Supongamos que un *beacon* deja de funcionar, el mismo debe reponerse en la misma posición; caso contrario, las aplicaciones móviles podrían brindar servicios inconsistentes, por ejemplo, al posicionar al usuario en un lugar distinto a la posición real del mismo. Esto guarda relación con lo que se plantea en [4], sobre la importancia de crear aplicaciones sensibles al contexto realmente usables.

3 Exploración de algunas APIs de proximidad

En esta sección se presenta la exploración realizada sobre el uso de dos APIs de *beacons* existentes, una para *PhoneGap* (desarrollo multiplataforma híbrido) [12] y otra para *React Native* (desarrollo multiplataforma interpretado) [13]. Para la exploración, se desarrollaron dos aplicaciones móviles, una en *PhoneGap* y otra en *React Native*; cabe mencionar que el foco está puesto en explorar el uso de las APIs y

cómo estas se comportan, y no se hace hincapié en el desarrollo de una aplicación a gran escala.

Para la exploración se usaron cuatro *beacons de posicionamiento* (*Location UWB Beacons*) de *Estimote* [14], los cuales se describieron en la Sección 2.1. Estos *beacons* pueden transmitir datos usando tanto protocolos de proximidad como de posicionamiento, ambos fueron descriptos en la Sección 2.1.1. *Estimote* provee una página de configuración (*Estimote Cloud)*, para setear diferentes valores a los *beacons* utilizados. Adicionalmente, *Estimote* brinda algunas aplicaciones *templates* que permiten probar algunas funcionalidades de los *beacons de posicionamiento*. Estas aplicaciones *templates* fueron usadas principalmente para poder comprender cómo funcionan estos *beacons*.

Cabe mencionar que las APIs de *beacons* usadas ([12] y [13]) funcionan sólo para el protocolo de proximidad de *iBeacon/AltBeacon*. Acorde a esto, se desarrolló la funcionalidad de las dos aplicaciones móviles para la exploración; las cuales monitorean la proximidad a los *beacons*; y al entrar en la proximidad de alguno de ellos, se brinda en la interfaz la información del mismo. Para esto, se aprovechó las características de *ranging* que provee el protocolo de proximidad de *iBeacon/AltBeacon* (como se describió en la Sección 2.1.1). Más allá de las diferencias que propone el desarrollo de aplicaciones *PhoneGap* y *React Native,* ambas APIs ([12] y [13]) no difieren en la interfaz de acceso a aquellos datos comunes.

La restricción de las APIs respecto al protocolo que usa (de proximidad) también impactó en cómo los cuatro *beacons de posicionamiento* fueron usados, es decir, solo se accedió a los datos brindados por el protocolo de proximidad de los mismos.

A continuación se presentan las principales características relacionadas al desarrollo de las dos aplicaciones móviles que usan APIs de *beacons* ([12] y [13]). Para más detalle de cómo funcionan en general *PhoneGap* y *React Native* se puede consultar [5], [6], [7] y [8]. En la Figura 2 se puede apreciar el código implementado en cada aplicación para dar comienzo al *ranging*.

a) PhoneGap

```
const uuid =
cordova.plugins.locationManager.BeaconRegion.
WILDCARD_UUID;
const major = undefined;
const minor = undefined;
const identifier = 'REGION1';
const beaconRegion = new
cordova.plugins.locationManager.BeaconRegion
(identifier, uuid, major, minor);
startRanging = () => {
  cordova.plugins.locationManager
    .startRangingBeaconsInRegion
    (beaconRegion)
    .fail((e) => {
      console.error('Error while starting
      ranging');
    })
    .done();
}
```

b) React Native

```
Beacons.detectIBeacons();
try {
  Beacons.setForegroundScanPeriod(2000);
  await Beacons.startRangingBeaconsInRegion
  ('REGION1');
  console.log(`Beacons ranging started
  succesfully!`);
} catch (error) {
  console.log(`Beacons ranging not started,
  error: ${error}`);
}
```

Fig. 2. Código para dar comienzo al *ranging*. 2.a) *PhoneGap*. 2.b) *React Native.*

En la Tabla 2 se puede observar la comparación de los datos obtenidos usando *ranging* con las dos APIs usadas. Por simplicidad, no se especifican en esta tabla los datos generales respecto a los *beacons*, ya que ambas APIs los brindan; estos datos son el UUID (*Universally Unique Identifier*) que identifica a un *beacon* y el *Major-Minor* que son valores numéricos que oscilan entre 0 y 65535, los cuales sirven para identificar a un *beacon* con mayor precisión.

Tabla 2. Comparación de los datos obtenidos usando *ranging* con las dos APIs de *beacons*.

Dato	Tipo	API para PhoneGap	API para React Native
Proximity (posibles valores: "*inmediate*", "*near*", "*far*")	String	si	si
RSSI (*fuerza de recepción de la señal*)	Double	si	si
TX *(fuerza de la señal de transmisión del beacon)*	Double	si	no
Distance *(en metros)*	Double	Calculable	si
Accuracy *(respecto de RSSI)*	Double	si	no

Cada una de las APIs usadas en los desarrollos solo permiten mostrar aquellos datos que se especifican en la Tabla 2. Por ejemplo, la aplicación desarrollada para *React Native* no puede acceder al TX o Accuracy; mientras que la aplicación *PhoneGap* puede mostrar la distancia siempre y cuando se implemente una función para tal fin, ya que esta es calculable. En la Figura 3 se puede observar el código de ambas aplicaciones para mostrar los datos obtenidos del *ranging*. Cabe destacar que solo se hace hincapié en el código relevante para mostrar las características mencionadas.

a) PhoneGap

```
function toString(beacon) {
  return (
    'Color: <span class="dot ' +
    this.getName(beacon.uuid) +
    '"></span> (' + this.getName(beacon.uuid) +
    ') <br> UUID: ' + beacon.uuid +
    '<br> major: ' + beacon.major +
    '<br> minor: ' + beacon.minor +
    '<br> Proximity: ' + beacon.proximity +
    '<br> rssi: ' + beacon.rssi +
    '<br> TX: ' + beacon.tx +
    '<br> Distance: ' +
    this.calculateDistance(beacon.rssi, beacon.tx) +
    ' meters <br> Accuracy: ' + beacon.accuracy +
    '<hr>'
  );
}
```

b) React Native

```
const beacon = ({ data }) => {
  return (
    <Text style={styles.text}>
      <ColorLabel uuid={data.uuid} />{'\n'}
      UUID: {data.uuid}{'\n'}
      major: {data.major}{'\n'}
      minor: {data.minor}{'\n'}
      Proximity: {data.proximity}{'\n'}
      rssi: {data.rssi}{'\n'}
      Distance: {data.distance}{'\n'}
    </Text>
  );
};
```

Fig. 3. Código para mostrar los datos del *ranging*. 3.a) *PhoneGap*. 3.b) *React Native*.

4 Discusión

En esta sección se presentan algunas lecciones aprendidas durante la exploración realizada. Se espera que estas contribuyan en la temática de los mecanismos de sensado indoor [1], [2], la cual sigue siendo un área abierta de investigación.

Como se mencionó en la Sección 3, se usaron las aplicaciones *templates* brindadas por *Estimote* para comprender como funcionan los *beacons*. Se detectó que estas aplicaciones no tenían un comportamiento estable, lo cual impactó en el tiempo invertido en comprender cómo funcionan los *Location UWB Beacons*.

En la Tabla 2 de la Sección 3 se pudo apreciar como dos APIs que usan el mismo protocolo de proximidad, difieren en algunos datos que proveen. Esto puede impactar a la hora de tener que elegir un desarrollo usando *PhoneGap* o *React Native*. Es decir, estas APIs no tienen un comportamiento unificado aun usando el mismo protocolo.

La curva de aprendizaje del desarrollo de aplicaciones *PhoneGap* y *React Native*, es diferente. *PhoneGap* requiere conocimiento Web estándar, mientras que *React Native* requiere además aprender, por ejemplo, *React* (librería para interfaces de usuarios). Cabe mencionar que la experticia que tenga el desarrollador en el uso de buenas prácticas impacta en el mantenimiento del código, esto está más guiado en *React Native* ya que al desarrollar se debe seguir cierta estructura. En el caso de contar con experiencia previa en el desarrollo de *PhoneGap* o *React Native*, solo se deberá aprender cómo funcionan las APIs de *beacons*, las cuales están bien documentadas. Sin embargo, lograr aplicaciones móviles sensibles al contexto realmente usables [4] dependerá de la experticia de los desarrolladores en la temática.

Se deja planteada la siguiente pregunta abierta: *¿es el desarrollo multiplataforma la mejor estrategia?*, por ejemplo, *Airbnb* [5] venía usando *React Native*, sin embargo anunció que migrará al desarrollo nativo. Acorde a lo planteado en [5], *¿tendrá que ver esta decisión de Airbnb con la incertidumbre que hay sobre cómo mantener aplicaciones móviles generadas a partir de desarrollos multiplataforma?*.

5 Conclusiones y Trabajos Futuros

En este trabajo se presentó la exploración de dos APIs de *beacons* existentes. Se desarrollaron dos aplicaciones móviles, una en *PhoneGap* y otra en *React Native*; las cuales usan estas APIs para detectar la proximidad a los *beacons*. Estas aplicaciones al entrar en la proximidad de un *beacon*, brindan la información del mismo. Este funcionamiento simple permitió poder explorar la información brindada por cada una de estas APIs, detectando que aun usando ambas el mismo protocolo de proximidad, las APIs difieren en algunos datos que proveen. Por ejemplo, un dato que difiere es la distancia, la cual solo es provista por la API de *React Native*. Sin embargo, la distancia es calculable por la API de *PhoneGap*. Además, el trabajo presentó un espacio de discusión en el cual se detallaron las lecciones aprendidas en la exploración realizada, en pos de contribuir a la temática relacionada al desarrollo de este tipo de aplicaciones.

Como trabajo futuro se espera poder armar guías que asistan a los desarrolladores tanto en el diseño como en el desarrollo de este tipo de aplicaciones. Para esto se

considerarán las lecciones aprendidas tanto en el área de aplicaciones móviles sensibles al contexto [3], [4] como de los mecanismos de sensado [1], [2].

Referencias

1. Davidson, P., Piché, R.: A survey of selected indoor positioning methods for smartphones. IEEE Communications Surveys & Tutorials 9 (2), 1347-1370 (2017)
2. Basiri, A., Lohan, E.S., Moore, T., Winstanley, A., Peltola, P., Hill, C., Silva P.F.: Indoor Location Based Services Challenges, Requirements and Usability of Current Solutions. Computer Science Review 24, 1-12 (2017)
3. Alegre, U., Augusto, J., Clark, T.: Engineering context-aware systems and applications: A survey. Journal of Systems and Software 117, 55-83 (2016)
4. Alegre-Ibarra, U. Augusto, J.C., Carl, E.: Perspectives on engineering more usable context aware systems. Journal of Ambient Intelligence and Humanized Computing, 1-17 (2018)
5. Martinez, M., Lecomte, S.: Towards the quality improvement of cross-platform mobile applications. In: IEEE/ACM 4th International Conference on Mobile Software Engineering and Systems, pp. 184-188. IEEE Press, New York (2017)
6. Biørn-Hansen, A., Ghinea, G.: Bridging the Gap: Investigating Device-Feature Exposure in Cross-Platform Development. In: 51st Hawaii International Conference on System Sciences, pp. 5717- 5724. HICSS, Hawaii, Manoa (2018)
7. Página de PhoneGap, https://phonegap.com (último acceso: 06/07/2018)
8. Página de React Native, http://www.reactnative.com (último acceso: 06/07/2018)
9. Parikh, M., Shetty, M., Divita Vora, P., Koshy, R.: Smart Shopping Using Beacons. International Journal of Current Engineering an Scientific Research 5 (4), 1-6 (2018)
10. Klakegg, S., van Berkel, N., Visuri, A., Huttunen, H.L., Hosio, S., Luo, C., Ferreira, D.: Designing a context-aware assistive infrastructure for elderly care. In: UbiComp '17, pp. 563-568. ACM, New York (2017)
11. Kaulich, T., Heine, T., Kirsch, A.: Indoor Localisation with Beacons for a User-Friendly Mobile Tour Guide. KI-Künstliche Intelligenz 31(3), 239-248 (2017)
12. Repositorio del plugin iBeacon para PhoneGap, https://github.com/petermetz/cordova-plugin-ibeacon (último acceso: 06/07/2018)
13. Repositorio del plugin iBeacon para React Native, https://github.com/MacKentoch/react-native-beacons-manager (último acceso: 06/07/2018)
14. Página de Estimote, https://estimote.com (último acceso: 06/07/2018)
15. Shen, H., Cheng, P.H.: Using area judgment to adjust Indoor Positioning. In: 2017 International Conference on Applied System Innovation, pp. 272-275. IEEE Press, New York (2017)
16. Mohebbi, P., Stroulia, E., Nikolaidis, I.: Indoor Localization: A Cost-Effectiveness vs. Accuracy Study. In: IEEE 30th International Symposium on Computer-Based Medical Systems, pp. 552-557. IEEE Press, New York (2017)
17. Página de iBeacon, https://developer.apple.com/ibeacon (último acceso: 06/07/2018)
18. Página de AltBeacon, https://altbeacon.org (último acceso: 06/07/2018)
19. Página Estimote Monitoring, https://community.estimote.com/hc/en-us/articles/360003252832-What-is-Estimote-Monitoring (último acceso: 06/07/2018)
20. Página de Eddystone, https://developers.google.com/beacons (último acceso: 06/07/2018)
21. Página Estimote Location Protocol, https://community.estimote.com/hc/en-us/articles/208546097-What-is-a-beacon-protocol-Can-beacons-broadcast-multiple-packets-simultaneously (último acceso: 06/07/2018)
22. Página Estimote Nearable Protocol, https://community.estimote.com/hc/en-us/articles/203323543-What-are-Estimote-Stickers- (último acceso: 06/07/2018)

Towards a Framework for Ontology-Based Data Access: Materialization of OWL Ontologies From Relational Databases

Sergio Alejandro Gómez[1,2] and Pablo Rubén Fillottrani[1,2]

[1]Laboratorio de I+D en Ingeniería de Software y Sistemas de Información (LISSI)
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
San Andrés 800 - Campus Palihue – Bahía Blanca, Buenos Aires, Argentina
Email: {sag,prf}@cs.uns.edu.ar
[2]Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC-PBA)

Abstract. The landscape of information systems applications is comprised of legacy components that many times rely on heterogeneous data sources, proprietary data formats, and a low-level knowledge representation. Only seasoned programmers who maintain those systems may interpret these data. Ontology-based Data Access is a novel approach for developing information systems where an ontology defines a high-level global schema of already existing data sources and provides a vocabulary for user queries. In this work, we report on the construction of a system that exports such data, represented as a legacy relational database, as an OWL ontology in accordance to the W3C Direct Mapping specification. We discuss several case studies that show how the proposed application works.

Keywords. Ontology-based data access, Ontology bootstrapping, Web Ontology Language, Relational databases.

1 Introduction

The landscape of information systems applications is comprised of legacy components that many times rely on heterogeneous data sources, often having proprietary data formats, that also are too complex, because of having a low-level, non-uniform knowledge representation whose interpretation is only understood by the database administrator or the seasoned programmers who maintain those systems. Another challenge in modern information systems is given by need of performing the integration of data scattered among many data sources, again in different formats. Therefore, both end-users and new programmers that have to query and maintain those systems face great difficulties when writing and testing new queries (a process that, in the case of having hundreds of tables, can take a time in the order of weeks).

In the last years, Semantic Web technologies in the form of data artifacts such as ontologies with their associated technologies such as ontology editors,

ontology reasoners and programming libraries have reached a maturity degree that is good enough for encouraging its use for solving the aforementioned challenges. Ontology-based Data Access (OBDA) is a novel approach for developing the new generation of information systems in whose paradigm an ontology defines a high-level global schema of already existing data sources and provides a vocabulary for expressing user queries [1]. The two approaches to OBDA include (i) virtualization (in which the datasources are accessed through query rewriting in SQL of a DL query) and (ii) materialization (in which an ontology is populated from the data sources possibly using SQL filters).

In this work, we present a prototypical application for the materialization of OWL ontologies from relational databases. The prototype has been coded in the Java programming language, and accesses an H2 database in order to export its tables in several OWL formats by using the OWL API. Our implementation follows the directives of the direct mapping specification [2] for generating assertional knowledge and the directives in [1] for expressing terminological knowledge. In Sect. 2, we review the main components of the problem of relational databases, ontology reasoning and OBDA. We present the theoretical framework upon which our implementation is based in Sect. 3, formalizing the rules for materialization and discussing implementation details. We review related work in Sect. 4 and finally conclude and suggest possible research avenues for future work in Sect. 5. This work extends state-of-the-art technologies by providing general definitions of how the ontologies have to be built. To the best of our knowledge, the direct mapping specification, although very clear, only provides examples with no formalization whatsoever. We think that this research could be useful for developers of modern information systems where multiple data sources might exist that need to be accessed in unified way with a simple query language (such us SPARQL) or simply using a DL reasoning engine (such us Hermit).

2 Background

Here, we briefly review the concepts of relational databases, ontologies, mappings between databases and ontologies, and query answering in the context of ontology-based data access. We base part of our presentation in [3].

Relational databases. In the relational data model, a database is composed of a schema and an instance, where the schema defines a set of tables with their attribute names and constraints such us primary and foreign keys. The instance populates the tables of the schema with tuples assigning values to the attributes of the tables. As it is customary, the schema of a table T with a primary key k and attributes $a_1, \ldots, a_n$ will be denoted as $T(\underline{k}, a_1, \ldots, a_n)$ and the instance of T is a set of p tuples denoted as $\{(k^1, a_1^1, \ldots, a_n^1), \ldots, (k^p, a_1^p, \ldots, a_n^p)\}$. When necessary, we will refer to the domains (i.e. the set of valid values for the attributes) of $k, a_1, \ldots, a_n$ as $D_k, D_1, \ldots, D_n$, resp.

Ontologies in Description Logics. An ontology is a machine-processable conceptualization of a portion of the world. *Description Logics* (DL) [4] are a well-known

family of knowledge representation formalisms used for describing ontologies. In this work, we will consider a very tight subset of DL, to which we will restrict our discussion, based on the notions of *concepts* (unary predicates or classes) and *roles* (binary predicates or properties). Concept descriptions are built from concept names $C, D, \ldots$ and roles names $p, q, \ldots$ using the constructors conjunction ($C \sqcap D$), disjunction ($C \sqcup D$) and negation ($\neg C$) and inverse role p^-. The empty concept is denoted by $\bot$ and the universal concept is denoted by $\top$. As it is usual, we will abbreviate $\exists p.\top$ by $\exists p$. A DL ontology $\Sigma = (T, A)$ consists of two finite and mutually disjoint sets: a *Tbox* T which introduces the *terminology* and an *Abox* A (assertional box) which contains facts about particular objects in the application domain. The Tbox contains inclusion axioms $C \sqsubseteq D$, where C and D are (possibly complex) concept descriptions, meaning that every individual of C is also a D. In this work, we will restrict ourselves to two forms of axioms, viz. $C \sqsubseteq \exists p$, meaning that every individual of type C is related through role p to some other individual, and $\exists p^- \sqsubseteq C$, meaning that individuals related by role p have to necessarily be related to an individual of a concept C. Objects in the Abox are referred to by a finite number of *individual names* and these names may be used in assertional statements $C(a)$, meaning the individual a is a member of concept C, and $p(a, b)$, meaning that the individual a is related to the individual b by role p.

Web Ontology Language. The Web Ontology Language OWL is a semantic markup language for publishing and sharing ontologies on the World Wide Web [5]. OWL provide syntactic constructors to express classes, properties, individuals, and data values. OWL ontologies can be used along with information written in RDF, and OWL ontologies themselves are primarily exchanged as XML documents. In OWL, objects are represented with Uniform Resource Identifiers, which essentially are generalized web addresses (e.g. `http://foo.org/`) that do not necessarily have to be hosted in some web server. As OWL ontologies are serialized as XML (or N3/Turtle) documents they could be edited with simple text editors if desired. However to provide support for both knowledge engineers and programmers, there are ontology editors such as Protégé and Java libraries such as the OWL API [6].

Mappings. Mappings define how ontological terms are related to terms occurring in the relational schema and are essentially view definitions of the form $\mathit{Class}(f_o(x)) \leftarrow \mathsf{SQL}(x)$, that declare how to populate classes with objects; to populate properties with object-object, of the form $\mathit{objectProperty}(f_o(x), f_o(y)) \leftarrow \mathsf{SQL}(x, y)$, and object-value pairs, of the form $\mathit{dataProperty}(f_o(x), f_d(y)) \leftarrow \mathsf{SQL}(x, y)$, where $\mathsf{SQL}(\cdot)$ and $\mathsf{SQL}(\cdot, \cdot)$ are SQL queries with one and two output variables, resp., and $f_o(\cdot)$ and $f_v(\cdot)$ are functions that cast values returned by the SQL queries into objects (i.e. URIs and literals). Classes are populated with URIs $f_o(x)$ computed from the values x returned by $\mathsf{SQL}(x)$. Properties can relate two objects or assign a value to an object; in the former case, with pairs of objects $f_o(x)$ and $f_o(y)$ and in the latter case assigning the value $f_v(y)$ to the object $f_o(x)$ when the query $\mathsf{SQL}(x, y)$ is computed.

Query answering. Given a data access instance $(\mathcal{D}, \mathcal{V}, \mathcal{O}, \mathcal{M})$, where $\mathcal{D}$ is a relational database, $\mathcal{V}$ is an ontological vocabulary, $\mathcal{O}$ is a set of ontological axioms over $\mathcal{V}$, and $\mathcal{M}$ is a set of mappings between $\mathcal{V}$ and $\mathcal{D}$. There are two approaches to answer a query Q over $\mathcal{V}$: (i) *materialization*: ontological facts are materialized (i.e. classes and properties participating in mappings are populated with individuals by evaluating SQL queries participating in mappings) and this gives a set of ontological facts $\mathcal{A}$ and then Q is evaluated against $\mathcal{O}$ and $\mathcal{A}$ with standard query-answering engines for ontologies, or (ii) *virtualization*: Q should be first rewritten into SQL using $\mathcal{O}$ and $\mathcal{M}$ and then SQL should be executed over $\mathcal{D}$. In this work, we will use the materialization approach.

3 Materialization of OWL Ontologies from Relational Databases

Materializing an OWL ontology from a relational database requires exporting the database contents as a text file in OWL format. For doing this, we need to export the schema information of each table as Tbox axioms and the instance data of the tables as Abox assertions. Here we formalize the process of exporting a single table having a single primary key, two tables participating in a one-to-many relationship, three tables in a many-to-many relationship and a table without a primary key. After that, we report implementation details.

3.1 Rules for Exporting Tables as Ontologies

We implemented the direct mapping specification from relational databases to OWL according the directions given by [1, 2]. Building an ontology from a database requires creating at least a class C_T for every table T, and for every attribute a of domain d in T we need two inclusion axioms $C_T \sqsubseteq \exists a$ and $\exists a^- \sqsubseteq d$. Primary key values k_i serve the purpose of establishing the membership of individuals to classes as Abox asssertions of the form $C_T(C\#k^j)$. For indicating that a^j is the value of attribute a, we will use a role expression of the form $C_T\#a(C_T\#k^j, C_T\#a^j)$. When it is clear from context, we might drop the prefix $C_T\#$ for simplifying our notation. A foreign key fk in table T_1 referencing a primary key field in table T_2 will also require to add two Tbox axioms $C_{T_1} \sqsubseteq \exists \mathsf{ref}_fk$ and $\exists \mathsf{ref}_fk^- \sqsubseteq C_{T_2}$ and an Abox assertion $\mathsf{ref}_fk(k^j, fk^t)$ for expressing that the individual named k^j in C_{T_1} is related to the individual named fk^t in C_{T_2}. Besides, in any case, if we want to consider a subset of a table for its mapping into an ontology, we might define an SQL query that will act as an SQL filter.

We will present how to map a table with a single primary key, two tables participating in a one-to-many relationship, three tables participating in a many-to-many relationship, and a table non-having a primary key. For each case, we give a formal specification of how the mapping must be done and present a concrete example. In all cases, the base URI for the ontology will be `http://foo.org`.

Definition 1 (Mapping of a table with a single primary key). *Let T be a table with schema $T(\underline{k}, a_1, \ldots, a_n)$ and instance $\{(k^1, a_1^1, \ldots, a_n^1), \ldots, (k^m, a_1^m, \ldots, a_n^m)\}$. To map T into DL, we have to create a class T and for each attribute a_i of domain D_i we have to add two axioms: $T \sqsubseteq \exists a_i$, indicating that every T has an attribute a_i, and $\exists a_i^- \sqsubseteq D_i$, meaning that the type of a_i is D_i. The assertional box for T will contain $\{C(k^1), \ldots, C(k^m)\}$. For every attribute a_i of the schema and instance value a_i^j, produce a property $a_i(k^j, a_i^j)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.*

Example 1. Let us consider a table for representing cars with the following schema: *Car(licensePlate, brand, model, weight, preowned, dateOfPurchase, numberOfDoors)*, populated with:

licensePlate	*brand*	*model*	*weight*	*preowned*	*dateOfPurchase*	*numberOfDoors*
ABC123	Chevrolet	Corsa	1000.1	FALSE	2010-10-01	4
CDE456	VW	Suran	1000.4	TRUE	2013-10-01	5

The OWL file generated by the system from the URI "http://foo.org" and SQL filter select * from Car where dateOfPurchase <= 2011-12-31. Then the name of the class would be *Car*. For example, for the attribute *brand* of type *string*, we will have two terminological axioms $Car \sqsubseteq \exists brand$ and $\exists brand^- \sqsubseteq String$. For saying that $ABC123$ is a car, we will have an assertion $Car(ABC123)$ and for establishing that the brand of $ABC123$ is Chevrolet, we will have an assertion $brand(ABC123, Chevrolet)$. Likewise, for saying that cars have an integer number of doors and that $ABC123$ has 4 doors, we will have $Car \sqsubseteq \exists numberOfDoors$, $\exists numberOfDoors^- \sqsubseteq Integer$ and $numberOfDoors(ABC123, 4)$. These DL assertions are expressed in OWL as depicted in Fig. 1.

```
<owl:Class rdf:about="http://foo.org#Car"/>

<owl:DatatypeProperty rdf:about="http://foo.org/Car#brand">
   <rdfs:domain rdf:resource="http://foo.org#Car"/>
   <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:about="http://foo.org/Car#numberOfDoors">
   <rdfs:domain rdf:resource="http://foo.org#Car"/>
   <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
</owl:DatatypeProperty>
...
<owl:NamedIndividual rdf:about="http://foo.org/Car/licenseplate=ABC123">
   <rdf:type rdf:resource="http://foo.org#Car"/>
   <Car:brand rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Chevrolet</Car:brand>
   ...
   <Car:numberOfDoors rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">4
   </Car:numberOfDoors>
    ...
</owl:NamedIndividual>
```

Fig. 1. OWL code for part of the *Car* class from Ex. 1

We now present how to map two tables participating in a one-to-many relationship.

Definition 2 (Mapping of a one-to-many relationship). *Let* $A(\underline{k_1}, a_1, \ldots, a_n)$ *and* $B(\underline{k_2}, b_1, \ldots, b_m, k_1)$ *be two tables participating in a one-to-many relationship where* k_1 *is both the primary key in* A *and a foreign key in* B*. Tables* A *and* B *are translated in DL according to Def. 1. Besides, the two axioms are added:* $B \sqsubseteq \exists \mathsf{ref}_k_1$ *and* $\exists \mathsf{ref}_k_1^- \sqsubseteq A$*. And for every tuple* $(k_1^i, a_1^i, \ldots, a_n^i)$ *of* A *related to a tuple* $(k_2^j, b_1^j, \ldots, b_m^j, k_1^i)$ *in* B*, an assertion* $\mathsf{ref}_k_1(k_2^j, k_1^i)$ *is added.*

Example 2 (Continues Ex. 1). Consider a one-to-many relation of table *Car* from Ex. 1 with a table $AutoPart(\underline{autoPartID}, name, carID)$, populated with:

autoPartID	*name*	*carID*
1	engine	ABC123
2	windshield	ABC123

The following assertions are added: $autoPartID(1)$, $name(1, engine)$, $carID(1, ABC123)$, $\mathsf{ref}_carID(1, ABC123)$, $autoPartID(2)$, $name(2, windshield)$, $carID(2, ABC123)$, $\mathsf{ref}_carID(2, ABC123)$. Notice that we actually abuse notation in $carID(1, ABC123)$ as the second argument is an string literal while in $\mathsf{ref}_carID(1, ABC123)$ it stands for an object identifier. See Fig. 2 to see how the assertions for the first tuple are codified in OWL.

```
<owl:Class rdf:about="http://foo.org#AutoPart"/>
...
<owl:ObjectProperty rdf:about="http://foo.org/AutoPart#ref-carID"/>
...
<owl:NamedIndividual rdf:about="http://foo.org/AutoPart/autopartid=1">
   <rdf:type rdf:resource="http://foo.org#AutoPart"/>
   <AutoPart:ref-carID rdf:resource="http://foo.org/Car/licensePlate=ABC123"/>
   <AutoPart:autoPartID rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">1
    </AutoPart:autoPartID>
   <AutoPart:carID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">ABC123
    </AutoPart:carID>
   <AutoPart:name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">engine
    </AutoPart:name>
</owl:NamedIndividual>
...
```

Fig. 2. Part of the OWL code for *AutoPart* class from Ex. 2

We now address the issue of mapping two tables in a many-to-many relationship. As it is customary in the relational data model, the relationship is reified an intermediate table containing the primary keys of the tables being related. We motivate the concept of *composite property name* for expressing tables having primary keys comprised of more than one key field.

Definition 3 (Mapping of a binary many-to-many relationship). *Let* $A(\underline{k_1}, a_1, \ldots, a_n)$ *and* $B(\underline{k_2}, b_1, \ldots, b_m)$ *be two tables in a many-to-many relationship* R *reified as* $R(\underline{k_1}, \underline{k_2}, r_1, \ldots, r_l)$*. Tables* A *and* B *are expressed in DL according to Def. 1. The primary key of table* R *is expressed as a* composite property name $k_1 \circ k_2$*. The axiom* $R \sqsubseteq \exists(k_1 \circ k_2)$ *is added as well as two axioms for* k_1 *and* k_2*:* $\exists \mathsf{ref}_k_1 \sqsubseteq A$ *and* $\exists \mathsf{ref}_k_2 \sqsubseteq B$*.*

Example 3 (Continues Ex. 2). Consider another table *Person* for representing people and yet another one for representing what person drives what car of table *Car* from Ex. 1, with schemas $Person(\underline{identification}, name)$ and $Drives(\underline{personID}, \underline{carID}, lastTime)$, resp., and populated with:

Person

identification	*name*
10	John

Drives

personID	*carID*	*lastTime*
10	ABC123	2018-06-10

The following axioms have to be added to the Tbox for representing people: $Person \sqsubseteq \exists identification$, $Person \sqsubseteq \exists name$, $\exists identification^- \sqsubseteq String$, and $\exists name^- \sqsubseteq String$. For representing who drives what car, we have to add: $Drives \sqsubseteq \exists personID$, $Drives \sqsubseteq \exists carID$, $Drives \sqsubseteq \exists lastTime$, $\exists personID^- \sqsubseteq String$, $\exists carID^- \sqsubseteq String$, and $\exists lastTime^- \sqsubseteq Date$. Finally, for defining the composite property built up of the concatenation of the primary key fields of the *Drives* table, the following axioms are added: $Drives \sqsubseteq \exists(personID \circ carID)$. For modeling the instances of table *Person*, the following assertions are added to the ontology: $Person(10)$, $identification(10, 10)$, $name(10, john)$. And for the table *Drives*, whose only tuple has the composite primary key "$10 \circ ABC123$", the following assertions are included in the ontology: $Drives(10 \circ ABC123)$, $personID(10 \circ ABC123, 10)$, $carID(10 \circ ABC123, ABC123)$, $\mathsf{ref}_personID(10 \circ ABC123, 10)$, $\mathsf{ref}_carID(10 \circ ABC123, ABC123)$, and $lastTime(10 \circ ABC123$, 2018-06-10). Part of the OWL code for modeling this situation is presented in Fig. 3.

```
<owl:Class rdf:about="http://foo.org#Drives"/>
...
// Object Properties
<owl:ObjectProperty rdf:about="http://foo.org/Drives#ref-carID"/>
<owl:ObjectProperty rdf:about="http://foo.org/Drives#ref-personID"/>
...
<owl:NamedIndividual rdf:about="http://foo.org/carID=ABC123;personID=10">
  <rdf:type rdf:resource="http://foo.org#Drives"/>
  <Drives:ref-carID rdf:resource="http://foo.org/Car/licensePlate=ABC123"/>
  <Drives:ref-personID rdf:resource="http://foo.org/Person/identification=10"/>
  <Drives:carID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">ABC123</Drives:carID>
  <Drives:lastTime rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
      2018-06-10T00:00:00</Drives:lastTime>
</owl:NamedIndividual>
```

Fig. 3. OWL code for part of the representation of table *Drives* from Ex. 3

In the case of tables non-having a primary key, the direct mapping specification requires to build blank nodes as the individual identifiers for each row of the table, we take a simpler approach by defining a *surrogate primary key* with auto-incrementing values.

Definition 4 (Mapping of tables non-having a key). *Let* $T(a_1, \ldots, a_n)$ *be a table, we will interpret it as if it was a table* $T(\underline{blankkey}, a_1, \ldots, a_n)$ *where*

blankkey is called a surrogate primary key *and an instance of* T $\{(a_1^1, \ldots, a_n^1), \ldots, (a_1^m, \ldots, a_n^m)\}$ *will be interpreted as* $\{(1, a_1^1, \ldots, a_n^1), \ldots, (m, a_1^m, \ldots, a_n^m)\}$.

Example 4. Consider a table *Measurement*(*time*, *temperature*) populated with:

time	*temperature*
2018-07-03T12:00:00	23.4
2018-07-03T13:00:00	25.6

Notice that in this case, we could have had *time* defined as the primary key of the table but it could have also be the case of having several measurements recorded by different instruments and that situation would have made that option infeasible. The Tbox is then composed by: $Measurement \sqsubseteq \exists blankkey$, $\exists blankkey^- \sqsubseteq Integer$, $Measurement \sqsubseteq \exists time$, $\exists time^- \sqsubseteq TimeStamp$, $Measurement \sqsubseteq \exists temperature$ and $\exists temperature^- \sqsubseteq Double$. From the table's contents, we obtain the following assertions: $Measurement(1)$, $time(1, 2018\text{-}07\text{-}03\ 12{:}00{:}00.0)$, $temperature(1, 23.4)$, $Measurement(2)$, $time(2, 2018\text{-}07\text{-}03\ 13{:}00{:}00.0)$ and $temperature(2, 25.6)$. Part of the OWL code for modeling this situation can be seen in Fig. 4.

```
<owl:NamedIndividual rdf:about="http://foo.org/Measurement/BLANKKEY=1">
  <rdf:type rdf:resource="http://foo.org#Measurement"/>
  <Measurement:temperature rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
     23.4</Measurement:temperature>
  <Measurement:time rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTimeStamp">
     2018-07-03 12:00:00.0</Measurement:time>
</owl:NamedIndividual>

<owl:NamedIndividual rdf:about="http://foo.org/Measurement/BLANKKEY=2">
  <rdf:type rdf:resource="http://foo.org#Measurement"/>
  <Measurement:temperature rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
     25.6</Measurement:temperature>
  <Measurement:time rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTimeStamp">
     2018-07-03 13:00:00.0</Measurement:time>
</owl:NamedIndividual>
```

Fig. 4. Individuals of the *Measurement* class of Ex. 4

3.2 Implementation Details

We now discuss implementation details of the application that we developed for bootstrapping ontology contents from a relational database. The application in its current state can process relational databases in H2 format. The access to a certain database is performed via JDBC, where the system, after establishing a connection with user provided parameters such as name of the database, user ID and password, can automatically obtain the database schema. OWL code generation is done via the OWL API [6]. In this option, the user can specify a base URI, the name of the table to be mapped as an ontology, an SQL filter for selecting a subset of the table's records, and can see the schema information of the table (e.g. attributes, primary and foreing keys). The application allows to

generate the ontologies in two OWL syntaxes, viz. the XML syntax as shown in Ex. 1–4 and the Turtle (i.e. Terse RDF Triple Language) syntax [7]. There is another option for generating an ontology for the contents of the entire database (where the user has to specify the name of the ontology and a base URI). An executable JAR file and a sample database with the contents of Ex. 1–4 can be downloaded from `http://cs.uns.edu.ar/~sag/obda-test`.

4 Related Work

We now discuss related work. MASTRO [8] is an Ontology-Based Data Access (OBDA) management system. Ontologies in MASTRO are specified through languages belonging to the DL-Lite family of lightweight DLs. The ontology is connected to external relational data management or data federation systems through a mapping establishing a semantic relation between SQL queries issued over the underlying databases and elements of the ontology. To access data, users can specify SPARQL queries over the ontology and make use of the query answering services provided by Mastro. Mastro is developed in Java. Our implementation is also implemented in Java what expedites its interaction with OWL API. MASTRO represents ontologies in OWL Lite, a dialect of OWL that, despite its efficiency, imposes too much restrictions on its knowledge representation capabilities. We aim towards more rich and flexible dialects of OWL that will eventually allow us to model richer data features such as the cardinality of relations.

BOOTOX [3] is a system that aims at facilitating ontology and mapping development by their automatic extraction (i.e., bootstrapping) from relational databases. It allows to control the OWL 2 profile of the output ontologies, bootstrap complex and provenance mappings, which are beyond the W3C direct mapping specification, and also it allows to import pre-existing ontologies via alignment. Our implementation exports the ontologies in several formats of OWL (viz., Turtle and XML) though we have not yet committed to specific OWL profiles because we plan on using standard DL reasoners (e.g., Hermit or Pellet).

RODI [9] is a benchmark for the evaluation of the practical utility of ontology-based data integration systems and their application in practice. RODI includes test scenarios from the domains of scientific conferences, geographical data, and oil and gas exploration. Scenarios are constituted of databases, ontologies, and queries to test expected results. Systems that compute relational-to-ontology mappings can be evaluated using RODI by checking how well they can handle various features of relational schemas and ontologies, and how well the computed mappings work for query answering. We plan on using RODI for extending the evaluation of our implementation.

5 Conclusions and Future Work

We have presented an application in a prototypical status for performing OBDA by allowing the user to materialize the contents of an H2 relational database

as an OWL ontology. Our main contribution is a clear and precise description of the mappings to guarantee that the database data are well interpreted as ontological data. Our application is programmed in Java and uses JBDC to access the database and the OWL API for generating the ontology. We showed how our application is capable of handling the cases of exporting a single table with or without a key field, two tables in a one-many-relationship and three tables implementing a many-to-many relationship. As future work, we plan to perform an experimental evaluation of its performance on different databases. We also plan to add more features, e.g. mapping generation, SQL unfolding and supporting more database formats, this will allow to include more precise user-defined specifications of each mapping such as implicit information present in neither the database's schema nor the instance but in the application usage patterns of the database.

Acknowledgments. This research is funded by Secretaría General de Ciencia y Técnica, Universidad Nacional del Sur, Argentina and by Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC-PBA).

References

1. Kontchakov, R., Rodríguez-Muro, M., Zakharyaschev, M.: Ontology-Based Data Access with Databases: A Short Course. In: Reasoning Web: Semantic Technologies for Intelligent Data Access of the LNCS. Volume 8067. Springer (2013) 194–229
2. Arenas, M., Bertails, A., Prud'hommeaux, E., Sequeda, J.: A Direct Mapping of Relational Data to RDF. W3C Recommendation 27 September 2012 (2012)
3. Jimenez-Ruiz, E., Kharlamov, E., Zheleznyakov, D., Horrocks, I., Pinkel, C., Skjæveland, M.G., Thorstensen, E., Mora, J.: BootOX: Practical Mapping of RDBs to OWL 2. In: the 14th International Semantic Web Conference. (2015) 113–132
4. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P., eds.: The Description Logic Handbook – Theory, Implementation and Applications. Cambridge University Press (2003)
5. Bao, J., Kendall, E.F., McGuinness, D.L., Patel-Schneider, P.F.: OWL 2 Web Ontology Language Quick Reference Guide (Second Edition) W3C Recommendation 11 December 2012 (2012)
6. Horridge, M., Bechhofer, S.: The OWL API: A Java API for OWL Ontologies. Semantic Web **2**(1) (2011) 11–21
7. Becket, D., Berners-Lee, T., Prud'hommeaux, E., Carothers, G.: RDF 1.1 Turtle. Terse RDF Triple Language. W3C Recommendation 25 february 2014 (2014)
8. Calvanese, D., Giacomo, G.D., Lembo, D., Savo, D.F.: The MASTRO system for ontology-based data access. Semantic Web **2**(1) (2011) 43–53
9. Pinkel, C., Binnig, C., Jiménez-Ruiz, E., Kharlamov, E., May, W., Nikolov, A., Bastinos, A.S., Skjæveland, M.G., Solimando, A., Taheriyan, M., Heupel, C., Horrocks, I.: RODI: Benchmarking Relational-to-Ontology Mapping Generation Quality. Semantic Web (2016) 1–26

A Graphical Web Tool with DL-based Reasoning Support over Orthogonal Variability Models

Angela Oyarzun[1] and Germán Braun[1,3], Laura Cecchi[1], and Pablo Fillottrani[2,4]

[1]UNIVERSIDAD NACIONAL DEL COMAHUE [2]UNIVERSIDAD NACIONAL DEL SUR
[3]*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)*
[4]*Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)*

Abstract Variability Management is one of the most challenging tasks in a Software Product Line (SPL) development. This is reflected in the way software is developed, maintained and extended. Therefore, automatic variability analysis has emerged in order to validate models in early development stages, avoiding affecting derived products quality. In this work, we present *crowd-variability*, a novel graphical tool designed for modelling and validating Orthogonal Variability Models (OVM) using Description Logics (DL)-based reasoning services. We describe the tool and demonstrate the usage of the first prototype along with examples of use. Currently, we are working to release the first beta version of *crowd-variability*.

Keywords: Software Product Lines, Orthogonal Variability Models, Description Logics, Graphical tools for modeling variability

1 Introduction

Software Product Lines (SPL) development involves one of the most challenging and important activities of the domain and application engineering phases, Variability Management [1]. During the first phase, variability is defined, modelled, implemented and validate, while in the second phase, such variability is instantiated in order to derive new products. In this context, Variability Management has a great impact on the way software is developed, extended and maintained. Orthogonal Variability Models (OVM) are used to define variability and relate it to more traditional models like class diagrams, use cases, among others. OVM validation plays a decisive role in detecting faults in a SPL in early development stages that, on the contrary, can limit the quality of the products derived. Due to this fact, automated variability analysis [2, 3, 4, 5] has emerged, which focuses on a set of techniques for translating and validating variability models by considering the anomalies or mismatches these models might contain. Therefore, checking consistency of variability models is a critical problem.

There are several works about automated variability analysis that propose different techniques and methods [5, 6], many of them based on Feature Model analysis, which is another language for variability. However, approaches for validating OVM diagrams have also been undertaken. In this sense, FaMa-OVM [6] is a tool that works with OVM specified using a textual format that are later analysed by SAT-solvers, but as

well as different SAT-based approaches, it presents certain limitations related to more restrictive logics and thus fail to reflect the finer logical structure of variability models.

Therefore, our group presented a proposal [7] focusing on the use of Description Logics (DL) [8] for tackling this issue. DL are a family of decidable logics that have been widely proven to give support to software engineering in order to improve the automated analysis. Our encoding is provided on an OVM-based language from a previously defined SeVaTax framework [9], aiming at analysing variability models properties and deriving products from an SPL. Thus, its models are formalised in DL $\mathcal{ALCI}$ [10] guaranteeing EXPTIME reasoning on SeVaTax models.

In this work, we present *crowd-variability*, a graphical tool for designing, visualising and checking consistency of OVM diagrams. This novel client-server tool provides graphical support for users modelling their diagrams and is integrated with automatic DL-based reasoning [11], using OWLlink [12] as communication protocol for off-the-shelf reasoning systems. Thus, users will visualise their OVMs in an on-line manner while are being modelled and edited. A first Web prototype[1] of this tool already runs on the client-server architecture, supports OWLlink and enables the satisfiability checking process on OVM graphical diagrams. Such client-server architecture is based on an ontology engineering environment, named *crowd*[2] [13, 14]. *crowd* is a multi-views software for graphical editing of ontologies, using standard languages such as UML, EER and ORM, and DL reasoning in order to give support to ontology engineering tasks.

This work is structured as follows. Section 2 introduces orthogonal variability models and describes each of their components. Section 3 illustrates and details the architecture of this tool. Section 4 presents the first prototype developed together with a simple example of use. Section 5 exposes a preliminary evaluation and discussions. To conclude the paper, section 6 elaborates on final considerations and directions for future works.

2 Orthogonal Variability Models

Variability identifies and models variable characteristics of the products derived from Software Product Lines [1]. Products are applications generated by properly selecting services from a variability model and its constraints. Variability can be modelled as part of traditional models such as class diagrams, use cases, feature models or it can be represented by a separated model. In this regard, OVMs have emerged, which are in charge of defining variability of a SPL, through a separated model and relating it to other diagrams.

OVMs consist of two fundamental elements: variation points and variants. These elements relate to each other through variability dependencies and/or constraint dependencies. Both elements and interactions are depicted in Table 1.

Elements:

– ***Variation Point***: it represents a variable object in real world.

[1] `http://crowd.fi.uncoma.edu.ar/ovm/`
[2] `http://crowd.fi.uncoma.edu.ar`

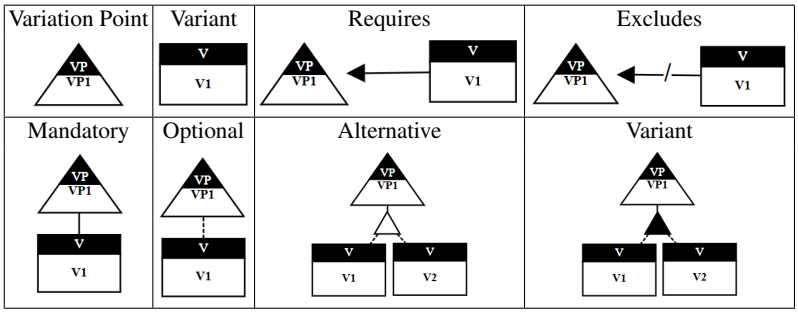


Table 1. OVM Graphical Components

– ***Variant***: it denotes how a variation point varies.

Constraint Dependencies:

– ***Requires***: it describes a relationship between variation points, variants or a variant and a variation point where the selection of an element requires the selection of the associated element to the first one.
– ***Excludes***: it describes a relationship between variation points, variants or a variant and a variation point where the selection of an element excludes the selection of the associated element to the first one.

Variability Dependencies:

– ***Mandatory***: it defines the selection of a variation point requires the selection of the associated variants.
– ***Optional***: it determines the selection of a variation point can require, but not necessarily, the selection of the associated variants.
– ***Alternative***: it specifies the selection of a variation point requires the selection of only one of the associated variants.
– ***Variant***: it defines the selection of a variation point requires the selection of at least one of the associated variants.

Example 1. Fig.1 depicts mandatory and alternative variability dependencies modelling the variability of a Remote Robotics Laboratory(RRL), which is defined through a programming language(PL) and the type of robot (R): Frankestito[15, 16] or Multiplo N6 Max[17]. The RRL variation point allows configuring four possible products: {RRL, R, Multiplo N6 Max, PL, Python}, {RRL, R, Multiplo N6 Max, PL, Blockly}, {RRL,R, Frankestito, PL, Python} and {RRL, R, Frankestito, PL, Blockly}. Therefore, as at least one product can be derived, this model is considered *consistent* [7].

3 *crowd-variability* Overview

crowd-variability is a graphical modelling tool being supported by Universidad Nacional del Comahue and Universidad Nacional del Sur of Argentina. Its main purposes are to

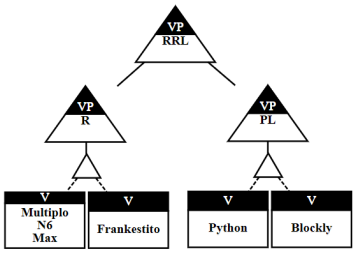


Figure 1. OVM example for a Remote Robotics Laboratory (RRL) modeled in *crowd-variability* tool.

enable users to design and visualise OVM diagrams, as well as validating them. This tool employs complete logical reasoning in order to verify the satisfiability of specifications. So as to apply automated analysis, a semantic definition of all elements of an OVM diagram must be provided. Therefore, each variant, variation point, variability dependency and constraint dependency is translated into a logic-based formalism. Finally, this automatic variability analysis is possible due to the fact that the tool is fully integrated with a powerful logic-based reasoning system acting as a background inference engine.

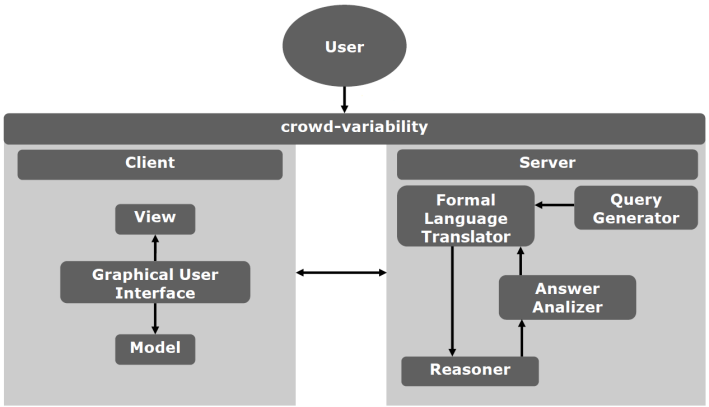


Figure 2. An overview of the *crowd-variability* client-server architecture.

An overview of the client-server architecture is shown in Figure 2. On the client side, users are able to create and edit their models through a graphical user interface (GUI). It offers a set of graphical primitives of the OVM language and functions required for manipulating these diagrams. The server component consists of modules in charge of translating graphical models into a formal language, generating queries to the reasoner, reason over OVM and processing the reasoner output. We explore each com-

ponent in turn.

Client. The Client side is composed of a View, a Model and a Graphical User Interface (GUI). The first two, View and Model are in constant interaction between them. The **View** possesses the visual representation of the OVM diagrams, widgets and events for handling them, while the **Model** contains an abstract representation of each primitive of the OVM language and is able to answer diverse methods that affect it. The corresponding models and its components are created by the external graphical library JointJS[3], which is a powerful open source JavaScript. JointJS has been chosen among others because it offers a wide variety of functionalities and focuses on the model that represents each element in user diagrams and the responses to actions on the view. Moreover, it facilitates diagrams translation into a JSON format, uses Backbone 2[4] to give structure to Web Applications by providing a model-view-controller architecture and it is capable of expanding its functions by creating and/or adding your own plug-ins.

On the other hand, the *crowd-variability* **Graphical User Interface** is an user-interface mashup [18], which is based on widgets aggregating sets of functionalities in a common graphical space. It is associated to a set of cross-view widgets, such as tool-bars and user management, and a set of specific-view ones, icons, among others. These widgets are orchestrated by the system, through event notifications generated by each one as the user interacts. Last but not least, the user interface component also provide us with functionalities such as user login, model store and interoperability with other tools through JSON importing and exporting in addition to OWLlink files.

Server. The Server side comprises a Formal Language Translator, a Query Generator, an Answer Analyzer and a Reasoner. The **Formal Language Translator** module takes a JSON representation of an user's model and translates it into a formal language and write it to an OWLlink file, which is later sent to the reasoner together with a set of queries. This queries are appended to the file after being generated by a **Query Generator** module, so that the reasoning system can determine the consistency of the model provided by the user. Afterwards, **Reasoner** module connects to an off-the-shelf reasoning system, p.e. Racer [19], for processing the OWLlink document and returns a new OWLlink document including the responses to these queries. Finally, the **Answer Analyzer** processes the reasoner output in order to give the conclusions to users. The output is a new JSON object indicating the consistency of the whole model and each one of this concepts.

4 First Prototype and Example of Use

The client side of *crowd-variability* first prototype runs in a Web browser and has been developed using CoffeeScript, a programming language that compiles into JavaScript, PHP and JointJS library, which has been used to build a new plug-in for OVM. The server side runs in an Apache server and has been developed using PHP. The prototype offers users the possibility of graphically visualising their OVM diagrams, while they are being designed and edited. It also supports the complete set of OVM primitives as

[3] https://www.jointjs.com/opensource

[4] `http://backbonejs.org/`

described in section 2. Once built the model, users may request its consistency evaluation. At the moment, *crowd variability* prototype allows OVM diagrams translation into $\mathcal{ALCI}$ DL. Currently, the integrated reasoner in this tool is Racer[11]. Table 2 shows some variability dependencies, its associated graphics and its $\mathcal{ALCI}$ DL encoding. For more details refer to [7].

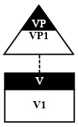
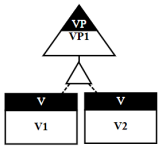
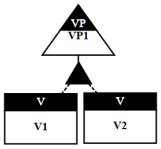
Expression	crowd-variability graphic	$\mathcal{ALCI}$ DL encoding
Optional Variability Dependency		$\exists vp1v1^{=1} \sqsubseteq VP1$ $V1 \sqsubseteq \leq 1vp1v1^{-} \sqcap \geq 1vp1v1^{-}$
Alternative Variability Dependency		$VP1 \sqsubseteq \exists vp1v1^{=1} \sqcap \exists vp1v2^{=1}$ $\bot \sqsubseteq \exists vp1v1 \sqcap \exists vp1v2$ $V1 \sqsubseteq \leq 1vp1v1^{-} \sqcap \geq 1vp1v1^{-}$ $V2 \sqsubseteq \leq 1vp1v2^{-} \sqcap \geq 1vp1v2^{-}$
Variant Variability Dependency		$VP1 \sqsubseteq \exists vp1v1^{=1} \sqcup \exists vp1v2^{=1}$ $V1 \sqsubseteq \leq 1vp1v1^{-} \sqcap \geq 1vp1v1^{-}$ $V2 \sqsubseteq \leq 1vp1v2^{-} \sqcap \geq 1vp1v2^{-}$

Table 2. OVM Variability Dependencies

Encoding has suffered some changes from the original one because of pragmatic reasons. In [7], each variant and variation point are represented by singleton concepts and dependencies are represented by roles. However, it does not present any axiom to ensure the singleton property of the classes. Therefore, in order to accomplish that, we have modified the codification by adding cardinalities.

The JSON object to be sent to the reasoner contains information about each element in the diagram. JSON object also includes data about the diagram itself together with options to manage the visualisation of the primitives. For variants and variation points, it records its name and position. For constraint and variability dependencies, it keeps information about its name, origin, targets and type of dependency. In particular, for alternative and variant variability dependencies, it registers its position. Fig. 4 presents the JSON representation and Fig. 5 the respective DL encoding in OWLlink format for the model in Fig. 3.

The *crowd-variability* Editor works on diagrams, which may contain only one OVM diagram. The diagrams are referred as models. Only one diagram is visible at a time and the editing of each of them is independent. The tool does not implement special visual techniques for handling very large diagrams yet.

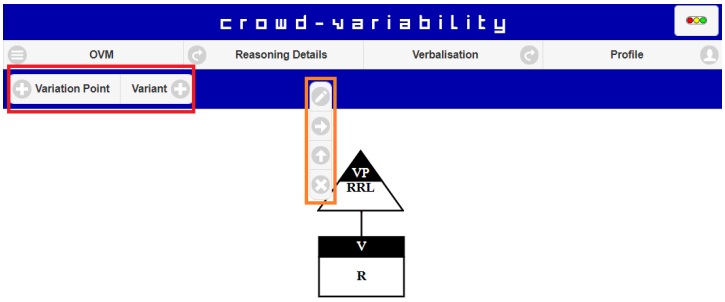


Figure 3. *Look and feel* of a mandatory dependency example in *crowd-variability*.


```
{"metadata":[{"tool":"crowd - variability",
              "owner":"GILIA - Universidad Nacional del Comahue"}],
 "graphops":[{"language":"ovm","version":"","displayRequires":"true",
              "displayExcludes":"true","displayMandatory":"true",
              "displayAlternative":"true","displayOptional":"true","displayVariant":"true"}],
 "variation_points":[{"name":"RRL","position":{"x":462,"y":74}}],
 "variants":[{"name":"R","position":{"x":365,"y":109}}],
 "constraint_dependencies":[],
 "variability_dependencies":[{"name":"m1","origin":"RRL","target":["R"],"type":"mandatory"}]}
```


Figure 4. *crowd-variability* JSON representation of model in Figure 3

.

```
//Mandatory Dependency Domain
//∃rrlr ⊑ RRL
<owl:SubClassOf>
  <owl:ObjectSomeValuesFrom>
    <owl:ObjectProperty IRI="rrlr"/>
    <owl:Class abbreviatedIRI="owl:Thing"/>
  </owl:ObjectSomeValuesFrom>
  <owl:Class IRI="rrl"/>
</owl:SubClassOf>
```

```
//RRL ⊑ ∃rrlr^=1
<owl:SubClassOf>
  <owl:Class IRI="rrl"/>
  <owl:ObjectIntersectionOf>
    <owl:ObjectSomeValuesFrom>
      <owl:ObjectProperty IRI="rrlr"/>
      <owl:Class IRI="owl:Thing"/>
    </owl:ObjectSomeValuesFrom>
    <owl:ObjectMaxCardinality cardinality="1">
      <owl:ObjectProperty IRI="rrlr"/>
    </owl:ObjectMaxCardinality>
  </owl:ObjectIntersectionOf>
</owl:SubClassOf>
```

```
//Mandatory Dependency Range
//∃rrlr⁻ ⊑ R
<owl:SubClassOf>
  <owl:ObjectSomeValuesFrom>
    <owl:ObjectInverseOf>
      <owl:ObjectProperty IRI="rrlr"/>
    </owl:ObjectInverseOf>
    <owl:Class abbreviatedIRI="owl:Thing"/>
  </owl:ObjectSomeValuesFrom>
  <owl:Class IRI="r"/>
</owl:SubClassOf>
```

```
//R ⊑≥ 1rrlr⁻ ⊓ ≤ 1rrlr⁻
<owl:Class IRI="r"/>
  <owl:ObjectIntersectionOf>
    <owl:ObjectMaxCardinality cardinality="1">
      <owl:ObjectInverseOf>
        <owl:ObjectProperty IRI="rrlr"/>
      </owl:ObjectInverseOf>
    </owl:ObjectMaxCardinality>
    <owl:ObjectMinCardinality cardinality="1">
      <owl:ObjectInverseOf>
        <owl:ObjectProperty IRI="rrlr"/>
      </owl:ObjectInverseOf>
    </owl:ObjectMinCardinality>
  </owl:ObjectIntersectionOf>
</owl:SubClassOf>
```

Figure 5. *crowd-variability* OWLlink representation of mandatory dependency in Figure 3

.

For depicting the current *look and feel* of *crowd-variability* editor, Fig. 3 models a mandatory variability dependency between a variation point RRL and a variant R. Both elements were created by pressing the buttons highlighted in red. The GUI includes tool-bars for variation points and variants that presents users' options to change their name(pencil button), add the respective constraint(right arrow button) or variability(up arrow button) dependencies and delete them(cross button). In particular, the dependency in Fig. 3 has been created by pressing the button with an up arrow of the RRL's tool-bar colored in orange, as variants do not have the option of creating variability dependencies.

5 Preliminary Evaluation

crowd-variability has been designed as an extensible and sustainable architecture. It has been developed using expansible graphical libraries and Web technologies. Therefore, it can be extended to other variability modelling proposals such as Feature Models (FM), Common Variability Language (CVL) and SeVaTax. These proposals have a similar expressiveness, however, there are great differences between them. Feature Models model variability through a hierarchical set of features and its relationships [20]. Common Variability Language is a domain-independent language for specifying and resolving variability [21]. SeVaTax is an extension of the orthogonal variability models [9].

From a graphical point of view, JointJS graphics library offers the possibility of expanding its functionalities by creating or/and adding new plug-ins, that specify each primitive of new variability modelling languages. Furthermore, new formalisation methods, like the one for OVM diagrams presented in [7], can be incorporated.

Currently, this tool is integrated with Racer reasoning system. Nonetheless, new back-end reasoners can be connected to the tool.

Hence, a relevant feature of *crowd variability* is to consider users' preferences and usages, allowing the selection of different approaches to model variability and distinct reasoners for the reasoning service.

6 Comparison with other tools

Currently, the increment of the complexity of information systems determine more complex OVM to be modeled. Therefore, manual management of such models becomes an impossible labor without an automated tool support. In this regard, we have surveyed existing tools in certain aspects like model specification and its integration with automatic reasoning systems.

FaMa-OVM[6] is an extensible tool for automated analysis of OVM diagrams, integrated with multiple off-the-shelf reasoners such as SAT4j, JavaBDD and Choco-Solver. However, its input model is specified in a textual format, which describes the model relationships, constraints, attributes and global attributes. These last two features correspond to OVM diagrams with attributes, also supported by FaMa-OVM, although the use of a textual format makes the model specification a difficult and error-prone task.

In [5], OVM models are formalised in VFD+ (based in FFD), a parametric construct designed to define the syntax and semantics of FODA-inspired Feature Diagrams(FD) languages. Nonetheless, in order to automate the analysis, the models in VFD+ are translated into CNF to be later sent to the SAT4j solver that determines the formula satisfiability. Furthermore, this tool does not offer any graphical front-end yet.

As it has been mentioned before, both SAT-approaches present certain limitations related to more restrictive logics and thus failing to reflect the finer logical structure of variability models. On these proposals, the relationships among services are missed when embedding variability models in CNF(Conjunctive Normal Form). As a consequence, the domain validation process involves several product-by-product SAT checks. For this reason, identifying inconsistency sources is hard in such approaches as has also been evaluated in [9].

7 Conclusions and Future Work

In this work, we have presented *crowd-variability*, a Web tool that integrates graphical support for orthogonal variability models design and automatic reasoning so as to validate such models. It pursues to solve those limitation presented in surveyed proposals based on SAT-solvers through the use of description logics. Moreover, we aim to avoid error occurrences in models design by focusing on a graphical user interface with a set of primitives of the OVM language and the necessary functions for diagrams construction and edition. Taking this into consideration, we have reached the first prototype of *crowd variability*, which allows simple OVM designs and its consistency analysis by logic reasoning systems. Finally, we address extensions tool capability in a preliminary evaluation.

In future works, we plan to continue evaluating *crowd-variability*, enhance and extend its functionalities. Currently, it determines when a model is valid, however, we also aim to be able to determine dead services, i.e. services that will never be derived for a product, valid product instantiations and to explain the inconsistencies found in the OVM.

References

1. Klaus Pohl, Günter Böckle, and Frank J. van der Linden. *Software Product Line Engineering: Foundations, Principles and Techniques*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
2. Fabricia Roos-Frantz, José A Galindo, David Benavides, Antonio Ruiz Cortés, and J Garcıa-Galán. Automated analysis of diverse variability models with tool support. *Jornadas de Ingenieria del Software y de Bases de Datos (JISBD 2014), Cádiz. Spain*, page 160, 2014.
3. David Benavides, Sergio Segura, and Antonio Ruiz-Cortés. Automated analysis of feature models 20 years later: A literature review. *Inf. Syst.*, 35(6):615–636, September 2010.
4. Matthias Kowal, Sofia Ananieva, and Thomas Thüm. Explaining anomalies in feature models. In *Proceedings of the 2016 ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences*, GPCE 2016, pages 132–143, New York, NY, USA, 2016. ACM.

5. A. Metzger, K. Pohl, P. Heymans, P. Y. Schobbens, and G. Saval. Disambiguating the documentation of variability in software product lines: A separation of concerns, formalization and automated analysis. In *15th IEEE International Requirements Engineering Conference (RE 2007)*, pages 243–253, Oct 2007.
6. Fabricia Roos Frantz, José Ángel Galindo Duarte, David Felipe Benavides Cuevas, and Antonio Ruiz Cortés. FaMa-OVM: A tool for the automated analysis of ovms. In *Proceedings of the 16th International Software Product Line Conference - Volume 2. SPLC '12, ACM, New York, NY, USA*, 2012.
7. Germán Braun, Matias Pol'la, Laura Cecchi, Agustina Buccella, Pablo Fillottrani, and Alejandra Cechich. A DL Semantics for Reasoning over OVM-based Variability Models. 2017.
8. Diego Calvanese, Maurizio Lenzerini, and Daniele Nardi. Description logics for conceptual data modeling. In *Logics for Databases and Information Systems*, pages 229–263. Kluwer, 1998.
9. M. Pol'la, A. Buccella, M. Arias, and A. Cechich. Sevatax: service taxonomy selection validation process for spl development. In *2015 34th International Conference of the Chilean Computer Science Society (SCCC)*, pages 1–6, Nov 2015.
10. T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, May 2001.
11. Volker Haarslev, Kay Hidde, Ralf Möller, and Michael Wessel. The racerpro knowledge representation and reasoning system. *Semantic Web Journal*, 2012.
12. Thorsten Liebig, Marko Luther, Olaf Noppens, and Michael Wessel. Owllink. *Semantic Web*, 2(1):23–32, 2011.
13. Christian Gimenez, Germán Braun, Laura Cecchi, and Pablo Fillottrani. Una arquitectura cliente-servidor para modelado conceptual asistido por razonamiento automático. In *XVIII Workshop de Investigadores en Ciencias de la Computación*, 2016.
14. Christian Gimenez, Germán Braun, Laura Cecchi, and Pablo Fillottrani. crowd: A tool for conceptual modelling assisted by automated reasoning - preliminary report. In *the 2nd Simposio Argentino de Ontologías y sus Aplicaciones SAOA '16 JAIIO '16*, 2016.
15. Eduardo Grosclaude, Rafael Zurita, Rodolfo del Castillo, Miriam Lechner, and José Riquelme. Designing a myro-compatible robot for education as copyleft hardware. In *XX Congreso Argentino de Ciencias de la Computación (Buenos Aires, 2014)*, 2014.
16. Rafael Zurita, Juan de la Fuente, Martín Bucarey, Daiana Bonet, Rodolfo del Castillo, Guillermo Grosso, Laura Cecchi, Jorge Rodríguez, et al. Mejorando las posibilidades de aprender a programar, ampliación del robot educativo multiplo n6 max a frankestito. In *XII Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET, La Matanza 2017).*, 2017.
17. Multiplo —Open Source Robotics Building System. Último acceso Julio 2018, website `http://www.robotgroup.com.ar/en/kits-de-rob%C3%B3tica/robots/n6-max-detail`.
18. Ahmet Soylu, Felix Mödritscher, Fridolin Wild, Patrick De Causmaecker, and Piet Desmet. Mashups by orchestration and widget-based personal environments: Key challenges, solution strategies, and an application. *Program*, 2012.
19. V. Haarslev and R. Möller. Racer system description. In R. Goré, A. Leitsch, and T. Nipkow, editors, *International Joint Conference on Automated Reasoning, IJCAR'2001, June 18-23, Siena, Italy*, pages 701–705. Springer-Verlag, 2001.
20. K. Kang, S. Cohen, J. Hess, W. Nowak, and S. Peterson. Feature-Oriented Domain Analysis (FODA) Feasibility Study. Technical Report CMU/SEI-90-TR-21, Software Engineering Institute, Carnegie Mellon University Pittsburgh, PA., 1990.
21. O. Haugen, B. Moller-Pedersen, J. Oldevik, G. K. Olsen, and A. Svendsen. Adding standardized variability to domain specific languages. In *2008 12th International Software Product Line Conference*, pages 139–148, 2008.

Registro emocional de personas interactuando en contextos de entornos virtuales

Jorge Ierache[1] , Ramiro Nicolosi[1] , Gonzalo Ponce[1], Claudio Cervino[2], Edgardo Eszter[2]

[1]Intituto de Sistemas Inteligentes y Enseñanza Experimental del la Robótica (ISIER-UM)
[2] Instituto de Fisiología y Neurociencias (IFiNe-UM)
Secretaria de Ciencia y Tectología
Universidad de Morón, Cabildo 134, (B1708JPD) Morón, Buenos Aires, Argentina
jierache@unimoron.edu.ar

Resumen. Se presentan un modelo distribuido y sus resultados que implementa el registro EEG para asistir a la inferencia del estado emocional de un individuo interactuando en entornos virtuales, registrando imágenes del entorno virtual y los datos EEG que permiten analizar las emociones y comportamiento de las personas en entornos virtuales que simulan de modo realista situaciones de la vida cotidiana o tareas laborales específicas, de riesgo, brindan un marco apropiado para, no sólo detectar, medir y analizar dichas emociones o respuestas de comportamiento, sino también para poder proyectarlas a situaciones ya no simuladas sino reales.

Palabras clave Interfase cerebro-computadora, Reconocimiento patrones emocionales, Estado biométrico-emocional, Entornos virtuales. Computación afectiva.

1 Introducción

En el contexto de la computación afectiva, el desarrollo del campo de las Interfaces Cerebro-Maquina [1], [2] conocidas por sus siglas en inglés de Brain Control Interface (BCI), se incrementó en los últimos años. Un BCI es un biosensor que registra EEG, señales eléctricas, no pensamientos reales, para traducir la actividad cerebral en acción y así entregar entradas concisas a juegos, robots y dispositivos que ejecutan aplicaciones de salud y bienestar, educativas y de investigación entre otras. Esto posibilita la comunicación entre las funciones mentales y cognitivas de quien la utiliza para luego ser procesadas, clasificadas e interpretadas por aplicaciones o dispositivos puntuales. La investigación de las interfases BCI se desarrolla en un campo científico multidisciplinario con aplicaciones que van desde la computación aplicada en el campo de las neurociencias, domótica, robótica y entretenimiento entre otras; con aplicaciones que van desde mover cursores en pantalla hasta determinar que disc jockey divierte más a la multitud que lo escucha [3], [4]. En trabajos previos del ISIER-UM [5], [6], [7], [8] orientados al control de Robots y el control de artefactos en el contexto de la domótica [9], como así también en la lectura emocional del usuario, enfocando la lectura de la excitación y meditación [10], se experimentó con BCI en particular con EMOTIV [11] integrando la respuesta de biopotenciales

eléctricos de individuos para el control de robots a través del electro-miograma, electro-encefalograma y electro-oculograma que son bioseñales eléctricas generadas por los patrones de actividad de los músculos, el cerebro y los ojos del usuario. Por otro lado, en el IFiNe se viene trabajando en el registro de actividad bioeléctrica cerebral [12], [13] y de otros parámetros fisiológicos, tanto en humanos como en animales de experimentación. Las Brain Control Interface, se utilizan para inferir estados emocionales. En este caso se capturan y analizan los registros de EEG. Actualmente se experimenta con el control de computadoras, dispositivos, sin embargo no se registran desarrollos de sistemas que asistan a la inferencia emocional con capacidades de explotación y desarrollo de integración abierta y transparente a los entornos virtuales, en especial de simulación y entrenamiento. Analizar las emociones y comportamiento de las personas en entornos virtuales que simulan de modo realista situaciones de la vida cotidiana o tareas laborales específicas, de riesgo, brindan un marco apropiado para, no sólo detectar, medir y analizar dichas emociones o respuestas de comportamiento, sino también para poder proyectarlas a situaciones ya no simuladas sino reales. Allí radica el valor de los simuladores como entorno para ensayos de esta naturaleza [14].

2 Problema

El adiestramiento en sistemas virtuales que recrean las escenas y situaciones, basadas en el mundo real, requieren en el contexto de computación afectiva el registro de información biométrica relacionada especialmente con el estado emocional, situación de excitación-relajación con el empleo de una interfaz Cerebro-Máquina (BCI), que permita convertir parámetros fisiológicos, cognitivos y emocionales del usuario (entre otras, la actividad eléctrica cerebral (EEG), parámetros de pulsaciones, presión arterial, frecuencia cardiaca) en especial frente a situaciones simuladas criticas como lo son los fallos en sistemas o plantas y su resolución a través de aplicación de procedimientos de emergencia, a fin de poder evaluar no solo la resolución del caso en forma aislada, sino también el comportamiento biométrico del individuo frente a la situación presentada en el contexto virtual, contribuyendo a logar un perfil y su historial comparativo en sus distintas sesiones de entrenamiento. Luego de observar las aplicaciones existentes y considerar los datos necesarios para analizar las reacciones de las personas dentro de un contexto de entorno virtual crítico, se plantea la necesidad de una aplicación que tenga una captura integral de la información biométrica-emocional de la persona en cada instante en el que interactúa en el entorno virtual.

3 Descripción de la solución

Se diseño una arquitectura de sistema que tiene como entradas principales por un lado la captura de las ondas cerebrales con el empleo del BCI y por el otro la toma de "capturas de pantalla" y las interacciones que el usuario realice con la computadora donde ejecuta el ambiente virtual (empleando, imágenes, simuladores, juegos, etc). Mientras se capturan las imágenes de dichas pruebas, al mismo tiempo el usuario utiliza el BCI de Neurosky para capturar sus ondas cerebrales en otra computadora dedicada. Previamente, el "Time" de ambas computadoras se sincronizará para que la información pueda vincularse correctamente. Esta información se guarda en base de datos dedicada de la aplicación, identificando cada instante con el "Time" sincronizado, que actúa en como valor univoco. La aplicación permite etiquetar los niveles de las ondas para luego poder analizarlo más fácilmente y poder ubicar los momentos en los que cada una de las ondas o niveles supera o es menor que ciertos valores o umbrales de señal ~~configurables .~~ configurables. Se representa en la figura 1 el modelo conceptual del sistema.

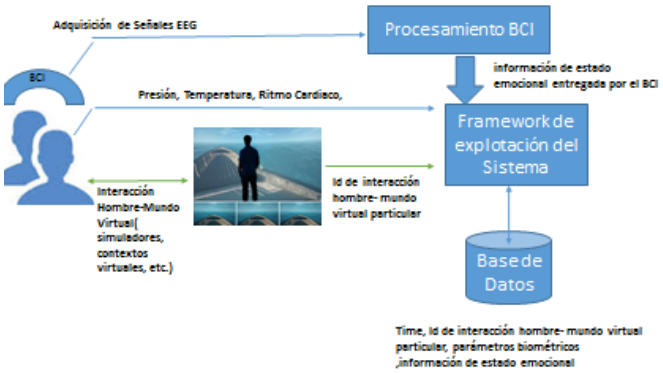


Fig. 1. Modelo conceptual del sistema de influencia del estado biométrico emocional de personas interactuando en mundos virtuales

El BCI empleado es el NeuroSky [15], sus algoritmos [16], brindan información en relación a estado de:

a) **Relajación ,** indica el nivel de "calma" o "relajación" mental, el valor varía de 0 a 100, y aumenta cuando los usuarios relajan la mente y decrece cuando están inquietos o estresados. El medidor cuantifica la capacidad de encontrar un estado interno de atención plena y puede ayudar a los usuarios a aprender cómo autocorregirse y encontrar el equilibrio interno para superar situaciones de~~l~~ tensión;

b) **Atención:** indica la intensidad del "foco" mental o "atención". El valor varía de 0 a 100. El nivel de atención aumenta cuando un usuario se enfoca en un solo pensamiento o en un objeto externo, y disminuye cuando se distrae. Los usuarios pueden observar su habilidad para concentrarse usando el algoritmo.

c) **Detección de parpadeo:** señala los parpadeos de un usuario. Un número más alto indica un parpadeo "más fuerte", mientras que un número más pequeño indica un parpadeo "más claro" o "más débil". La frecuencia de parpadeo a menudo se correlaciona con nerviosismo o fatiga. Los parpadeos son similares a los de un sistema binario on / off estándar y, por lo tanto, son valiosos para los controles que requieren respuestas definitivas. Por ejemplo, en aplicaciones de comunicación, un abrir y cerrar de ojos significa no, dos significan sí, dar a las personas con necesidades especiales una forma sencilla de comunicarse;

d) **Esfuerzo mental:** mide la carga de trabajo mental al realizar una tarea. Cuanto más difícil es el cerebro de un usuario en una tarea, mayor es el valor. El algoritmo funciona bien con tareas físicas (por ejemplo, dibujo) y mentales (por ejemplo, recitado), y se puede usar para el seguimiento continuo en tiempo real y comparaciones entre ensayos para medir los efectos de la multitarea, la variabilidad de la carga de trabajo y más. El algoritmo se puede usar para rastrear los efectos de diversas cargas cognitivas sobre la capacidad de aprender y proporcionar retroalimentación para la auto-mejora del usuario.

Se presenta a continuación una descripción de las principales clases desarrolladas, en este orden, la aplicación de captura de bioseñales cuenta con una interfaz la cual se puede ver en la figura 2, esta, manipula los datos necesarios para adquisición de los datos biométricos (PORT, ROW_TIME, POORSIGNAL, ATTENTION RELAJACION, EEGPOWER, EegPowerDelta, EegPowerThetha, EegPowerAlpha1, EegPowerAlpha2, EegPowerBeta1, EegPowerBeta2, EegPowerGamma1, EegPowerGamma2) que entrega el BCI, para esto cuenta con las variables locales donde van a ser almacenadas cada una de estas.. También almacenara la fecha y hora de ejecución precisa para que pueda hacerse la integración con los datos de la aplicación de captura de imágenes. A su vez cuenta con datos extras que serán, el nombre de la prueba, el puerto a utilizar y una breve descripción, la cual la aplicación deja al usuario que los rellene por pantalla. Para la programación de la aplicación de captura de imágenes, como se observa en la Figura 3, tenemos tres interfaces, las cuales son: capturaDeImagen, VisualizacionDeBD y RegistroSeleccionado. La primera es la que manipula toda la aplicación, su funcionalidad es la de capturar las imágenes y guardarlas en la base de datos, la cual al inicializar la interfaz genera las conexiones correspondientes. Posee dentro la locación donde se almacenarán las imágenes, un identificador único el cual se reinicia por cada sección. La sección es un número que indica el paquete actual de las capturas. Dentro de la misma interfaz tenemos un cronómetro interno, el cual gobierna el tiempo que hay entre una captura y la otra. La segunda interfaz, VisualizacionDeBD, solamente posee una conexión a la base correspondiente y un componente que permite la visualización de la tabla que se le indique de la base de datos en forma de grilla. Una cualidad que se le agrego permite al usuario seleccionar un registro, se abra la tercera interfaz para ver más en detalle el contenido del registro. Para la tercer y ultima interfaz, RegistroSeleccionado una ventana simple permite la muestra todos los campos ordenadamente y una pre visualización de la captura, además, dicha interfaz, permite modificar el campo Descripción y también poder darle doble click a la imagen y poder verla con la aplicación predeterminada de imágenes que posea la computadora donde se corre dicha aplicación.

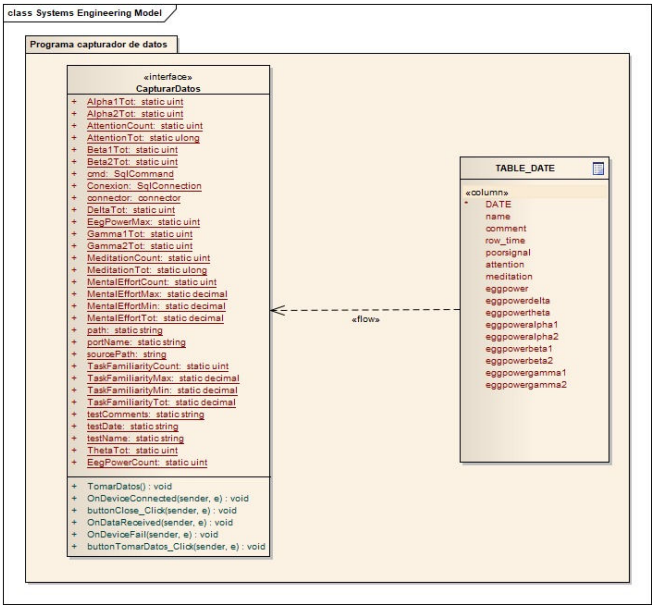


Fig 2. Diagrama del sistema capturador de ondas cerebrales

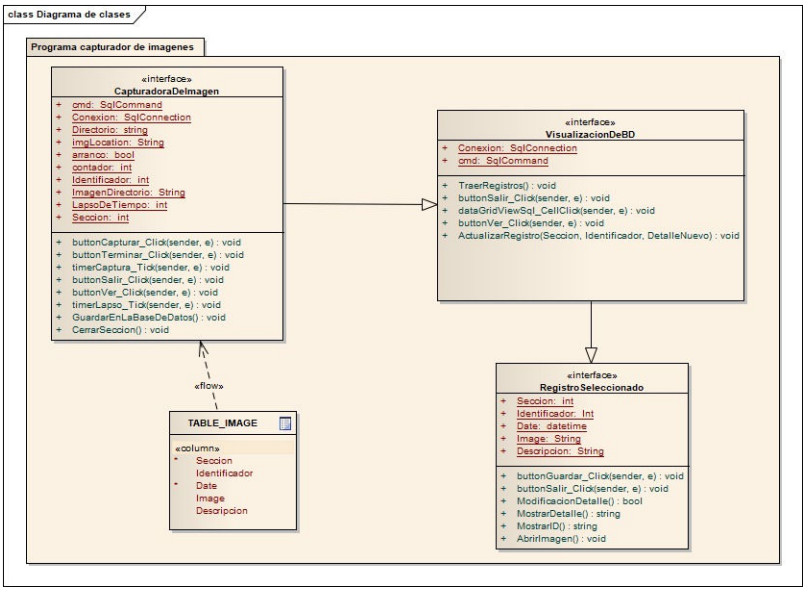


Fig 3. Diagrama de clases de sistema capturador de imágenes

El framework de registro de bioseñales correspondiente al BCI se implementó en el lenguaje C#, utilizando la librería ThinkGear.dll la cual posee las funciones necesarias para interpretar las señales. Esta aplicación registra los metadatos de atención y relajación en un rango de entre 0 y 100. Asimismo, también registra los datos en hz de las ondas cerebrales. La estructura de los datos que registra es la siguiente: HORARIO_DE_PC, TEST_NAME, TEST_COMMENT, PORT, ROW_TIME, POORSIGNAL, ATTENTION, RELAJACION (también llamada Meditación), EEGPOWER, EegPowerDelta, EegPowerThetha, EegPowerAlpha1, EegPowerAlpha2, EegPowerBeta1, EegPowerBeta2, EegPowerGamma1, EegPowerGamma2. Se puede ver un ejemplo en la figura 4

	HORARIO_DE_PC	TEST_NAME	TEST_COMMENT	PORT	ROW_TIME	POORSIGNAL	ATTENTION	MEDITATION	EGGPOWER	EegPowerDelta	EegPowerTheta
333	09/07/2018 07:36:4...	PruebaDinamica	PruebaDinamica	COM3	1531175808	99724	50	37	2892387	89080	22464
334	09/07/2018 07:36:5...	PruebaDinamica	PruebaDinamica	COM3	1531175810	96158	56	54	2892387	23262	22745
335	09/07/2018 07:36:5...	PruebaDinamica	PruebaDinamica	COM3	1531175811	96823	48	43	2892387	78090	36763
336	09/07/2018 07:36:5...	PruebaDinamica	PruebaDinamica	COM3	1531175812	97495	54	50	2892387	12748	11021
337	09/07/2018 07:36:5...	PruebaDinamica	PruebaDinamica	COM3	1531175813	94362	77	44	2892387	110626	50807
338	09/07/2018 07:36:5...	PruebaDinamica	PruebaDinamica	COM3	1531175814	94947	64	53	2892387	32808	17087
339	09/07/2018 07:36:5...	PruebaDinamica	PruebaDinamica	COM3	1531175815	94585	64	51	2892387	21721	11220
340	09/07/2018 07:36:5...	PruebaDinamica	PruebaDinamica	COM3	1531175816	94358	74	44	2892387	285978	11508
341	09/07/2018 07:36:5...	PruebaDinamica	PruebaDinamica	COM3	1531175817	92227	74	44	2892387	20235	9206
342	09/07/2018 07:36:5...	PruebaDinamica	PruebaDinamica	COM3	1531175818	92466	77	48	2892387	134145	16346
343	09/07/2018 07:36:5...	PruebaDinamica	PruebaDinamica	COM3	1531175819	9125	74	54	2892387	8713	10072
344	09/07/2018 07:37:0...	PruebaDinamica	PruebaDinamica	COM3	1531175820	89891	75	61	2892387	24194	12459
345	09/07/2018 07:37:0...	PruebaDinamica	PruebaDinamica	COM3	1531175821	92086	66	69	2892387	47144	66218
346	09/07/2018 07:37:0...	PruebaDinamica	PruebaDinamica	COM3	1531175822	91896	67	60	2892387	46277	25733
347	09/07/2018 07:37:0...	PruebaDinamica	PruebaDinamica	COM3	1531175823	90183	61	54	2892387	13695	15919

Fig 4. Base de datos de la aplicación de captura de bioseñales

El sistema de captura de imágenes del entorno virtual en el que interacciona el usuario también se ha implementado en el lenguaje C# y registra en la base de datos secciones, identificadores, fecha, la ubicación de la imagen y una descripción de la misma. La estructura de los datos que se registran es la siguiente: SECCION, IDENTIFICADOR, DATE, IMAGE, DESCRIPCION. Se puede observar un ejemplo en la figura 5: La base de datos que captura de imágenes se implementó en SQL Server.

	SECCION	IDENTIFICADOR	DATE	IMAGE	DES
233	4	40	09/07/2018 07:22:14.946 p.m.	D:\Universidad\TESIS\Capturas\image-636667717349...	Car
234	4	41	09/07/2018 07:22:15.949 p.m.	D:\Universidad\TESIS\Capturas\image-636667717359...	Car
235	4	42	09/07/2018 07:22:16.961 p.m.	D:\Universidad\TESIS\Capturas\image-636667717369...	Car
236	4	43	09/07/2018 07:22:17.958 p.m.	D:\Universidad\TESIS\Capturas\image-636667717379...	Car
237	4	44	09/07/2018 07:22:18.943 p.m.	D:\Universidad\TESIS\Capturas\image-636667717389...	Car
238	4	45	09/07/2018 07:22:19.961 p.m.	D:\Universidad\TESIS\Capturas\image-636667717399...	Car
239	4	46	09/07/2018 07:22:20.964 p.m.	D:\Universidad\TESIS\Capturas\image-636667717409...	Car
240	4	47	09/07/2018 07:22:21.955 p.m.	D:\Universidad\TESIS\Capturas\image-636667717419...	Car
241	4	48	09/07/2018 07:22:22.959 p.m.	D:\Universidad\TESIS\Capturas\image-636667717429...	Car
242	4	49	09/07/2018 07:22:23.942 p.m.	D:\Universidad\TESIS\Capturas\image-636667717439...	Car
243	4	50	09/07/2018 07:22:24.950 p.m.	D:\Universidad\TESIS\Capturas\image-636667717449...	Car
244	4	51	09/07/2018 07:22:25.969 p.m.	D:\Universidad\TESIS\Capturas\image-636667717459...	Car
245	4	52	09/07/2018 07:22:26.954 p.m.	D:\Universidad\TESIS\Capturas\image-636667717469...	Car
246	4	53	09/07/2018 07:22:27.958 p.m.	D:\Universidad\TESIS\Capturas\image-636667717479...	Car
247	4	54	09/07/2018 07:22:28.974 p.m.	D:\Universidad\TESIS\Capturas\image-636667717489...	Car

Fig 5. Base de datos de la aplicación de captura de imágenes

4 Pruebas y Resultados Obtenidos

El desarrollo de las pruebas tiene como objetivo el registro de EEG, imágenes asociadas y estados de relajación y atención del individuo, durante su actuación en un entorno virtual. La experimentación inicial abordará situaciones de casos de borde contribuyendo con el estado de relajación o en situaciones que no contribuyan con el estado de relajación. Se diseñaron las pruebas considerando dos casos iniciales .El primero corresponde a un ambiente estático donde se utilizan un conjunto de fotos, imágenes y sonidos que contribuyan con el estado de "Relajación". Estos elementos irán variando en ciertos periodos de tiempo, momentos de relajación y momentos generados por imágenes de que no contribuyen con el estado de relajación. Las imágenes empleadas pueden ser seleccionadas por los directores del experimento. En particular las pruebas basales se realizaron con diversos usuarios, se presenta el resultado de dos usuarios, para los cuales se emplearon diez imágenes de relajación y otras diez imágenes que buscan alterar la relajación, con un intervalo temporal de 5 segundos entre imágenes, ver set de imágenes en [17]. El segundo, caso de pruebas el usuario utiliza un entorno dinámico representado por un juego virtual "Counter Strike Global Ofensive", ver las imágenes capturadas en [18], con todas las tareas y eventos que esto representa durante un promedio de 180 segundos de juego, donde se registra el grado de atención del usuario, además de los parámetros del EEG. En la figura 6 se pueden observar los gráficos con los resultados de las pruebas estáticas de dos usuarios y sus picos máximos de relajación, acompañado por sus correspondientes imágenes fig 7a primer usuario y fig 7 b segundo usuario. En la figura 8 se pueden observar los resultados de las pruebas dinámicas de dos usuarios con sus respectivos gráficos, se muestran el momento en el que los usuarios tuvieron el mayor grado de atención mientras ejecutaban el juego, acompañado por la captura del sistema de las imágenes correspondientes, al primer usuario en fig 9 a y del segundo usuario en figura 9b

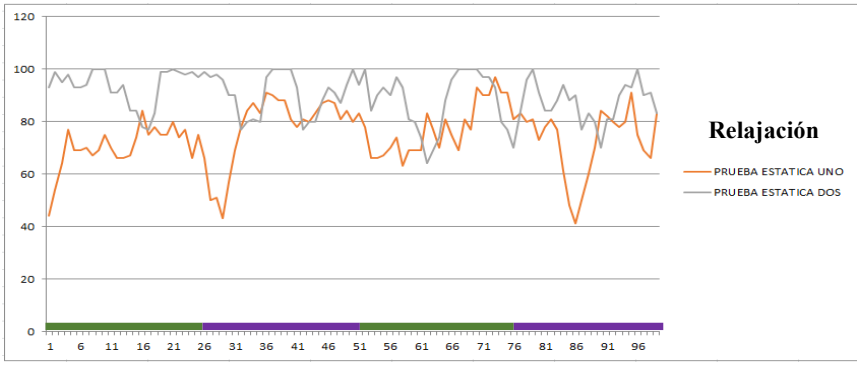


Fig 6 Resultados de la prueba estática de relajación. En la línea de tiempo inferior (segundos) se ve en verde los momentos de las imágenes de relajación y en violeta los

momentos en donde cada usuario se encontraba viendo imágenes que afectan el estado de relajación.



Fig 7 a Imagen primer usuario



Fig 7 b Imagen segundo usuario

Pruebas de ambiente estático correspondientes a las imágenes de máxima relajación de cada usuario

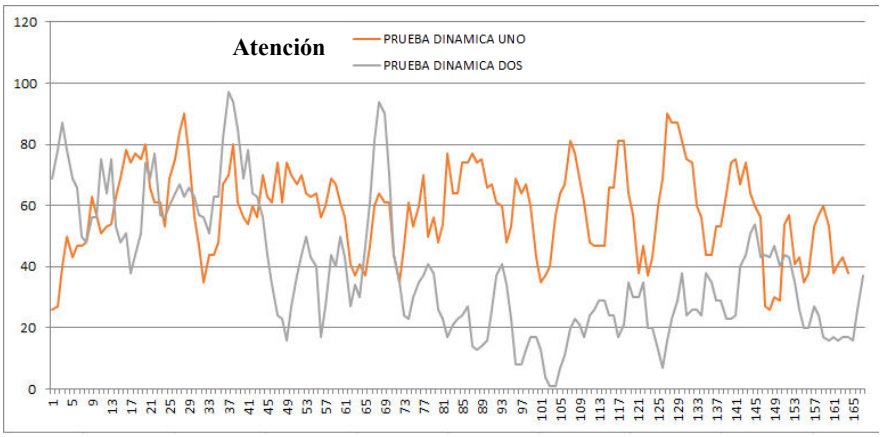


Figura 8: Resultados de la prueba dinámica. Eje X expresado en segundos





Fig 9 a captura de imagen primer usuario Fig 9b captura de imagen segundo usuario

Pruebas de ambiente dinámico correspondiente a las imágenes capturadas con un pico máximo en el nivel de atención de cada usuario.

5 Conclusiones y futuras líneas de trabajo

Los resultado correspondientes a las pruebas basales fueron exitoso en términos de las funcionalidades de registros de parámetros biométricos de EEG y sus metadatos en correspondencia con el contexto de ambiente virtual del usuario en términos de la captura de imágenes correlacionadas con sus valores biométricos, esto contribuye con el desarrollo der ambientes de computación afectiva en función del perfil emocional del usuario , además de brindar información valiosa en particular en ambientes de entrenadores / simuladores en relación al estado emocional del usuario, en particular si estos sistemas son de misión crítica. Como futuras líneas de trabajo se encuentra en desarrollo inicial la integración de parámetros fisiológicos del usuario. En el marco de las futuras líneas de investigación se explorara la integración de una "api" que captura imágenes de rostros de la cara de la persona y permite dar más información sobre el estado emocional.

6 Agradecimiento

Los autores del presente trabajo agradecen el apoyo brindado por la SECYT-UM en el marco del proyecto PING 01-001-17

7 Referencias

1. Hamadicharef, "Brain Computer Interface Literature- A bibliometric study", in 10th International Conference on Information Science, Signal Processing and their Applications, Kuala Lumpur, 2010, pp. 626-629.

2. J. del R Millán, "Brain-computer interfaces," in Handbook of Brain Theory and Neural Networks, 2nd ed, M.A. Arbib, Ed. Cambridge, MA: MIT Press, 2002

3. M. A. Lebedev and M. A. L. Nicolelis, "Brainmachine interfaces: Past, present and future," Trends in Neurosciences, vol. 29, no. 9, pp. 536-546, September 2006.

4. R. Wolpaw, D. J. McFarland, "Brain-computer interface research at the Wadsworth center," IEEE Trans. Rehab. Eng., vol. 8, pp. 222–226, 2000

5. Ierache, J., Dittler M., Pereira G., García Martínez R, "Robot Control on the basis of Bio-electrical signals" XV CACIC 2009, UNJu, ISBN 978-897-24068-3-9

6. Ierache J, Pereira G, Iribarren J, Sattolo I, "Robot Control on the Basis of Bio-electrical Signals": "International Conference on Robot Intelligence Technology and Applications" (RiTA 2012). Series Advances in Intelligent and Soft Computing of Springer.

7. Ierache., J, Pereira.,G, Sattolo.,I ,Guerrero., A, D'Altto J, Iribarren,. J. Control vía Internet de un Robot ubicado en un sitio remoto aplicando una Interfase Cerebro-Máquina ".XVII CACIC 2011, UNLP, ISBN 978-950-34-0756-1,p 1373-1382.

8. Ierache J, Pereira G, Iribarren J "Demostración de los resultados en la integración de Interfases Lectoras de Bioseñales aplicadas al Control de un Robot" VII TEYET 2012 UNNOBA, 2012, demos educativas. ISBN 978-987-28186-3-0.

9. Ierache., J, Pereira., G, J, Iribarren Framework for Brain Computer Interface implemented to control devices in the context of home automation XIX CACIC 2013, CAECE Mar del Plata, ISBN 978-897-23963-12

10. Ierache J, Nervo F, Pereira G, Iribarren J Estado emocional centrado en estímulos, aplicando Interfase cerebro-maquina, XX Congreso Argentino de Ciencias de la Computación (Buenos Aires, 2014) ISBN 978-987-3806-05-6

11. Emotiv Epoc Vigente 2015 http://www.emotiv.com/.

12. Cervino, C. 1999. MS. Nuevo ritmo alfa y actividad gamma en los bulbos olfatorios de un armadillo sudamericano: cuantificación y perspectivas futuras. Premio "Prof. Dr. Eduardo de Robertis", al mejor trabajo de Neurociencia Básica. 455 pág.. Septiembre 1999.

13. Affanni, JM. y CO. Cervino. 2000. Actividad Bioeléctrica Cerebral. En: Fisiología Humana de Houssay (7ma ed.), A. Houssay y C. Cingolani (eds.). Buenos Aires: Ed. El Ateneo. 1120 pág.: 1015-1039.

14 Jorge Ierache, Claudio Cervino, Edgardo Eszter, Daniel Fortin, Alvaro Castro Menna Actas del XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018), Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste, 26 y 27 de abril de 2018, ISBN: 978-987-3619-27-4, pp 785-789.

15 http://neurosky.com/ vigente julio 2018

16 http://neurosky.com/biosensors/eeg-sensor/algorithms/ vigente julio 2018

17 https://drive.google.com/open?id=10lA3cSkxpkIBEriRa1R6nedm8G6qfer- vigente julio 2018

18 Counter Strike Global Ofensive https://drive.google.com/open?id=13yntTzZG3HxQA951UZpFXqDRIozPFeSV vigente julio 2018

Análisis de las Tecnologías de la Información y la Comunicación y su innovación en la industria

Alicia Mon, Horacio René Del Giorgio

Universidad Nacional de La Matanza – Departamento de Ingeniería e Investigaciones Tecnológicas - Florencio Varela 1903 (B1754JEC) - San Justo, Buenos Aires, Argentina
alicialmon@gmail.com , hdelgiorgio@unlam.edu.ar

Abstract

En el presente artículo se exponen los resultados parciales de una investigación en curso cuyo objetivo reside en la creación de un índice de TICs (Tecnologías de la Información y las Comunicaciones) para detectar los tipos de tecnologías instaladas en la actualidad en la industria, analizar cuánto valor agrega el uso de estas tecnologías en las diferentes ramas industriales y divisar las necesidades de desarrollo, implementación e innovación de software en las cadenas de valor.

Para ello se ha propuesto una tipificación de tecnologías presentes en la industria, diferenciándolas entre productos software, equipamiento hardware, e infraestructura y comunicaciones, de modo que permita detectar cuáles son los productos que la industria tiene implementados, según las funciones para las que se los utiliza.

Finalmente se ha realizado una validación de las taxonomías con un grupo de expertos en el área de TICs y de software en particular, mediante una serie de entrevistas, a efectos de especificar la tipificación propuesta para establecer diferentes niveles de desarrollo.

Keywords: TICs, Índice de TICs, Desarrollo Tecnológico, Industria.

1 Introducción

El uso de las TICs (Tecnologías de la Información y las Comunicaciones) impacta directamente en la estimulación de la producción industrial y ofrece un considerable potencial para el crecimiento de las diversas ramas productivas [1], dado que generan mayor valor agregado en los diferentes procesos, permitiendo mejorar los niveles de productividad y competitividad.

Las TICs en general y la implantación de software en particular tienden a facilitar un reordenamiento de los procesos productivos, de logística y distribución, así como el control sobre las cadenas de comercialización, generando un mayor valor agregado sobre el producto final.

La incorporación de nuevas tecnologías en los sectores industriales requiere de un profundo conocimiento sobre la capacidad existente, es decir que, sin información relativa a las TICs instaladas y utilizadas en los diferentes procesos no es posible definir necesidades de incorporación tecnológica para generar una reconversión en las cadenas de valor.

Sin embargo, poner en acción dicho potencial depende crucialmente de la realización de profundos cambios en la estructura productiva, reorganización de los negocios, desarrollo de capital humano y las estrategias de promoción de las políticas públicas.

Estas consideraciones son válidas tanto para los países en desarrollo como para los países más avanzados.

La diversa bibliografía sobre el desarrollo productivo y los desarrollos tecnológicos [2] [3] [4] no brinda con precisión una forma específica de medir los diferentes niveles de desarrollo industrial según las TICs que tiene implementadas y el impacto que generan en los niveles de productividad, así como en las estrategias de innovación requeridas por la industria.

En lo que respecta a la Industria, el uso de las TICs puede referirse a tareas específicas implicadas en la creación de un producto (tecnologías de producto), a tareas involucradas en el desarrollo de un proceso productivo (tecnologías de proceso), a las prácticas implicadas para la operación de las distintas funciones de una unidad productiva (tecnologías de gestión), o bien a las prácticas realizadas para garantizar la correcta apropiación de las competencias por parte de consumidores y usuarios (tecnologías de uso) [5].

En el siguiente apartado se exponen un conjunto de Tipologías ordenadas en base a diferentes taxonomías [6] que permiten analizar las tecnologías implementadas en cada área. Luego, se presentarán los resultados de la validación de dicha tipificación, a partir de un estudio realizado con expertos.

2 Metodología

La definición de las taxonomías derivadas de la tipificación de TICs propuesta ha sido validada con expertos de la industria del software (Cámara de Empresas del Software - CESSI), de la industria local (Unión Industrial Argentina - UIA) y de organismos gubernamentales (Secretaría de Industria de La Matanza) a través de la realización de entrevistas en profundidad y un cuestionario sistematizado en el cual se presentaban las diferentes taxonomías construidas y cuyos resultados se exponen a continuación.

Para la elaboración de una taxonomía en la Industria, publicada en otros artículos por el grupo de investigación [7] [8] [9], se ha confeccionado una tipificación de los procesos industriales definiendo las funciones de una empresa tomando como base la Cadena de Valor de Porter. El modelo propuesto permite analizar el desempeño de una empresa organizando el análisis en relación con el conjunto de actividades primarias y de apoyo, siendo cada una de éstas fuente potencial de ventajas competitivas en costos o diferenciación, y cuyas interrelaciones permiten lograr un mayor valor diferencial emergente que pueda ser apreciado y reconocido por los compradores, en comparación con otras ofertas de la competencia.

Cada uno de los tipos de TICs agrupados en la taxonomía presentada diferencia entre productos software, hardware y comunicaciones, aporta valor según la función o el área de negocio en la que esté implementado, es decir que, se ha definido el conjunto de productos específicos utilizados para cada función de los procesos industriales presentada.

A partir del cruce de las dos taxonomías desarrolladas (TICs y Procesos Productivos) se ha propuesto la generación de instrumentos que permitan evaluar el nivel de inserción de TICs en las diferentes áreas o procesos industriales, separándolas por niveles de desarrollo tecnológico, según cuenten con tecnología básica o tecnología más avanzada tendiente a la transformación de la industria 4.0.

3 Tipificación propuesta para las TICs

La primera taxonomía desarrollada, que se expone en el presente artículo, se propone sistematizar el agrupamiento de TICs en 3 categorías de productos con el mismo nivel de jerarquía, interacción y dependencia entre sí, que corresponden a diversas áreas de desarrollo tecnológico, pero que se necesitan y se complementan en forma directa. En este sentido, la taxonomía organiza y diferencia entre productos software, equipos o hardware y comunicaciones o infraestructura.

Cada categoría incluye en su interior un conjunto de funciones y subcategorías de productos específicos que cumplen un tipo de función específica.

La siguiente Figura expone la Taxonomía con la tipología de productos software agrupados por tipo de función que cumplen o permiten que se realice, independientemente del área funcional de la industria en la que se encuentren implementados.



Figura 1 - Taxonomía de software. Fuente: Elaboración propia

En la siguiente Figura se expone la Taxonomía en la cual se ha definido la tipología de equipamiento y hardware, agrupados por tipo de función que cumplen o permiten que se realice, independientemente del área funcional de la industria en la que se encuentren implementados.

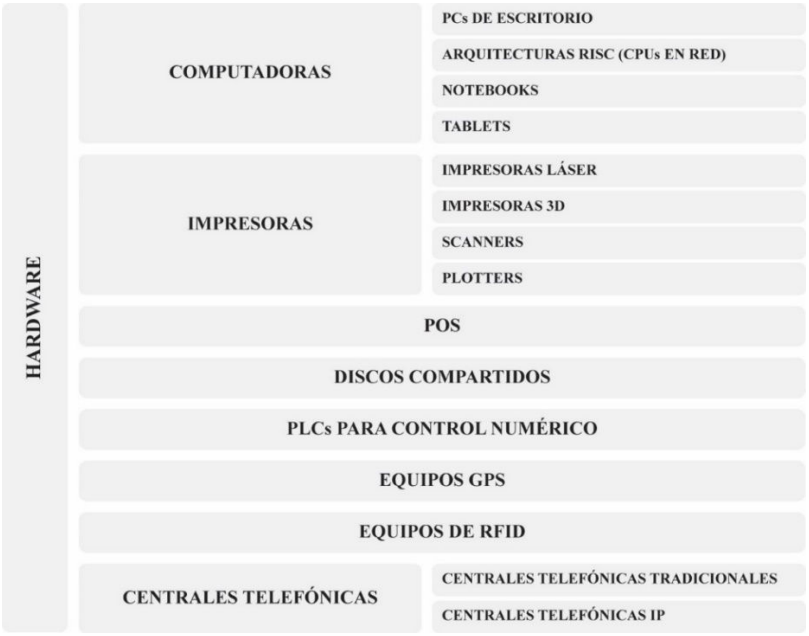


Figura 2 - Taxonomía de Equipamiento y Hardware. Fuente: Elaboración propia

En tanto que en la Figura que sigue, se expone la Taxonomía en la cual se ha definido la tipología de infraestructura y comunicaciones agrupadas por tipo de función que cumplen o permiten que se realice, independientemente del área funcional de la industria en la que se encuentren implementados este tipo de productos.

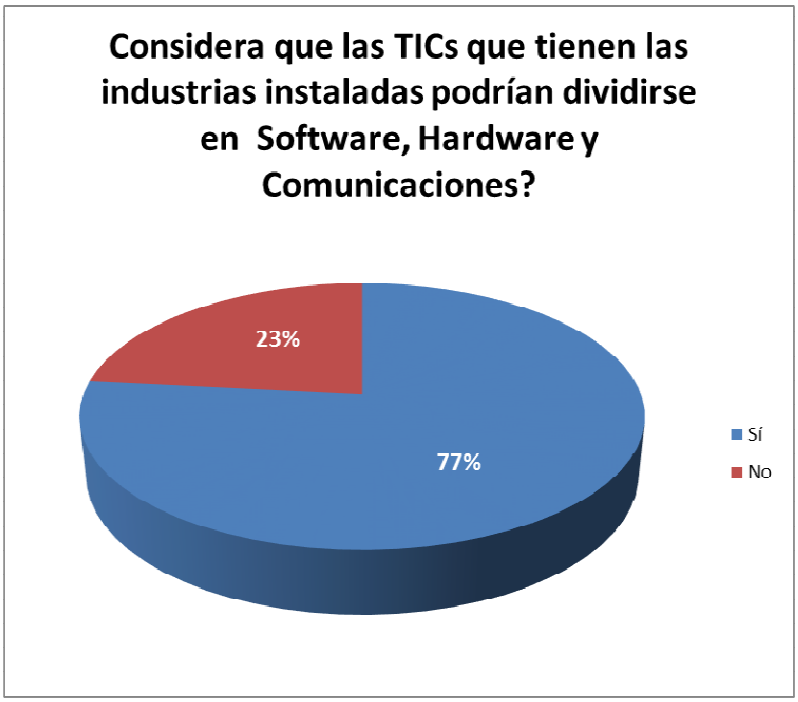


Figura 3 - Taxonomía de Infraestructura y Comunicaciones. Fuente: Elaboración propia

4 Validación y Mejora

Tal como se mencionó anteriormente y para cumplir con los objetivos propuestos, se consultó a 30 expertos de la industria del software (Cámara de Empresas del Software - CESSI), de la industria local (Unión Industrial Argentina - UIA) y de organismos gubernamentales (Secretaría de Industria de La Matanza) a través de la realización de un cuestionario estructurado y sistematizado, a efectos de validar la tipificación de TICs propuesta en el apartado anterior. El proceso de validación incluyó una guía de pautas de indagación que permitió conocer razones y motivos de la inclusión / exclusión de indicadores seleccionados para la construcción de un índice de TICs que permita la evaluación de diferentes niveles de desarrollo tecnológico que contengan las industrias de acuerdo con las tecnologías implementadas.

A continuación, se presentan los resultados de dicha consulta.



Cuadro 1

Tal como se puede observar en el Cuadro 1, una gran mayoría está de acuerdo con esta división propuesta, razón por la cual se dará por validada.

En el caso negativo (de desacuerdo, o inclusive, aun estando de acuerdo), se han obtenido, entre otras, las siguientes propuestas para considerar:

- La industria TIC está realizando su convergencia final a Cloud. Cualquier aplicación de IT estará soportada por la misma infraestructura Cloud.
- En algunos casos se podría simplificar Hardware y Comunicaciones en una sola categoría.
- Hardware y Software ya se comercializan como servicio. Para el cliente lo importante es el servicio más que el Hardware y Software que corra.
- Respecto de las Comunicaciones es un poco más difícil. Dentro de este grupo puede haber Hardware, Seguridad Informática y Servicio de proveedores.
- NFV (Network Functions Virtualization)

Respecto de los comentarios surgidos de esta primera validación, más allá de que hay diversas opiniones en las que aparecen zonas grises en la forma de clasificar, se

puede observar que han realizado comentarios acerca de la Virtualización, aunque por el momento puede que esto sea aún embrionario a nivel de una empresa industrial. Del mismo modo, cuando se mencionan las opciones de Hardware y Software como Servicio, nuevamente es válido el comentario, pero nuevamente pareciera ser muy embrionario para detectarse aún en las tecnologías implementadas en una Industria. De hecho, se ha mencionado al Software como Servicio cuando se describieron los Servicios Cloud. Estas observaciones, aportan en la conformación del grupo de TICs más evolucionadas que pueden encontrarse en las industrias más desarrolladas que se representan en el índice de evaluación que se está construyendo.



Cuadro 2

Para esta pregunta, se expuso la Figura 1, y se puede observar que una gran mayoría considera que no es necesario incluir algún otro tipo de Software, razón por la cual se da por validada esta clasificación.

En el caso negativo, de aquellos entrevistados que consideran que haría falta la inclusión de algún elemento, o inclusive de quienes están de acuerdo con la clasificación, se han obtenido, entre otras, las siguientes propuestas para considerar:

- En referencia a la arquitectura de Cloud, estaría faltando el elemento de gestión u "orquestador"
- El Método de Elementos Finitos es un modelo matemático para resolver cuestiones difíciles de modelar, más propio de un ambiente de Investigación y Desarrollo que de una Industria o Fábrica.
- Software para sistemas embebidos
- Mediciones con relación a la eficiencia energética
- Sistemas SCADA

Nuevamente surgen observaciones sobre tecnologías Cloud, que ya se han tratado en los párrafos anteriores. No obstante, en este caso han aparecido algunos cuestionamientos para tener en cuenta. Es el caso del “Método de Elementos Finitos”, que será considerado para su continuidad o exclusión en una próxima etapa. Con

respecto a “Software para Sistemas Embebidos”, quizás se trate de algo muy específico, aunque muy utilizado en la industria, razón por la cual también se lo tendrá en cuenta. También se da el mismo caso en los Sistemas SCADA, aunque los mismos están cubiertos, de alguna manera, mediante los PLCs, que sí han sido tratados en la taxonomía expuesta. Lo que definitivamente es un hallazgo es lo referente a las “mediciones con relación a la eficiencia energética”. Sin duda se tendrá en cuenta en una próxima revisión de la taxonomía de productos Software.



Cuadro 3

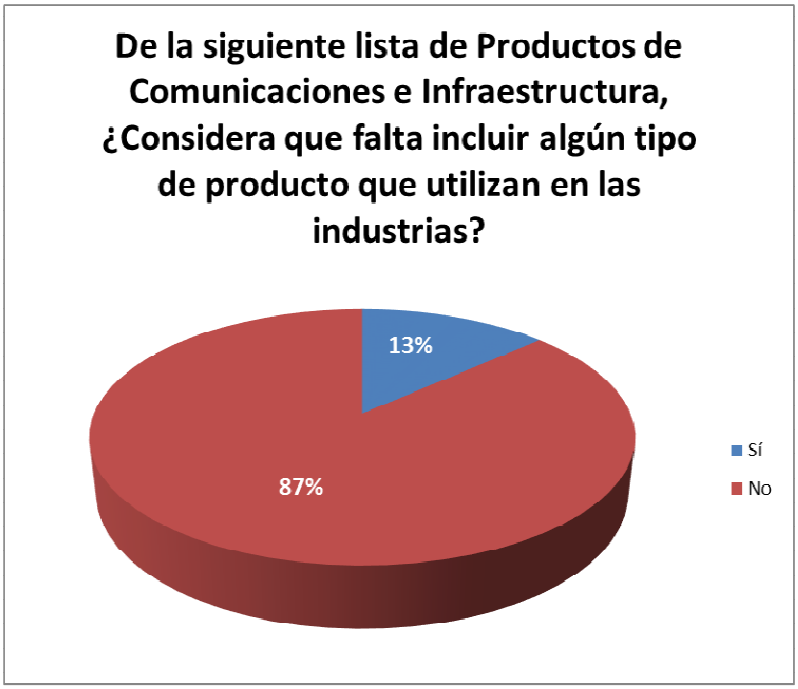
En esta pregunta, se expuso la taxonomía presentada en la Figura 2. La validación de esta tipología permite observar una paridad de opiniones, donde los que sugieren que faltan incluir elementos es apenas mayor. Del análisis de los aportes realizados por este grupo de expertos, se puede observar que es probable que hayan surgido algunas confusiones debido a las zonas grises que se han mencionado al inicio de este análisis, dado que proponen opciones para incluir elementos que ya han sido considerados en la parte de Infraestructura/Comunicaciones (Figura 3).

En el caso negativo (de quienes consideran que haría falta la inclusión de algún elemento, o inclusive de quienes están de acuerdo con la clasificación), se han obtenido, entre otras, las siguientes propuestas para considerar:

- Smartphones, Servidores (*comunicaciones/infraestructura*)
- Dispositivos móviles (*comunicaciones/infraestructura*)
- Celulares (*comunicaciones/infraestructura*)
- Sistemas de Comunicaciones Unificadas (*software*)
- Sistema de Cámaras de seguridad y control de acceso (*software*)
- Equipamiento para control y seguimiento de personas y/o bienes. (*software*)
- Medidores, Sensores, todo lo relacionado a medición (*software*)
- Herramientas de testeo (*software*)
- En PC, incluir "Raspberry pi". Y como otra clasificación agregar: "Sistemas embebidos para automatización y control".

- En una arquitectura de Cloud el sistema completo es soportado por uno o dos Data Centers, como minino. Esto incluye fundamentalmente servidores con blades multicore, storage, switches y SDN routers.
- Incluir los PLCs en "Infraestructura", pero además se deberían considerar equipos de Control Numérico. También, ya que se invocó la Arquitectura RISC, se podría tener en cuenta la SPARC, que utilizan los equipos SUN. Finalmente, las Centrales Telefónicas deberían estar en el listado de Infraestructura.

En la mayoría de los casos se puede observar que se trata de elementos que ya se han tenido en cuenta en la taxonomía de Comunicaciones/Infraestructura o bien de productos Software. En esos casos se hizo una aclaración entre paréntesis y en itálica al lado de cada propuesta, indicando donde estaba incluida en la taxonomía original. Asimismo, nuevamente surge la opción de la Virtualización, algunos puntos de vista (también derivados de estos "grises" que aparecen) sobre qué debería estar en *Infraestructura* más que en *Hardware*, o viceversa. Lo que sin duda se puede rescatar para un análisis posterior es la consideración de Raspberry Pi y de Equipos de Control Numérico.



Cuadro 4

Sobre la pregunta acerca de la tipología de Comunicaciones e Infraestructura se expuso la Figura 3. y en este caso, según se puede observar, el 87% considera que no es necesario incluir algún otro tipo de productos para Comunicaciones e Infraestructura, razón por la cual también se dará por validada esta clasificación.

En el caso negativo, de quienes consideran que haría falta la inclusión de algún elemento, o inclusive de quienes están de acuerdo con la clasificación, se han obtenido, entre otras, las siguientes propuestas para considerar:

- Virtualización.
- Protocolos y software para comunicaciones industriales.
- Más allá de incluir, sería "excluir" el concepto de Internet de las Cosas que está más asociado al usuario final que a la industria en sí.
- Conexiones MPLS/L2L entre sucursales

Nuevamente aparece la mención a la Virtualización, tema que ya se ha tratado en párrafos anteriores. En el caso de los "Protocolos y software para comunicaciones industriales", si bien pareciera ser algo demasiado específico, se puede considerar en una futura revisión. Con respecto a la observación sobre "Internet de las Cosas", será tomado como referencia para detallar con mayor nivel de profundidad cuales son los equipos o la infraestructura necesaria para el desarrollo de Internet de las Cosas, así como extrapolarlo con el concepto de "Internet del Usuario Final". Finalmente, respecto a las conexiones MPLS/L2L entre Sucursales, constituye una observación para una nueva revisión, aunque se considera que simplemente podría definirse como "Conexiones entre Sucursales", para independizarse de los protocolos de comunicaciones utilizados.

5 Conclusiones y Trabajo Futuro

En el presente artículo se ha expuesto una propuesta de Tipificación de las TICs para la evaluación del nivel de desarrollo tecnológico de implementación y uso en la Industria, y la misma se ha sometido a una validación a través de expertos en el tema.

Respecto de los resultados de dicha validación, y observando los porcentajes de los cuadros presentados que avalan el trabajo realizado en este documento, se dan por validadas las clasificaciones propuestas.

Sin embargo, se sugiere tener en cuenta, para una posterior revisión de las taxonomías, los siguientes ítems.

- Virtualización.
- Servicios Cloud.
- Software para sistemas embebidos (Software)
- Mediciones con relación a la eficiencia energética (Software)
- Sistemas SCADA (Software)
- Raspberry Pi (Hardware)
- Equipos de Control Numérico (Hardware)
- Protocolos y software para comunicaciones industriales (Infraestructura)

A partir de la validación, se concluye que el modelo de taxonomías elaborado es adecuado para la medición de las diferentes TICs implementadas en la industria, independientemente que deban ser ampliados y mejorados algunos tipos específicos con los aportes de los expertos.

Por otra parte, se concluye que es necesario diferenciar entre tecnologías existentes y tendencias del desarrollo tecnológico, para poder evaluar en el contexto real de uso cuáles son los niveles de desarrollo tecnológico por ramas de industria según tengan implementadas tecnologías que llevan mucho tiempo de uso en el mercado, tecnologías que llevan un tiempo medio en el mercado o bien las últimas tecnologías que registra el mercado dentro de cada uno de sus tipos a los fines de detectar las necesidades de implementación y las capacidades para innovación .

6 Referencias

[1] Ministerio de Ciencia y Tecnología (2009) - Libro Blanco de la Prospectiva TIC - Proyecto 2020. Buenos Aires.

[2] Ca' Zorzi, A. (2011) - Las TIC en el desarrollo de la PyME: algunas experiencias de América Latina. Centro Internacional de Investigaciones para el Desarrollo - En colaboración con Fondo Multilateral de Inversiones/Banco Interamericano de Desarrollo.

[3] Saavedra García, M. L.; Tapia Sánchez, B. (2013) - El uso de las tecnologías de información y comunicación TIC en las micro, pequeñas y medianas empresas (MIPyME) industriales mexicanas. Enl@ce, 85-104.

[4] Yoguel, G.; Novick, M.; Milesi, D.; Roitter, S.; Borello, J. (2004). Información y conocimiento: la difusión de las tecnologías de información y comunicación en la industria manufacturera argentina. CEPAL, 139-156

[5] Novick, M.; Ritondo, S. (2013). El desafío de las TIC en Argentina. Crear capacidades para la generación de empleo. Santiago: CEPAL.

[6] Zubieta, R.; Villadeamigo, J.; Cianci, L. (2013). Índices de Nivel Tecnológico – Su papel en una Estrategia de Desarrollo. PIUBAD - Simposio VIII. Buenos Aires. Obtenido de http://www.uba.ar/archivos_secyt/image/SIMPOSIO%20VIII%20Documento.pdf

[7] Mon, A.; Del Giorgio, H.; De María, E. (2017). La inserción de las TICs en el desarrollo industrial de La Matanza. 1er Congreso Latinoamericano de Ingeniería (CLADI). Entre Ríos. Argentina.

[8] Del Giorgio, H.; De María, E. Figuerola, C.; Mon, A.; Querel, M. (2016). Evaluación de software y TICs para el desarrollo industrial, el caso Argentino. IEEE, 11 Congreso Colombiano de Computación (11 CCC). Popayán, Colombia.

[9] Mon, A.; Del Giorgio, H.; Querel, M. (2017). Evaluación de software para el desarrollo industrial. Workshop de Investigadores en Ciencias de la Computación (WICC 2017). Buenos Aires, Argentina.

Técnicas de control de concurrencia en base de datos: implementación en un sistema de gestión

Guillermo Arduino, Pedro L. Alfonzo

Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, 9 de Julio 1449, Corrientes, Argentina.
{gaarduino@gmail.com, plalfonzo@hotmail.com}

Resumen. Las transacciones concurrentes en una aplicación que se implementa en una estructura cliente servidor, es de vital importancia, dado que mantener la coherencia, integridad y durabilidad de los datos que son consumidos no es una tarea trivial. La elección de una técnica de manejo de concurrencia que permita la serialización de las transacciones y la coherencia de las operaciones realizadas sobre dichos datos, debe ejecutarse de acuerdo a la naturaleza del problema que se esté tratando de resolver. En el presente artículo, se presenta la aplicación técnicas de concurrencia optimista, al modulo "autogestión de compras" en un Sistema de Gestión de Ventas. Los resultados demuestran que con la aplicación de la técnica mencionada y de acuerdo a operaciones que no suponen un bloqueo, o que el bloqueo resultaría costoso en el rendimiento del sistema, es decir, en tiempo y recursos, se consigue mantener la integridad y coherencia de los datos.

Palabras clave: Transacciones concurrentes, ACID, software de gestión, sistema de gestión de base de datos, base de datos.

1 Introduction

Se define Transacción en un sistema de gestión de base de datos (SGBD) como un conjunto de operaciones que se realizan como una sola unidad lógica de trabajo. Dicha unidad debe cumplir con cuatro propiedades básicas, conocidas como atomicidad, coherencia, aislamiento y durabilidad (ACID), para poder ser clasificada como una transacción [1][2].

1.1 Control de concurrencia

El control de las transacciones debe residir en las aplicaciones, primordialmente al especificar el inicio y fin de las mismas, en puntos que aseguren la coherencia lógica de los datos. Por lo tanto, se deben explicitar las secuencias de modificaciones de datos que los dejan en un estado coherente en relación a las reglas de negocios de la organización [3]. Lo expuesto se logra cumpliendo con las propiedades ACID. A continuación se sintetizan cada una de ellas [4]: i) Atomicidad: una transacción debe

ser una unidad atómica de trabajo, es decir, que se realicen todas sus modificaciones en los datos o no se realice ninguna; ii) Coherencia: cuando una transacción finaliza debe dejar todos los datos en un estado coherente. Deberán aplicarse todas las reglas a las modificaciones de la transacción permitiendo de esa manera mantener la integridad de los datos; iii) Aislamiento: las modificaciones realizadas por transacciones simultáneas se deben bloquear de las llevadas a cabo por otras transacciones simultáneas; iv) Durabilidad: el SGBD asegura que perduren los cambios realizados por una transacción que termina con éxito.

Por otra parte, las transacciones pueden encontrarse en los siguientes estados [3]: i) Activa: durante su ejecución; ii) Parcialmente comprometida: después de ejecutar la última instrucción; iii) Fallida: imposible de seguir ejecutándose; iv) Abortada: transacción que retrocedió y dejó la base de datos restaurada a su estado anterior; v) Comprometida: la transacción ha finalizado correctamente.

En ocasiones los usuarios tienen acceso a los datos de manera simultánea, es decir, leen o modifican los mismos datos al mismo tiempo. Cuando esto no se controla pueden suceder algunas de las siguientes situaciones problemáticas: i) Actualizaciones perdidas: cuando dos o más transacciones seleccionan los mismos datos. Con la última actualización se sobrescriben las actualizaciones realizadas por otras transacciones produciendo perdida de datos [4]; ii) Dependencia no confirmada (lectura no actualizada): cuando una transacción selecciona datos que están siendo actualizadas por otra transacción, lo que conlleva a que la segunda transacción acceda a datos que todavía no han sido confirmados [4]; iii) Análisis contradictorios (lectura irrepetible): cuando una transacción obtiene acceso a los mismos datos varias veces y en cada ocasión accede a datos diferentes. Esto se parece a la dependencia confirmada, la diferencia radica en que los datos ya han sido confirmados por la segunda transacción, con lo cual la lectura se hace irrepetible [1], [4]; iv) Lecturas fantasmas: es una situación particular que se produce en un SGBD cuando se ejecutan dos consultas idénticas al mismo tiempo y la recopilación de los datos devueltos por la segunda consulta es diferente a lo que se obtiene con la primera [1]. Por lo tanto, cuando varios usuarios intentan modificar datos en una base de datos al mismo tiempo, debe implementarse un sistema de control de forma que las modificaciones realizadas por un usuario no interfiera en forma negativa a la de otro usuario. Esto se denomina control de simultaneidad [4].

La teoría del control de simultaneidad tiene dos clasificaciones: i) Control de simultaneidad pesimista [1], [4]: un sistema de bloqueos impide que los usuarios modifiquen los datos de forma que afecte a otros usuarios. Cuando un usuario lleve a cabo una acción que da lugar a que se aplique un bloqueo, los demás usuarios no podrán realizar acciones que crearían conflictos con el bloqueo, hasta que sea liberado. Se utiliza principalmente en entornos donde existen conflictos por la obtención de datos, y en los que el costo de la protección de los mismos con bloqueos es menor que el de revertir las transacciones si se producen conflictos de simultaneidad; ii) Control de simultaneidad optimista [2][5][6]: no se bloquean los datos cuando se acceden a ellos para la lectura. Cuando un usuario realiza una actualización de datos, el sistema comprueba si otro usuario los modificó después de la lectura. Si se actualizaron los datos, se produce un error. Normalmente, el usuario que recibe el error revierte la transacción y comienza de nuevo. Se utiliza

principalmente en entornos donde hay pocos problemas de contención por la obtención de datos y en los que el costo de revertir ocasionalmente una transacción es menor que el de bloquear los datos cuando son leídos.

1.2 Técnicas de control de concurrencia

A continuación se mencionan las técnicas con la que se logran evitar problemas de concurrencia aplicadas a la simultaneidad de transacciones en un SGBD. Como se indicó anteriormente entre las técnicas de concurrencia existen dos grandes grupos [1], [3], [4]: i) Control de simultaneidad pesimista y ii) Control de simultaneidad optimista. En cada grupo se incluyen diferentes técnicas que permiten manejar esa simultaneidad de manera eficiente [1], [7]:

a) Control de simultaneidad pesimista:
   - Técnicas basadas en bloqueos (Fig. 1).
   - Técnicas Multi-Granulares (Fig. 2).
   - Técnicas basadas en marcas temporales (Fig. 3)).

b) Control de simultaneidad optimista.
   - Técnicas basadas en Validación.
   - Técnica original de control de concurrencia optimista.

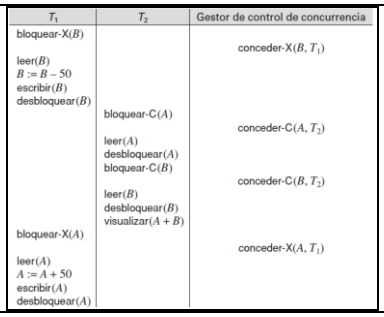


Fig. 1. Técnicas basadas en bloqueos (Fuente: [1])

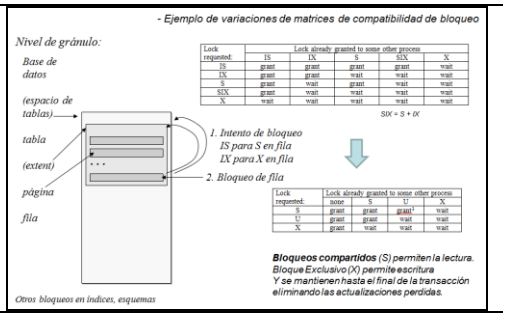


Fig. 2. Técnicas Multi-Granulares (Fuente: [7])

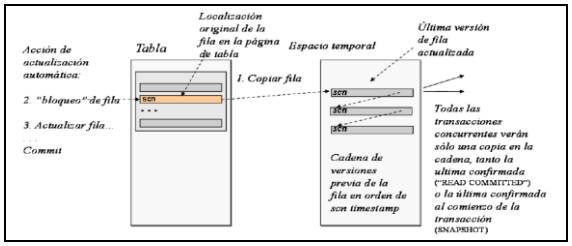


Fig. 3. Técnicas basadas en marcas temporales (Fuente: [7]).

1.3 Las base de datos y las transacciones concurrentes

A continuación se describe la técnica de concurrencia optimista basada en validación utilizado en el presente trabajo. Como se expresó anteriormente se presume que todas las transacciones finalizan, es decir, quedan en un estado transaccional comprometido, pero sujetas a una validación final. Las fases que debe cumplir una transacción son las siguientes [1], [4]: i) Fase de lectura: las transacciones se almacenan de forma temporal sin actualizar la base de datos; ii) Fase de validación: se realiza una prueba para verificar si es factible copiar esas variables locales temporales en la base de datos sin que se produzca alguna transgresión de secuencialidad; iii) Fase de escritura: si no se ha producido ninguna violación de secuencialidad las operaciones se escriben en la base de datos.

Cada transacción puede combinar las tres fases citadas, siempre y cuando respeten ese orden de secuencia en la ejecución de las transacciones concurrentes. Por lo tanto, para validar las transacciones se debe conocer el momento en que tienen lugar las fases. Para ello, se establece el orden de secuencia a través de las técnicas basadas en marcas temporales [1]. En esta técnica la probabilidad de conflictos entre transacciones es muy baja y se obtiene un tiempo de respuesta más rápida y evita los retrocesos en cascada, dado que las escrituras son sólo después de que la transacción que ejecuta la escritura se encuentre comprometida. Sin embargo, existe una posibilidad de que las transacciones extensas nunca sean comprometidas, debido a largas secuencias de transacciones cortas conflictivas que provocan reinicios de aquella transacción. Para evitar esto es necesario bloquear las transacciones conflictivas de forma temporal permitiendo que la transacción afectada sea comprometida [1], [2].

Las aplicaciones prácticas sobre transacciones concurrentes han sido estudiados como temas recientes [1], [2], [4], [5]. Es importante mantener la integridad de un SGBD impidiendo que las transacciones no finalicen en un estado intermedio [2]. En el caso en que se deba cancelarse la transacción, deberá quedar el SGBD en su estado original [4].

Por lo expuesto, en este trabajo se propone implementar transacciones concurrentes en SGBD que permita el acceso de multiples Clientes al Servidor, aplicando control de concurrencia optimista en un módulo de “módulo de autogestión de compras” de un sistema de gestión, garantizando las propiedades ACID.

2 Metodología

La metodología utilizada en la construcción del sistema de gestión se basó en las siguientes etapas:
Etapa 1: Revisión y estudio de técnicas de transacciones concurrentes en base de datos [1], [2], [4], [7].
Etapa 2: Selección de un método de transacción concurrente en base de datos. Se optó por un esquema de validación, que es un método de control de concurrencia optimista, que garantiza el acceso a múltiples usuarios [1], [4], [7].
Etapa 3: Selección y estudio de herramientas [8], [9], [10], [11].

Etapa 4: Desarrollar un sistema de gestión, eligiéndose para tal fin el modelo de prototipo perteneciente al ciclo de vida del modelo Evolutivo [12]. A continuación se menciona algunas actividades desarrolladas:

- Se definieron las funciones principales del sistema: i) Administración y control de usuarios; ii) Administración y ventas y iii) Administración y control de stock.
- Se establecieron los perfiles de usuarios. i) Administrador: gestiona y administra el Sistema; ii) Vendedor: recibe y notifica el pedido terminado y entregado; iii) Cajero: registra los cobros diarios por las ventas; iv) Cliente: realiza el pedido por pantalla táctil de los diferentes tipos de productos.
- Se organizó el sistema bajo el modelo de Cliente-Servidor.
- Se diseño de la base de datos. Se definieron adecuadamente las tablas y sus relaciones.
- Se definieron las funcionalidades del sistema teniendo en cuenta los requerimientos. Se utilizó el Lenguaje de Modelado Unificado (UML) para describirlas, a partir de los requerimientos, utilizándose: i) Casos de Uso; ii) Conversaciones y diagramas de secuencias para detallar la implementación de los Casos de Uso.

Etapa 5: Desarrollo e implementación del módulo de "autogestión de compra" del sistema desarrollado, aplicando un método de transacción concurrente.

Se construyó el prototipo del módulo de "autogestión de compras" con el diseño de interfaces gráficas que permiten interactuar a los actores intervinientes en la concreción de la compra de un producto determinado, como así también la modificación y cancelación del mismo (Fig. 4).

Las conexiones de base datos y las transacciones aplicadas al SGBD se desarrollaron en lenguaje SQL, y se utilizó como gestor Microsoft SQL Server Express 2005. El módulo se construyó en base a los requerimientos y con la modalidad iterativo incremental. En la Tabla 1 se enumeran las funcionalidades desarrolladas en dos etapas iterativas al módulo de autogestión al cual luego se implementó la técnica optimista.

Tabla 1. Funcionalidades incorporadas al módulo de autogestión.

Incrementos	**Funcionalidad**
Primero	a)Seleccionar Productos; b) Seleccionar Gusto; c) Confirmar Compra
Segundo	a)agregar Nuevo producto; b) Agregar Insumos; c) Eliminar Producto

En relación a la evaluación del prototipo se realizaron diversas pruebas que permitieron hacer una evaluación del mismo a fin de detectar y depurar errores que pudieran surgir en el desarrollo del módulo de "autogestión de compras" al cual se aplicó la técnica de transacciones concurrentes. Se aceptó el prototipo siendo implementado el módulo de "autogestión de compra" con la primera iteración y las primeras funcionalidades.

En una segunda iteración se agregaron las funcionalidades restantes que surgieron de la interacción de los usuarios con el módulo mencionado. En esta iteración se implementó la técnica de concurrencia optimista basada en Validación al módulo de "autogestión de compra", para poder controlar los posibles conflictos transaccionales

que generen inconsistencias en la base de datos. Una técnica basada en bloqueos impulsaría una sobrecarga innecesaria del sistema.

En este sentido se implementaron tres fases bien definidas[1],[7]: i) Fase de lectura: donde se leen los productos seleccionados, sin actualizar la base de datos; ii) Fase de validación: al momento de ejecutar la funcionalidad de "Confirmar la Compra" se realizó una prueba de validación, para determinar la modificación de la base de datos en la operación escribir sin alterar ni vulnerar la secuencialidad; iii) Fase de escritura: al confirmarse la compra del producto y la transacción fue correcta en la fase de validación entonces las modificaciones son aplicadas a la base de datos, si la situación explicada anteriormente fuera contraria, se retrocede como si la transacción nunca se hubiese realizado.

Con la aplicación de la técnica mencionada anteriormente y teniendo en cuenta la naturaleza del módulo de autogestión desarrollado, se buscó que las transacciones se ejecuten presumiendo que todas fueron capaces de finalizar la ejecución y ser validadas al final.

3 Resultados

A los efectos de ilustrar la técnica objeto de estudio, en esta sección se describe la aplicación de la técnica de transacciones concurrentes optimistas al módulo "gestión de compras" (Fig. 4), teniendo en cuenta las consideraciones de crear el modelo Cliente Servidor y la forma de poder establecer la conexión a la base de datos, como así también la configuración de la misma para aceptar este tipo de transacciones.

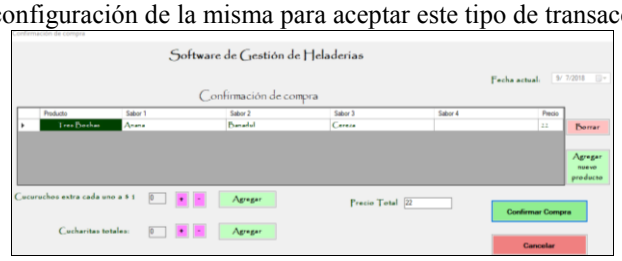


Fig. 4. Pantalla del módulo "Gestión de compras"

Las características del Servidor de Aplicación se observa en la Tabla 2 las cuales también son compartidas por el Servidor de base de datos, los cuales afines demostrativos conviven en el mismo host.

Tabla 2. Características del host servidor

Característica	Descripción
Procesador	AMD C50 DOBLE NÚCLEO 1.0 GHZ
Memoria Ram	DDR 2 4GB
Red Lan de Prueba	a) SSID: Arduino. b) Rango: 192.168.1.1 - 192.168.1.63 c) Asignación de IPs: Manual
Dirección IP	192.168.1.50/24
Sistema Operativo Nativo	Windows 7 32 bits Servipack 1
Nombre de Equipo	Usuario-PC

En la Tabla 3 se visualizan las características correspondientes a los Hosts donde se ejecutó la aplicación cliente para las pruebas de la técnica objeto de estudio. Para lo cual, se utilizaron equipos de características simples ya que al aplicación cliente en si no presenta mayores requerimientos de hardware.

Tabla 3. Características del host cliente

Característica	Descripción
Procesador	AMD ATLHON X II 255 3.01 GHZ
Memoria Ram	DDR 2 4GB
Red Lan de Prueba	a) SSID: Arduino. b). Rango: 192.168.1.1 - 192.168.1.63 c) Asignación de IPs: Manual
Dirección IP Equipo 1	192.168.1.2/24
Dirección IP Equipo 2	192.168.1.10/24
Sistema Operativo Nativo	Windows 7 32 bits Servipack 1
Nombre de Equipo 1	Usuario-PC-1
Nombre de Equipo 2	Usuario-PC-2

Para poder configurar la conexión remota al Servidor SQL Server, se estableció la habilitación de conexión remota al mismo y el modo de Autenticación Mixta con un usuario y contraseña en SQL Server. Posteriormente se establecieron los elementos que componen la cadena de conexión utilizada en la aplicación para enlazar la base de datos con la misma, así como el usuario de SQL Server para el acceso remoto, como ser: i) el Data Source; ii) el InitialCatalog; iii) la Seguridad iV) Mixed Authentication (SQL Authentication) y v) el Network Address.

Establecidos estos parámetros se procedió a habilitar la base de datos con aislamiento Snapshot [13].

Se utilizó este tipo de aislamiento porque representa el uso de la técnica de concurrencia optimista sin utilización de bloqueos. Para ello verifica que la transacción tenga una versión coherente de aquellos datos con la que la misma se inicia. Es como si se tomará una instantánea de cada transacción.

Por lo expuesto se logra que no se bloqueen las escrituras de datos de otras transacciones y tampoco son bloqueadas las lecturas. Para poder utilizarlo se debió elegir la opción ALLOW_SNAPSHOT_ISOLATION establecida en modo ON. Si se realiza una operación UPDATE en una tabla y al instante se emite una operación de selección para esa misma tabla los datos modificados se incluyen en el conjunto de resultados obtenidos.

Para comprender el nivel de aislamiento, en la Fig. 5 se observa donde las versiones de filas actualizadas de cada transacción se mantienen en una tabla temporal. Cada transacción es marcada con un número único de secuencia, los cuales son registrados para cada versión de fila. La transacción trabaja con la versión de fila más reciente, la cual posee un número de secuencia anterior al número de secuencia propio de la transacción ejecutada, y omite cualquier versión de fila más nueva que haya sido creada con posterioridad a su comienzo.

A los efectos de aplicar la técnica de concurrencia optimista, se estableció un acceso restringido como se visualiza en la Fig. 6, basándose en conectividad por protocolo TCP/IP. Lo cual genera que cualquier modulo cliente que intente acceder a la base de datos no podrá hacerlo mientras el Servidor de Aplicación este inactivo.

Cuando este tome actividad se comunicará con el Servidor de base de datos, en este caso SQL Server el cual iniciará las conexiones pertinentes y ejecutará las peticiones de transacciones que le sean requeridas.

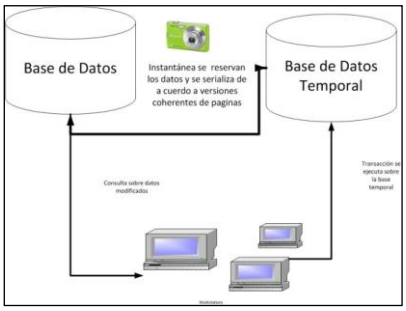


Fig. 5. Aislamiento Snapshot

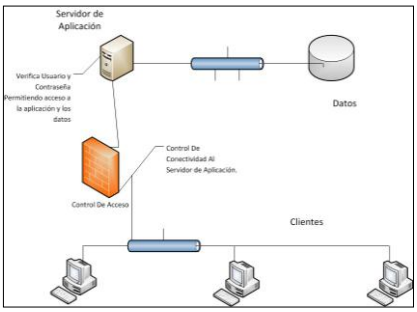
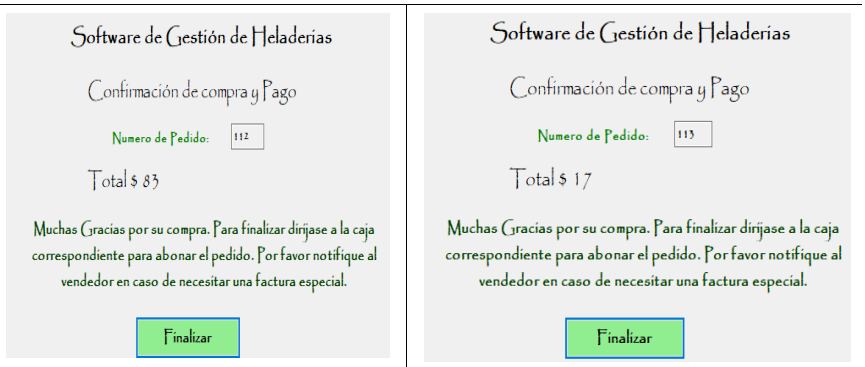


Fig. 6. Control de acceso Cliente Servidor

Una vez que se ingresó al sistema y establecida la conexión se procedió a ejecutar las pruebas de transacciones concurrentes, probando el camino exitoso, es decir, que dos transacciones se pueden ejecutar concurrentemente, superando las validaciones y confirmando ambas transacciones como puede observarse en la Fig. 7. En el caso que una transacción no supere la etapa de validación, la transacción queda anulada y las propiedades ACID se cumplen, como puede observarse en la Fig. 8.

Adicionalmente, se almacenaron en la base de datos información de las transacciones que han cumplimentado las validaciones y han sido confirmadas, como así también aquellas que no lo hicieron. En la Fig. 9 se puede observar a modo de ejemplo las transacciones efectuadas por fecha, disponiendo además, de las opciones de filtrar por dirección IP o por estado (completas o incompletas).



a) Primera terminal remota b) segunda terminal remota

Fig. 7. Mensajes de transacciones exitosas en las terminales.

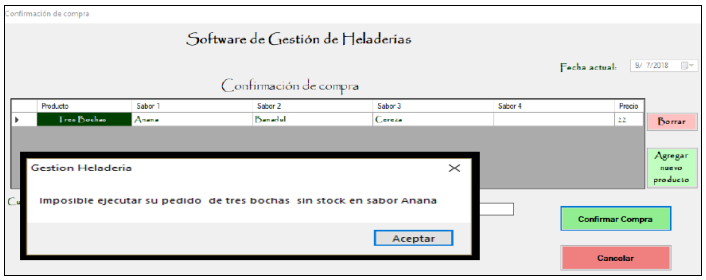


Fig. 8. Transacción fallida con rollback sin superar validación.

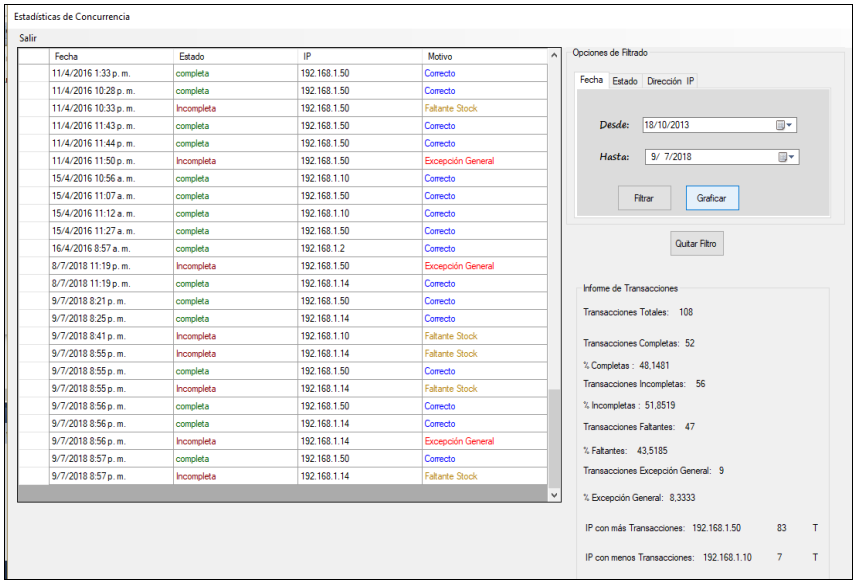


Fig. 9. Lista de transacciones filtradas por fecha.

4 Conclusiones

El estudio de las transacciones concurrentes en una aplicación que se implemente en una estructura cliente servidor, es de vital importancia, dado que mantener la coherencia y durabilidad de los datos que son consumidos desde un SGBD no es una tarea trivial. La elección de una técnica de manejo de concurrencia que permita la serialización de las transacciones y la coherencia de las operaciones realizadas sobre dichos datos, debe realizarse de acuerdo a la naturaleza del problema que se esté tratando de resolver.

En el software desarrollado la naturaleza misma del problema fue la lectura de múltiples datos por las distintas instancias del módulo de “autogestión de compra”, sin necesidad de hacer operaciones de actualización hasta la confirmación y facturación del pedido, lo cual conllevo a que al ser gran parte de las operaciones

aplicadas de lectura, los datos o las tuplas que conforman los registros en la base de datos no fueran necesariamente bloqueadas, sino que simplemente se permitiera su lectura y tratamiento y luego de un proceso de validación las operaciones que modificaran dichos datos fueran aceptadas o si hubiera algún impedimento dichas transacciones se anularan y se volviera la base de datos su estado original. Esto es en definitiva lo que se llama una Técnica de Concurrencia Optimista, siendo lo que se aplicó en este trabajo. Lo que se logra es que las conexiones entre cliente y servidor sean mucho más fluidas con una mayor atención de clientes en menor tiempo, siendo innecesaria una conexión persistente con el servidor, como así también el menor consumo de recursos del mismo.

Lo expuesto, puede observarse en los resultados obtenidos al mostrar como una operación que fue aceptada mucho después que otra, pero que por su tratamiento de los datos que capturó en forma temprana le correspondió insertarse antes, a pesar de que otra transacción haya finalizado primero, lo cual significa que retrasa las actualizaciones permitiendo mantener la coherencia de los datos contenidos en la base de datos (Fig. 9).

Referencias

1. Silberschatz, A., Korth, F., Sudarshan, S.: Fundamentos de Base de Datos, Madrid: Mcgraw-Hill/Interamericana De España, S. A. U., 2002.
2. Simultaneidad optimista. https://msdn.microsoft.com/es-es/library/aa0416cz%28v=vs.110%29.aspx?cs-save-lang=1&cs-lang=vb#code-snippet-4.
3. Neira Cisterna, M.: Métodos de Optimización de Consultas para el Lenguaje SQL, https://docplayer.es/13940333-Metodos-de-optimizacion-de-consultas-para-el-lenguaje-sql.html
4. SQL Server Transaction Locking and Row Versioning Guide, https://technet.microsoft.com/en-us/library/jj856598(v=sql.110).aspx
5. Frati, F. E.: Evaluación de técnicas de detección de errores en programas concurrentes. Trabajo Final presentado para obtener el grado de GRID, Facultad de Informática Universidad Nacional de La Plata, La Plata Pcia de Bs As Argentina, 2014.
6. Buckle, C. E., Urriza, J. M., Paez, F.: Transitando hacia las Bases de Datos de tiempo real, 39JAIIO - JII, Universidad Nacional de La Patagonia San Juan Bosco - Puerto Madryn, Argentina, 2010.
7. Martti, L., Dervos, D. A., Silpiö, K.: SQL Transactions Teoría y ejercicios en la práctica, Universidad de Málaga: DBTech VET, 2014.
8. Enterprise Architect , http://www.sparxsystems.com/products/ea/7.5/
9. Start Uml, http://staruml.io/faq
10. Microsoft Visual Studio, https://www.microsoft.com/en-us/download/details.aspx?id=23507
11. Microsoft SQL Server 2005 Express Edition, http://www.microsoft.com/es-ar/download/details.aspx?id=21844
12. Somerville, I.: Ingeniería del Software, Madrid: Pearson Educación S.A., 2005.
13. Set Transaction Isolation Level, https://msdn.microsoft.com/es-es/library/ms173763%28v=sql.120%29.aspx

Prototipo de dispositivo de ubicación a través de puntos estratégicos para personas no videntes o con discapacidad visual

Guillermo Arispe [1,3], Claudio Aciti [1,2], Matías Presso[1,2,3] y José Marone [2]

[1] Universidad Nacional de Tres de Febrero, Sede Caseros I.
Valentín Gómez 4828, Caseros (B1678ABJ), Buenos Aires, Argentina
[2] Universidad Nacional del Centro de la Provincia de Buenos Aires.
Pinto 399, Tandil (7000), Buenos Aires, Argentina
[3] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires
Calle 526 entre 10 y 11, La Plata (1900), Buenos Aires, Argentina
{guillermo.arispe@gmail.com, caciti@exa.unicen.edu.ar, matiaspresso@gmail.com, marone@exa.unicen.edu.ar}

Abstract. El presente artículo describe el diseño y desarrollo de un sistema destinado a guiar a personas no videntes o con discapacidad visual a través de diferentes entornos en una ciudad. El sistema consiste en una aplicación móvil que orienta al usuario por medio de audios y un conjunto de dispositivos electrónicos geoposicionados en puntos estratégicos de una ciudad. Dichos dispositivos brindan información del entorno a la aplicación de manera inalámbrica aportando accesibilidad y simpleza. Para este trabajo se implementó un prototipo totalmente funcional, donde la aplicación móvil fue desarrollada para Android, mientras que los dispositivos fueron creados sobre la plataforma Arduino.

Keywords: Discapacidad Visual, Ciudades Inteligentes, Ciudades Inclusivas.

1 Introducción

Con el desarrollo económico y el proceso de urbanización generalizado, cada vez es mayor el porcentaje de la población mundial que vive en las ciudades. Cómo están diseñadas estas ciudades impactan en cómo se relacionan sus ciudadanos y cómo se desarrolla la actividad económica y social. Es decir, cómo sea la ciudad en la que vivimos influye en la calidad de vida de sus habitantes. En la actualidad está adquiriendo una notable trascendencia el concepto de ciudades inteligentes, aquellas que usan la tecnología para incrementar la eficiencia y sostenibilidad de sus servicios.

Pero, también es necesario tener en cuenta las ciudades inclusivas y accesibles. El diseño y funcionalidad de la ciudad en la que vive una persona con discapacidad física o intelectual va a facilitar o impedir su inclusión en la sociedad, sus posibilidades de encontrar trabajo y de ser lo más independiente posible. [1][2][3]

Existen tres aspectos fundamentales en una ciudad que impactan en la vida de las personas con discapacidad: [4]

Los aspectos físicos permiten que la persona pueda desplazarse de un lugar a otro en la ciudad. Esto es fundamental para poder ir al trabajo, a estudiar, realizar actividades de ocio, ir al médico, es decir, para poder realizar las actividades cotidianas. Algunos elementos básicos son unas veredas anchas (para sillas de ruedas), lisas, con rampas, sin postes en lugares no habituales; o estaciones de trenes con ascensores o escaleras automáticas; semáforos con indicaciones de voz; carteles con información en braille y con imágenes simples. Estos elementos permiten que las personas con discapacidad se muevan libremente y de forma segura por la ciudad.

Los aspectos económicos tienen en cuenta a las personas con discapacidad en el diseño de oportunidades laborales y en el diseño de espacios y productos de consumo. Como ejemplos se pueden citar puestos de trabajo adaptados; edificios accesibles y con baños adaptados para sillas de ruedas; espacios sin colores estridentes, o insonorizados; centros comerciales con puntos de información y señalización sencilla. Esto permite que las personas con capacidades diferentes, físicas e intelectuales, puedan aportar como trabajadores y consumidores.

Los aspectos sociales son aquellos que permiten que todos los ciudadanos compartan los mismos lugares, facilitando la interacción y la participación en la comunidad. Por ejemplo, parques con hamacas adaptadas; museos con guías formados sobre discapacidad; colegios de integración etc.

Si bien adaptar una ciudad a las personas con discapacidad supone cuantiosas inversiones, los beneficios pueden ser disfrutados por todos los ciudadanos. Las rampas en veredas y acceso a edificios permitirán una movilidad más sencilla a personas mayores y a padres con carritos de bebé, por ejemplo. La señalización sonora o con imágenes sencillas permitirá a niños y ancianos ubicarse fácilmente. Los lugares públicos accesibles y adaptados permitirán una mayor riqueza social a todos los ciudadanos. Medidas sencillas que permiten que todos los habitantes de una ciudad puedan aportar y disfrutar de ella.

Las personas no videntes o con problemas severos de visión tienen grandes problemas de movilidad por la ciudad y disponen de una autonomía reducida, lo que hace que en pocas ocasiones abandonen por voluntad propia entornos que conocen bien. Están limitadas en la realización de tareas y esto afecta su calidad de vida, así como sus posibilidades de interacción con el mundo que los rodea[1].

Desplazarse por un centro urbano puede ser una tarea complicada y a veces hasta abrumadora. La ansiedad, estrés y frustración que genera intentar trasladarse por una ciudad que no está adaptada a sus necesidades genera que miles de personas no videntes a nivel global opten por no salir de sus casas si pueden evitarlo.

Aún así, con el bastón como instrumento básico para orientarse y con un perro guía los más afortunados, las personas no videntes se encuentran con obstáculos que no siempre pueden detectar con facilidad. Ejemplos de estos obstáculos son las obras tanto planificadas como urgentes/accidentales que alteran sus entornos conocidos. [5]

En el año 2014, la OMS (Organización Mundial de la Salud) realizó un estudio [6] el cual muestra que existen alrededor de 285 millones de personas no videntes en el mundo, de las cuales 39 millones son ciegas y 246 millones presentan baja visión.

[1] https://www.livestrong.com/article/241936-challenges-that-blind-people-face

Para estas personas, tareas cotidianas como sortear baches inesperados en veredas, realizar una simple compra en un supermercado, ingresar o salir en una estación de subte, tomar un colectivo y bajar del mismo en el lugar correspondiente, cruzar una calle o simplemente caminar hacia su destino se tornan difíciles y, en la mayoría de las ocasiones, hasta peligrosas.

Al ritmo que avanza la tecnología, las ciudades inteligentes van a ser cada vez más frecuentes, incorporando tecnología en diferentes ámbitos que las transforman en localidades más eficientes en el uso de sus recursos y mejoran los servicios entregados. En este tipo de ciudades se busca solucionar los principales problemas a los que se ven enfrentados los ciudadanos cada día, mejorando su calidad de vida.

2 Objetivos

2.1 Objetivo general

El principal objetivo del trabajo es suministrar una herramienta que permita guiar mediante información obtenida de geolocalización y dispositivos colocados en puntos estratégicos a la persona no vidente a través de diferentes ambientes, pudiendo ser estos conocidos o desconocidos para la persona que utilice la aplicación. No se busca reemplazar sus herramientas actuales como el bastón o los perros guías, sino complementar las mismas.

2.1 Objetivos específicos

Desarrollar una aplicación simple, lo más elemental posible y con una interfaz que pueda ser utilizada por personas no videntes o con discapacidad visual.

Detectar la ubicación del usuario de manera continua para saber cuando deben buscarse placas alrededor.

Realizar la conexión entre el teléfono con Android y la placa Arduino, cuando ambos dispositivos estén próximos.

Procesar los datos obtenidos para brindar información de ubicación y guía a la persona no vidente o con discapacidad visual.

3 Alcance y Limitaciones

El prototipo fue desarrollado en Android y se usaron placas Arduino en diferentes puntos estratégicos. Los dispositivos móviles necesarios deben contar con Bluetooth, sistema de GPS y poseer una versión del sistema operativo dentro del rango 4.0.3 hasta 7.1.2. Por lo tanto, los usuarios no tienen que contar con un dispositivo específico de gama alta con características que se vean reflejadas en el precio del mismo.

Entre las principales limitaciones que posee el prototipo, se encuentran: la imposibilidad de conectar múltiples dispositivos móviles al mismo tiempo al módulo Bluetooth de la placa Arduino ubicada en un punto estratégico, la baja intensidad de la señal GPS recibida desde satélites y la duración de la batería del dispositivo.

4 Solución Propuesta

La solución propuesta es una aplicación para dispositivos móviles con sistema operativo Android que ayuda a personas no videntes o con discapacidad visual, brindándoles información de guía y ubicación a través de placas o microcontroladores Arduino que se encuentran ubicados en puntos estratégicos (Fig 1.).

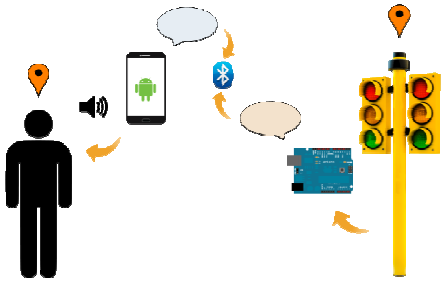


Fig. 1 Diagrama de funcionamiento, tomando como ejemplo de punto estratégico un semáforo

Estos puntos estratégicos pueden ser, por ejemplo:

Semáforos: En un semáforo se puede obtener información del estado actual, del segundero, una georeferenciación del lugar, el ancho de la calle, si es mano y/o contramano, etc.
Señalización de obras: Las obras realizadas por el municipio, que tienen que romper la vereda o la calle, pueden colocar una placa para que le brinde información a la persona no vidente que se acerca. Esta placa es usada en el lugar mientras dure la obra y luego puede ser reutilizada en otra.
Góndolas de supermercados: Puede brindar información de los productos de las góndolas, la ubicación de las cajas, la salida, los baños, etc
Estaciones de subte: Puede brindar información relacionada a la estación del subte: entradas, salidas, para qué estación se dirige, cuánto resta para el próximo subte, etc
Paradas de colectivo: Puede brindar información relacionada a la parada: a qué línea de colectivo pertenece, hacia donde vá, el recorrido, cuánto falta para que llegue.

Antes de establecer la comunicación Bluetooth y comenzar a interactuar con el microcontrolador, la aplicación detecta de manera continua la actividad del usuario, la

cual consiste en saber si el usuario está en movimiento o no. Cuando el usuario no está en movimiento o vuelve a su estado de reposo, la aplicación desactiva el Bluetooth, si estuviese activado, para reducir el consumo de batería del dispositivo. Para el caso contrario, cuando detecta que el usuario comienza a moverse, obtiene la ubicación del individuo. Luego de tomar al menos dos mediciones, compara la distancia recorrida por el usuario y si la diferencia es de al menos 10 metros, habilita el Bluetooth y comienza a buscar una placa cercana a la cual conectarse y solicitarle información. Si no encuentra ninguna, luego de los próximos 10 metros intenta nuevamente. Una vez encontrada alguna de las placas ubicadas en algún punto estratégico, inicia el proceso de conexión. En caso de fallar, reintenta establecer la conexión 3 veces más. Si luego de estos intentos la aplicación no logra conectarse a la placa, notifica al usuario que ha ocurrido un error y que no es posible brindarle información en ese momento. En caso de lograr una conexión exitosa, envía el primer comando para sincronizarse con la placa y espera la confirmación de la misma. Una vez realizado el handshake, la aplicación solicita la información del contexto a la placa y al recibir dicha información -en formato JSON- la interpreta y convierte para comunicarla al usuario.

En cada una de las etapas de la comunicación, la aplicación informa mediante notificaciones de voz al usuario todo lo que ocurre, desde que se intenta establecer una conexión hasta que la misma finaliza. Esto se realiza debido a que puede surgir un error en alguna de las etapas como por ejemplo, el establecimiento de la conexión. Si no se notifica al usuario, este no se entera que ocurrió un error y puede suponer que todavía se está realizando el intento de conexión y se lo deja esperando.

Teniendo en cuenta la discapacidad visual de los usuarios a los que está dirigido este prototipo, todo el proceso antes mencionado es realizado automáticamente en segundo plano sin que el usuario tenga que interactuar con la aplicación.

Además, la aplicación ofrece la posibilidad al usuario de realizar la búsqueda de placas cercanas de manera manual, mediante un movimiento de "Shake"[2] (Ver Fig. 2). Es decir que si el usuario toma su dispositivo y lo agita de un lado a otro, la aplicación comenzará a buscar placas cercanas a las cuales conectarse.



Fig. 2 Indicación de movimiento de "Shake".

Otra característica importante que posee la aplicación es que se ejecuta automáticamente cuando el dispositivo es reiniciado, para seguir brindando información al usuario sin que este tenga que iniciarla manualmente.

[2] Del inglés: batido

4.1 Configuración del Hardware

La conexión entre el microcontrolador o placa Arduino y el módulo Bluetooth HC-05 se realiza de acuerdo al siguiente esquema:

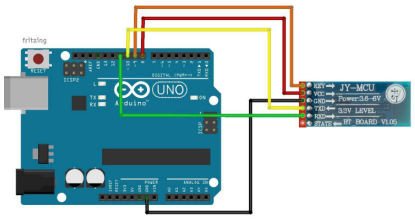


Fig. 3 Esquema de conexión entre la placa y el módulo.

La descripción de las conexiones se pueden observar en la Tabla 1.

Módulo HC-05	Arduino Uno
Tx - Transmisor	Pin 10 Receptor SoftwareSerial
Rx - Receptor	Pin 11 Transmisor SoftwareSerial
GND	GND
VCC	3,3 V

Tabla 1 Descripción de conexiones entre la placa Arduino y el módulo HC-05.

4.2 Software Embebido en microcontrolador

Para la placa Arduino se desarrollaron dos softwares en el lenguaje de programación Processing utilizando el IDE oficial provisto por Arduino. Uno de ellos se encarga de realizar la configuración del módulo Bluetooth. Este primer componente de software se instala, realiza la configuración y luego es reemplazado por el segundo, el cual se encarga de la comunicación Bluetooth, procesar los comandos recibidos desde la aplicación Android y enviar las respuestas correspondientes a la misma.

El programa utiliza la librería SoftwareSerial provista por Arduino y configura el módulo Bluetooth para que trabaje como un puerto serie, es decir, que la transmisión se realiza de manera secuencial, bit a bit. Una vez establecido el puerto serie, realiza el envío de comandos AT al módulo para configurar sus diferentes parámetros. Luego de enviar un comando, espera a que el módulo responda para saber si la configuración que se intenta hacer fue realizada con éxito. Si es así, continúa con el siguiente comando.

El segundo programa de los que se mencionaron anteriormente cuenta con dos métodos principales y obligatorios en todo programa de Arduino. Estos son, setup() y loop(). El primero, es la primer función que se ejecuta y donde se realiza la configuración inicial del módulo Bluetooth (configurándolo como puerto serie de la misma manera que el primer programa). En esta configuración, se establece el

nombre de la red BLINDY_PARTNER, se configura la contraseña de la misma, se establece el rol del módulo como esclavo para que los dispositivos que actúan como maestros se conecten y se cambia la velocidad de modulación a 115200 baudios, ya que es la velocidad necesaria para recibir correctamente los caracteres desde la aplicación Android.

El segundo método, como precisamente su nombre lo indica, es un loop o bucle que se ejecuta un número infinito de veces hasta que el microcontrolador es apagado o reiniciado. En este bucle, se verifica todo el tiempo si hay algo para leer. En el momento que se encuentra un carácter, se ejecuta la función procesarComando (char comando) que se encarga de tomar el comando recibido desde la aplicación, procesarlo y enviar una respuesta de acuerdo al mismo.

4.3 Aplicación Android

La aplicación fue desarrollada en el entorno o IDE Android Studio utilizando principalmente el lenguaje de programación Java y XML para los pocos elementos visuales que posee la aplicación. Sus componentes principales son: Actividad Principal, Servicio Principal y Broadcast Receiver

Actividad Principal: La ActividadPrincipal es la pantalla que se muestra al iniciar la aplicación. La misma no posee ningún componente con el cual el usuario pueda interactuar, ya que se busca manejar todo de manera automática.

Cuando el usuario instala la aplicación e ingresa por primera vez, se le solicita que conceda los permisos para utilizar el GPS y administrar la configuración Bluetooth. Estas acciones son las únicas que requieren intervención por parte del usuario. Luego, la ActividadPrincipal disminuye el brillo de la pantalla al mínimo para consumir menos batería y ejecuta el ServicioPrincipal, el cual comienza su ejecución en segundo plano.

Servicio Principal: Es iniciado desde ActividadPrincipal, y continúa ejecutándose en segundo plano aunque el usuario cierre la aplicación o cambie a otra.

A continuación se describen los componentes y las acciones del ServicioPrincipal, el cual puede considerarse el elemento más importante de la aplicación.

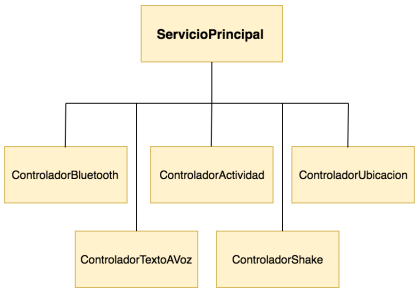


Fig. 4 Diagrama de componentes de ServicioPrincipal.

Cada uno de los controladores que componen el ServicioPrincipal tiene asignada una responsabilidad específica.

El ControladorActividad tiene la responsabilidad de detectar la actividad del usuario que, como se mencionó anteriormente, consiste saber cuando está en movimiento. Para esto utiliza Google Awareness API[3] la cual permite acceder a una variedad de señales contextuales como por ejemplo, el uso de batería o memoria, la ubicación actual del dispositivo, las condiciones climáticas a partir de esa ubicación, y entre otras, la actividad del usuario utilizada en este proyecto.

Al detectar que el usuario está en movimiento, el ServicioPrincipal comienza a obtener la ubicación del dispositivo mediante el ControladorUbicacion para luego calcular la distancia recorrida por el usuario. Si esta distancia es de al menos 10 metros, utilizando el ControladorBluetooth enciende el Bluetooth del dispositivo y realiza una búsqueda para encontrar placas Arduino cercanas a las que va a intentar conectarse para obtener información del contexto y notificarsela al usuario.

Las placas a las cuales se conecta la aplicación son aquellas que pertenecen a la red denominada BLINDY_PARTNER. La Fig. 5, describe el proceso de comunicación. Al encontrar una inicia el proceso de conexión y una vez realizado con éxito envía un CMD_INIT (0) para comenzar la comunicación. A este comando la placa Arduino responde con un EACK indicando que está sincronizada y lista para comenzar a transmitir su información. El primer caracter, 'E', es utilizado dentro de la aplicación para diferenciar si la respuesta es un mensaje de estado o de información, para el cual se utiliza el caracter 'I'. Al recibir el ACK, la aplicación envía un segundo comando: CMD_REQUEST_INF (1) para solicitar información sobre el contexto donde está ubicada la placa.

Esta información es enviada en formato JSON a la aplicación que, al recibirla, la parsea y la convierte a voz para transmitirla al usuario a través de los altavoces del dispositivo o los auriculares. Esta conversión es responsabilidad del ControladorTextoAVoz. Con esta información, la persona no vidente o con discapacidad visual puede tener una idea del contexto que lo rodea y tomar una decisión a partir de ello. Una vez transmitida la información al usuario, la aplicación envía un último comando: CDM_FIN (9), el cual indica a la placa que finaliza la

[3] https://developers.google.com/awareness/

comunicación y como respuesta a esto, la placa responde con un EFIN que la aplicación utiliza para proceder a la desconexión.

Luego de finalizar la comunicación, el Bluetooth es deshabilitado para que el mismo no siga consumiendo energía de la batería del dispositivo y de esta forma hacer un uso eficaz de la misma.

Broadcast Receiver: Este componente se utiliza para la situación en la que el dispositivo es reiniciado, permitiendo a la aplicación ejecutarse por sí sola y seguir brindando información de asistencia al usuario sin que este tenga que hacerlo manualmente.

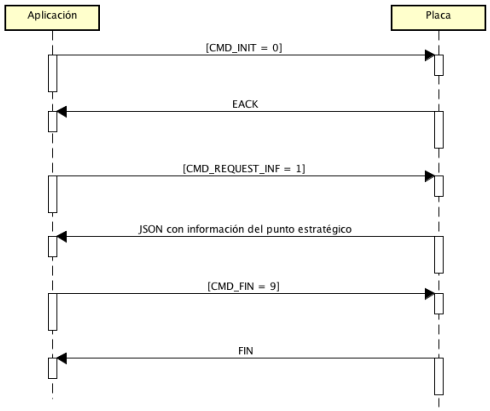


Fig. 5 Diagrama de secuencia - Protocolo de comunicación.

5 Conclusiones

La aplicación móvil fue probada en diferentes dispositivos y funcionó muy bien, sin importar el dispositivo Android que se utilice, siempre y cuando se cumplan los requisitos mínimos mencionados en el apartado de Alcance.

De acuerdo a las pruebas realizadas, se observó que el rango de conexión Bluetooth no representa un problema. La tecnología Bluetooth aplicada en la mayoría de los dispositivos actuales, posee un rango de cobertura que va desde los 60 metros en dispositivos con Bluetooth 4.0 y hasta 240 metros en dispositivos con Bluetooth 5.0. Como se mencionó, cada 10 metros recorridos por la persona la aplicación realiza la búsqueda de alguna placa cercanas a las cual conectarse.

La información almacenada en las placas es totalmente editable, por lo que no es un impedimento cargarla en diferentes idiomas para poder utilizar la herramienta en cualquier ciudad del mundo.

Luego de desarrollar las diferentes etapas del prototipo, se concluye que se ha desarrollado una herramienta capaz de brindar información contextual que permita ubicar y guiar a una persona a medida que se desplaza.

6 Trabajos Futuros

Entre los trabajos futuros se destacan, la opción de controlar la aplicación a través de comandos de voz; Permitir conexiones Bluetooth simultáneas para atender varias peticiones al mismo tiempo; Realizar un mapa de puntos estratégicos que permitirá al usuario conocer cuál es el punto más cercano donde se encuentra colocada una placa Arduino y de esta forma guiarlo hasta el mismo para poder brindarle información del ambiente y orientarlo hacia su destino; Y la funcionalidad "Me perdí". Mediante solicitudes de voz, el usuario podría indicarle a la aplicación que está perdido y la misma lo guiaría al punto estratégico más cercano para poder orientarlo y que retome el camino hacia su destino. Para esto, es necesario registrar los puntos estratégicos donde el usuario se conectó a una placa Arduino durante todo su trayecto.

7 Agradecimientos

Los autores agradecen a la CIC PBA, donde M. Presso pertenece a la carrera de Profesional de Apoyo a la Investigación.

Referencias

1. Erin Brady, Meredith Ringel Morris, Yu Zhong, Samuel White, Jeffrey P. Bigham (2013). Visual Challenges in the Everyday Lives of Blind People. Investigación de Microsoft. Universidad de Rochester, Estados Unidos.
2. Hersh, Marion y Johnson, Michael (2010). Assistive technology for visually impaired and blind people. s.l. : Springer, 2010
3. Bernardelli Claudio, Alvarez Alejandra (2017). La movilidad y la ciudad inteligente: accesibilidad en la cadena de información y comunicación en el transporte público y su entorno, para la inclusión de las personas con discapacidad y adultos mayores. Universidad de Buenos Aires (UBA, FADU).
4. David Korngold, Martin Lemos y Michael Rohwer de BSR en conjunto con financiación de AT&T. Smart Cities for All: A Vision for an inclusive, Accessible Urban Future
5. Blindness: Challenge and Achievement. World Access For The Blind. https://waftb.net/blindness-challenge-and-achievement (https://waftb.net/)
6. Organización Mundial de la Salud, Ceguera y Discapacidad Visual: http://www.who.int/mediacentre/factsheets/fs282/es/
7. Documentación oficial de Activities en Android https://developer.android.com/reference/android/app/Activity.html?hl=es-419
8. Documentación oficial de Services en Android https://developer.android.com/reference/android/app/Service.html?hl=es-419
9. Documentación oficial de Broadcast Receivers https://developer.android.com/reference/android/content/BroadcastReceiver.html

Towards a Visual SPARQL-DL Query Builder

Christian Gimenez†, Germán Braun†, Laura Cecchi†, and Pablo Fillottrani‡

†Universidad Nacional del Comahue
Buenos Aires 1400, Neuquén Capital, Argentina
‡Universidad Nacional del Sur
Avenida Alem 1253, Bahía Blanca, Buenos Aires, Argentina
{christian.gimenez,german.braun,lcecchi,}@fi.uncoma.edu.ar
prf@cs.uns.edu.ar

Abstract. Querying ontologies is an every-day activity that users need. This interaction will improve when the query is more expressive and easier to develop. For this purpose, a visual query language is an ideal mean for users and ontology engineers for creating queries taking advantage of the easy-to-understand and low time and cost characteristics, specially, for users which does not know textual query languages. On the other side, SPARQL-DL is a powerful and expressive textual query language for OWL-DL based ontologies that can combine TBox/ABox/RBox queries. Considering the advantage of both, we present in this work a visual query language that can be interpreted as SPARQL-DL sentences and thus being used for querying ontologies for its structure and/or instance information. Altogether, we use this idea to create a modified version of *crowd*, a Web modelling tool with reasoning support, that enables to implement and tests the presented graphical language along with the needed SPARQL-DL support for solving queries with the user's provided OWL 2 ontologies in any of its linearisations.

Keywords: Ontology, VQL, Semantic Web, SPARQL-DL, crowd

1 Introduction

The Semantic Web [1, 2] includes numerous ontologies that can be mixed semantically. These ontologies provide support for data and meaning for each knowledge base that an organisation can provide in its Web page. The usability of this information is key for these organisations which queries these data with no limitation in expressiveness. Specially, using both levels of DL: structural and assertional. Once created the ontologies, ontology engineers require a query formulation tool that easily express the user's information needs in terms of queries over ontologies. Examples of this are OptiqueVQS [3], OntoVQL [4], etc.

Not less important is to provide tools to users and domain experts for creating and accessing such information they need in a proper way. Querying using visual techniques provides a layer of abstractions that helps to understand easily the entire ontology. *crowd*[1] [5] is a Web tool for graphical ontology modelling created

[1] http://crowd.fi.uncoma.edu.ar

for achieving that purpose. It has been designed for being extendable to other visual language and for supporting reasoning services such as consistency checking. Nowadays, it provides support for standard conceptual modelling languages like UML and EER. Currently, ORM [6] and OVM [7] is under development. Also, another version of *crowd* has been released with support for the KF-metamodel [8, 9] which allows the user to create a model in one language, and through an intermediate representation which summarises common elements, creates a representation of the same model in EER, ORM or UML trying to preserve the semantics of the original.

SPARQL-DL is a query language (QL) for OWL-DL [10] ontologies significantly more expressive than existing DL QL by allowing combined ABox, RBox and TBox queries. It is aligned with SPARQL [11] to improve interoperability of applications on the Semantic Web and to allow to be covered by the standard reasoning services OWL-DL reasoners provide [12].

Aiming at integrating SPARQL-DL in an ontology engineering environment, this paper introduces a graphical language that preserves similarities to UML for creating visual SPARQL-DL queries. Furthermore, it presents a modification of the *crowd* architecture that allows visual SPARQL-DL queries. Implementing a query language in an already developed tool for creating ontologies via visual languages is not only the next logical step, but also a relevant one as its approach to ontology engineers mean to understand and study their ontologies and also help to create queries that their users require for satisfying their information needs. Moreover, it is possible to provide more reasoning services like anti-pattern based searches, in order to look for modelling issues in the ontology by obtaining a subset of the input concepts and rules and displaying them visually. However, a tool that supports both scenarios, design and query in a graphical way, and an implementation of a SPARQL-DL-based visual language is not available to our best of our knowledge.

The rest of the work is structured as follows: Section 2 introduces some well-known textual and visual query languages that has been researched through this work. Section 3 describes the UML-like language we propose for representing SPARQL-DL queries, the mapping rules and an example of query. Later, section 4 describes *crowd* architecture and the implementation needed for providing support to the proposed visual language and the reasoner for executing the queries. Section 5 shows possible applications of for this language. Finally, in section 6 summarises the paper and discuss future works.

2 Textual and Visual Query Languages

Textual query languages for ontologies provides a mean to retrieve information and structure for the user. The most common of these languages is OWLlink [13], the successor of DIG [14]. OWLlink provides a neutral mechanism for accessing OWL and OWL 2 reasoner functionality through a declarative interface whose protocol can be bound to a concrete transport mechanism like HTTP/1.1 and XML. OWLlink provides a wide spectrum of queries we can make, but they can

only affect the ABox, the TBox or the RBox separately. On the other hand, nRQL[15] is another DL query interface focused on retrieving ABox individuals that satisfy specific conditions which extends the expressivity of Racer [16] query language. However we cannot use it to retrieve structural definition from the ontology. Finally, SPARQL-DL is a powerful and expressive query language for OWL-DL that can combine ABox/TBox/RBox queries. It aligns with SPARQL and is simple enough to be easily built on top of existing OWL-DL reasoners [12].

In regard to visual query languages (VQL), we can find a variety of languages focused on querying OWL-DL ontologies. They present different ways of querying an ontology, but they does not provide a SPARQL-DL representation nor match its expressiveness. To our knowledge, there are no visual language created in base of SPARQL-DL publicly available.

In OntoVQL [4], queries are modelled by a graph connected style where concepts and individuals are nodes and roles are drawn by edges. Operators as "and" and "or" can be used for connecting more than one sub-query. Similarly to others visual languages, like SEWASIE, GRQL and OZONE, all these provides expressiveness enough for querying about the ABox only through a database or a nRQL query mapping.

GrOWL-Query [17] allows us to query about an ontology by representing the underlying DL semantics of OWL ontologies and replacing individuals by variables. Knowledge about OWL and DL is needed for using this language augmenting the learning curve.

A visual query system like OptiqueVQS also provides a visual language along with different techniques that facilitates the query formulation. The ability to query the ontology through SPARQL only allows OptiqueVQS to query about the ABox.

3 A UML-like Graphical Language for SPARQL-DL

A UML-like language is selected for preserving the objective of being accessible to the user and easy to understand. This section describes a UML-like language and a mapping from some of its graphical primitives to a SPARQL-DL scheme, which allows us to compute the query.

3.1 A UML-Like Visual Query Language

We define a UML-like visual language in which classes, associations, generalisations are primitives that represents concepts, roles and inclusion in DL, respectively. Objects in UML are depicted similarly as classes, where the name is combined with the instance name, a colon (":") and the Classifier name, all of them underlined [18]. This primitive is for representing ABox instances in DL, providing a way for querying about both TBox and ABox.

Associations can have names which we can use for representing roles names but multiplicity is not considered in this preliminary version. Class or instances

classifier names are used as concepts names in the DL representation. All names can be replaced with variables for requesting a set of answers that matches any element in the TBox and/or ABox that meets the UML diagram representation. Variables are written prefixed with a question mark. Depicting an UML primitive named without a variable is considering as asking for existence of that element in the input ontology. For example, an UML class named "Project" will represent the concept Project in the ontology and the output will be asserted as true if founded in the input TBox or false otherwise.

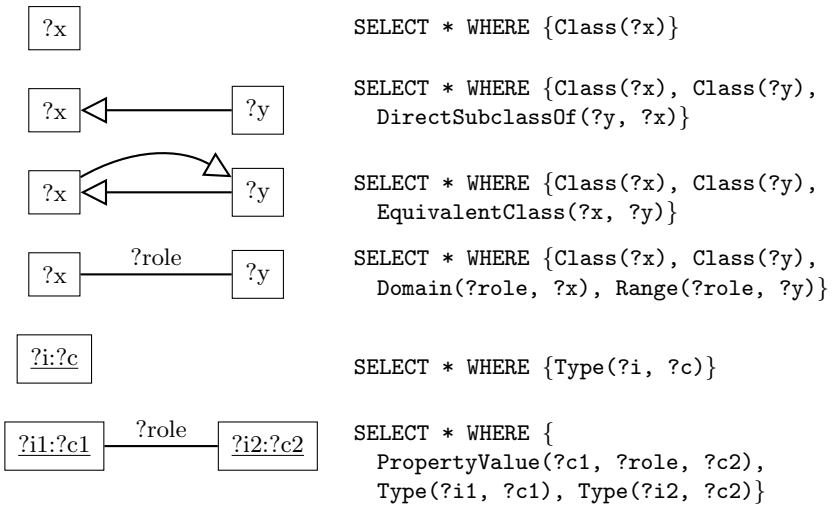


Fig. 1. Summary of the UML primitives mapping to SPARQL-DL.

3.2 SPARQL-DL Encoding

SPARQL-DL is used as a textual language for representing queries in *crowd* VQL. This language posses two types of queries: ASK sentence for querying about existence or assertions, which returns a true/false answer and SELECT queries for retrieving a set of elements that matches the provided condition.

On Figure 1, a summary of UML primitives with variables are depicted with their respective SPARQL-DL encoding. Using variables in the graphical query is considered as a query for retrieving a set of elements that matches the user's condition, and thus, we use a SELECT sentence. The conditions are determined by the query atoms, which are constructed depending on the primitives used in the diagram. Classes are interpreted as `Class()` in SPARQL-DL, allowing the engine to search for concepts in the input. A generalisation requires the representation in SPARQL-DL of the two classes or objects involved and a `DirectSubclassOf()` sentence for searching a direct subclass relation between these two elements. For

equivalence, the sentence `EquivalentClass()` is used. Associations are mapped into a `Domain()` and `Range()` query atoms for requesting the classes involved in the role.

Instances in the ABox can be queried by the objects UML representation, where we can use the name of the instance and class in a `Type(?i, ?c)` SPARQL-DL sentence which requests the instance "?i" of a concept "?c". Roles also can be mapped by a combination of three query atoms: a `PropertyValue()` and two `Type()`, the first one for searching the classes involved in the role, and the rest for retrieving the instance that can participate on it.

ASK queries are used when the graphical query provides names without variables. In this case, they are represented similarly as the SELECT queries but providing the name string.

The SPARQL-DL encoding order which we currently support emphasise the primitives related to querying the TBox. First, classes are mapped, then generalisations and finally associations. Next, primitives related to the ABox are considered: objects, generalisations between objects and roles (associations between objects).

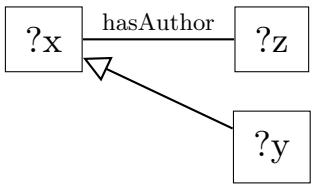


Fig. 2. A query expressed in UML class diagram model.

3.3 Example UML to SPARQL-DL

In the example presented on Figure 2 the user creates a model which requests all classes involved in the model, where a class "X" is in relation named `hasAuthor` with another class "Z" and is a parent of class "Y". The conversion of this model to a SPARQL-DL sentence will start by creating a query atom for each class and then for each relation. Then, it will detect the use of variables and create a SELECT statement with all of them. The sentence generated will be as follows:

```
SELECT ?x,?z,?y WHERE
{
  Class(?x), Class(?z), Class(?y), DirectSubclassOf(?y,?x),
  Domain(?x,:hasAuthor), Range(?z,:hasAuthor)
}
```

Afterwards, the SPARQL-DL reasoner will be used for processing the query with the user ontology. All the results will be displayed on the user interface.

4 *crowd* for VQL

The *crowd* architecture, depicted in Figure 3, consists in a client-server style where the client provides a Web interface following the view-model architecture provided by the BackboneJS and JointJS libraries. This allows the Javascript interface to generate a JSON representation of the user's model for sending to the server. The server side provides reasoning services while creating a DL encod-

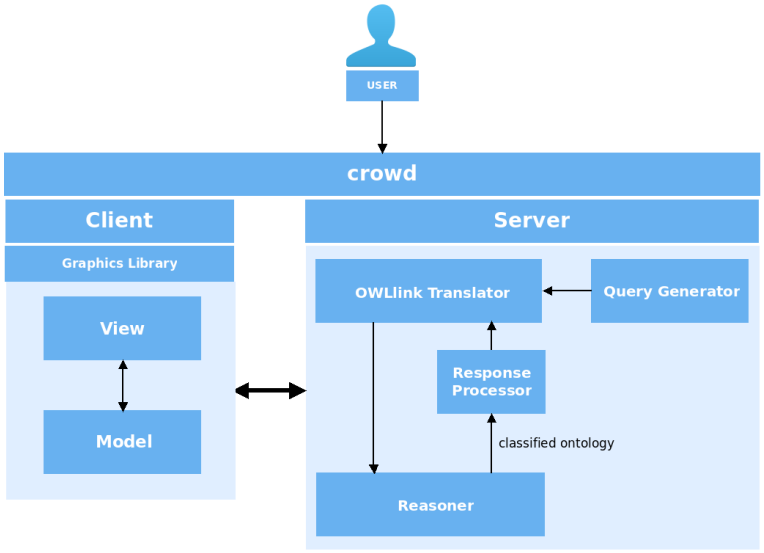


Fig. 3. *crowd* architecture.

ing from the JSON input. The module responsible for this is called "OWLlink Translator", which first creates an ontology in OWL 2 according to the method described by [19, 20]. For providing consistency checking services, a "Query Generator" module defines the necessary queries for the reasoner depending on the user's model and appends them to the OWLlink [13] document with the OWL 2 ontology already generated. The "Reasoner" module executes and maintains the program which use the OWLlink document as input, elaborating the answer which will be interpreted by the "Response Processor" creating a JSON understandable by the interface.

In Figure 4, the overall visual query processing is depicted. The visual query model designed by the user and the ontology are imported from the front-end interface. Both elements are sent to the back-end, the first in JSON format for creating the SPARQL-DL encoding in the SPARQL-DL Processor. Finally, the Reasoner receives both inputs, the ontology in OWL 2 syntax and the query, processes them and generates the output results in JSON or XML which can be used for sending back to the interface for displaying to the user.

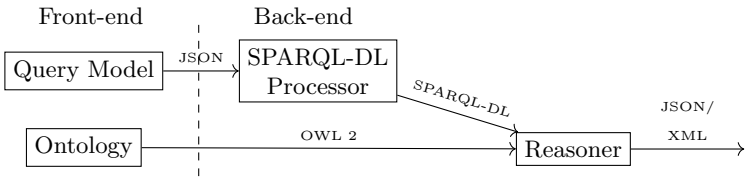


Fig. 4. The back-end steps for processing the query modelled by the user.

For implementing this process, some features has to be added to the initial version of *crowd*. The mapping rules described before were implemented for creating the SPARQL-DL encoding. The SPARQL-DL reasoner is called similarly as the current one with its own parameters and input files. Finally, the output is captured and processed by an extended version of the Response Processor for sending to the interface for the user consumption.

Currently, the graphical expressiveness of *crowd* VQL is constrained to UML classes, instances, generalisations and binary associations. Remains the rest of UML class diagram representation and SPARQL-DL encoding for future development.

5 Applications

Visual techniques for accessing data facilitate the design of queries for users and developers, reducing the learning curve and also abstract the complexity of the underlying logical query language. Moreover, it is possible to use the UML-like syntax for searching ontology patterns [21] and anti-patterns [22, 23] represented in a visual context and thus helping the engineer to raise the quality of the ontology design. In the same direction, other applications are related to the bottom-up construction of ontologies by means of non-standard services such as last common subsumers and most specific concepts [24]. Lastly, reasoners can also search for a possible pattern in a top-down way by querying the existence of part of the pattern into the input ontology, and thus suggesting the missing elements which completes them [25, 26]. For example, consider the pattern "List 2" presented in Figure 5. Part of the pattern is queried by using the UML class diagram depicted into the input ontology. The missing elements of the pattern can be suggested to the user for appending to the input ontology according to the results of the query.

Searching for anti-patterns represented in a graphical context can avoid situations that normally result in inconsistencies. This anti-patterns can be imported or drawn as a query in a UML-like diagram, making it easier to the user to remember and to understand. For example, searching for the logical anti-pattern called "Synonym or Equivalence" (SOE) explained in [28] requires to look into the ontology for the expression $C1 \equiv C2$, which latter should be treated by the user removing C2 according to the pattern solution. The Figure 6 displays a UML representation for the expression.

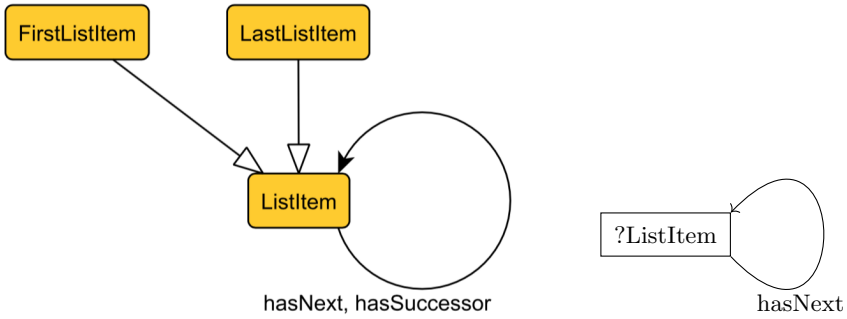


Fig. 5. List pattern obtained from [27] and the UML-like query used for searching the pattern.

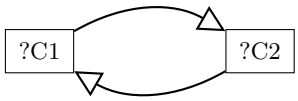


Fig. 6. UML-like diagram for querying the expression $C1 \equiv C2$.

6 Conclusions and Future Works

In this work, we proposed a graphical visual query language using an UML-like notation in order to keep the consistency with the standard conceptual modelling languages and reduce the curve of learning drastically. As an intermediate language, we propose the use of SPARQL-DL a textual language, whose expressiveness is greater than the current implemented, like OWLlink, nRQL and DIG, by mixing the ABox, RBox and TBox in the same query. Moreover, we modified our Web tool named *crowd* for supporting this graphical language and the import of ontologies in OWL-DL, creating an environment suitable for designing the query and testing it with an ontology. The back-end has been adapted for supporting the encoding of the graphical models with the SPARQL-DL queries and for executing it with the provided OWL 2 ontology. A running prototype is available following URL `http://crowd.fi.uncoma.edu.ar/sparqldl/`.

In future works, we expect to extend *crowd* with an interface for displaying the results of the reasoning system in a tabular and graphical view instead of the textual output. The rest of the UML class diagrams primitives not presented in this work are considered for next releases along with their SPARQL-DL encoding. Also, an expansion of this conceptual modelling language for supporting more SPARQL-DL query atoms are under study for future implementations. Finally,

the inclusion of Object Constraint Language (OCL) [29] for augmenting the expressiveness of the query is considered for future study.

References

1. Grigoris Antoniou and Frank van Harmelen. *A Semantic Web Primer, 2Nd Edition (Cooperative Information Systems)*. The MIT Press, 2 edition, 2008.
2. Dean Allemang and James A. Hendler. *Semantic Web for the working ontologist effective modeling in RDFS and OWL*. Morgan Kaufmann/Elsevier, Waltham, MA, 2011.
3. Ahmet Soylu, Evgeny Kharlamov, Dimitry Zheleznyakov, Ernesto Jimenez Ruiz, Martin Giese, Martin G. Skjaeveland, Dag Hovland, Rudolf Schlatte, Sebastian Brandt, Hallstein Lie, and Ian Horrocks. OptiqueVQS: a visual query system over ontologies for industry. *Semantic Web*, (in press), 2017.
4. Amineh Fadhil. Ontovql: Ontology visual query language. Master's thesis, Concordia University Montreal, Quebec, Canada, September 2008.
5. Christian Gimenez, Germán Braun, Laura Cecchi, and Pablo Fillottrani. crowd: A Tool for Conceptual Modelling assisted by Automated Reasoning - Preliminary Report. In *the 2nd Simposio Argentino de Ontologías y sus Aplicaciones SAOA '16 JAIIO '16*, 2016.
6. Federico Solorza, Germán Braun, Laura Cecchi, and Pablo Fillottrani. Hacia la Formalización de un Lenguaje Visual Unificador de UML, EER y ORM2. In *Workshop de Investigadores en Ciencias de la Computación 2018*, 2018.
7. Ángela Oyarzun, Germán Braun, Laura Cecchi, and Pablo Fillottrani. Una Herramienta Gráfica con Razonamiento basado en DL para el Análisis de Modelos OVM. In *Workshop de Investigadores en Ciencias de la Computación 2018*, 2018.
8. C. Maria Keet and Pablo Rubén Fillottrani. An ontology-driven unifying metamodel of UML Class Diagrams, EER, and ORM2. *Data & Knowledge Engineering*, 2015.
9. Pablo R. Fillottrani and C. Maria Keet. KF metamodel formalization. *CoRR*, abs/1412.6545, 2014.
10. Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, and Sebastian Rudolph, editors. *OWL 2 Web Ontology Language: Primer*. W3C Recommendation, December 2012. Available at `http://www.w3.org/TR/owl2-primer/` last visited July 2018 (Internet archive `http://archive.today/H7G0H`).
11. W3C OWL Working Group. *SPARQL 1.1 Query Language*. W3C Recommendation, 21 March 2013. Available at `http://www.w3.org/TR/sparql11-query/`.
12. Ilianna Kollia, Birte Glimm, and Ian Horrocks. SPARQL query answering over OWL ontologies. In Grigoris Antoniou, Marko Grobelnik, Elena Simperl, Bijan Parsia, Dimitris Plexousakis, Pieter De Leenheer, and Jeff Pan, editors, *The Semantic Web: Research and Applications*, pages 382–396, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
13. Thorsten Liebig, Marko Luther, Olaf Noppens, and Michael Wessel. Owllink. *Semantic Web*, 2(1):23–32, 2011.
14. S. Bechhofer, R. Moller, and P. Crowther. The DIG Description Logic Interface. In *In Proc. of International Workshop on Description Logics (DL2003)*, 2003.
15. Volker Haarslev, Ralf Möller, and Michael Wessel. Querying the semantic web with Racer + nRQL. In *In Proceedings of the KI-2004 International Workshop on Applications of Description Logics (ADL'04*, 2004.

16. V. Haarslev and R. Möller. Racer system description. In R. Goré, A. Leitsch, and T. Nipkow, editors, *International Joint Conference on Automated Reasoning, IJCAR'2001, June 18-23, Siena, Italy*, pages 701–705. Springer-Verlag, 2001.
17. Sergey Krivov, Richard Williams, and Ferdinando Villa. Growl: A tool for visualization and editing of OWL ontologies. *J. Web Sem.*, 5(2):54–57, 2007.
18. Object Management Group (OMG). OMG unified modelling language version 2.5, 2015.
19. Daniela Berardi, Diego Calvanese, and Giuseppe De Giacomo. Reasoning on UML class diagrams. *Artif. Intell.*, 168(1-2):70–118, 2005.
20. Germán Alejandro Braun, Laura Andrea Cecchi, and Pablo R. Fillottrani. Integrating graphical support with reasoning in a methodology for ontology evolution. In *Proceedings of the Joint Ontology Workshops 2015 Episode 1: The Argentine Winter of Ontology co-located with the 24th International Joint Conference on Artificial Intelligence (IJCAI 2015), Buenos Aires, Argentina, July 25-27, 2015.*, 2015.
21. Steffen Staab and Rudi Studer. *Handbook on Ontologies*. Springer Publishing Company, Incorporated, 2nd edition, 2009.
22. Tiago Prince Sales and Giancarlo Guizzardi. Ontological anti-patterns. *Data Knowl. Eng.*, 99(C):72–104, September 2015.
23. Catherine Roussey, Oscar Corcho, and Luis Manuel Vilches-Blázquez. A catalogue of owl ontology antipatterns. In *Proceedings of the Fifth International Conference on Knowledge Capture*, K-CAP '09, pages 205–206, New York, NY, USA, 2009. ACM.
24. Franz Baader. Least common subsumers and most specific concepts in a description logic with existential restrictions and terminological cycles. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, IJCAI'03, 2003.
25. Pablo Rubén Fillottrani and C. Maria Keet. Patterns for heterogeneous TBox mappings to bridge different modelling decisions. In Eva Blomqvist, Diana Maynard, Aldo Gangemi, Rinke Hoekstra, Pascal Hitzler, and Olaf Hartig, editors, *The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I*, volume 10249 of *Lecture Notes in Computer Science*, pages 371–386, 2017.
26. Germán Alejandro Braun and Laura Andrea Cecchi. Extension rules for ontology evolution within a conceptual modelling tool. In *Proceedings of the 1st Argentine Symposium on Ontologies and their Applications co-located with 44 Jornadas Argentinas de Informática (44JAIIO), Rosario, Argentina, September 2-3, 2015.*, 2015.
27. David Carral, Pascal Hitzler, Hilmar Lapp, and Sebastian Rudolph. List 2. `http://ontologydesignpatterns.org/wiki/Submissions:List_2` visited on July 2018 (Internet archive `http://archive.today/vOHit`).
28. Catherine Roussey and Oscar Corcho. Synonym or equivalence. `http://ontologydesignpatterns.org/wiki/Submissions:SynonymOrEquivalence_(SOE)` visited on July 2018 (Internet archive `http://archive.today/OO8AI`).
29. Object Management Group. *Object Constraint Language (Version 2.4)*. Pearson Higher Education, 2014.

Workshop de Procesamiento de Señales y Sistemas de Tiempo Real (WPSTR)

Coordinadores

Osvaldo Sositto (UNLaM) - Horacio Villagarcia Wanza (UNLP) - Emanuel Frati (UNdeC) - Jorge Ierache (UM)

Temas

Señales y Arquitecturas orientadas al tratamiento de señales
Sensores. Redes de Sensores
Sistemas embebidos. Aplicaciones IoT
Robótica Autónoma. Robótica situada. Drones
Sistemas Industriales. Industrias 4.0. Manufactura Integrada por Computadora. Visión por computadora
Modelos de Sistemas de Tiempo Real
Ingeniería de Software para STR
Verificación y Validación de STR
Redes, protocolos y Seguridad en STR
Sistemas de Tiempo Real duro

Detección de carriles en caminos rurales no pavimentados

Marcelo Tosini, Lucas Leiva

Departamento de Computación y Sistemas
Facultad de Ciencias Exactas
Universidad Nacional del Centro de la Provincia de Buenos Aires
Tandil – Buenos Aires – Argentina
{mtosini,lleiva}@exa.unicen.edu.ar

Resumen. El trabajo presenta un procedimiento para la detección de caminos rurales, particularmente aquellos de tierra, grava, ripio o estabilizado, que por sus características constructivas no presentan delimitaciones laterales parejas, ni demarcación central o lateral. Este tipo de caminos es frecuente en varios países que, por su infraestructura vial, aún tienen un gran porcentaje de carreteras consideradas como pertenecientes a la red terciaria –sin pavimentar ni delinear-. El procedimiento propuesto de detección de bordes de los caminos utiliza una estrategia de análisis primario de los canales de matiz, saturación e intensidad de la imagen a fin de detectar cuál de ellos provee, en cada caso, mejor información. Posteriormente se aplica una secuencia de filtros de convolución que faciliten la detección de las curvas que mapean ambos bordes laterales mediante una transformada de Hough adaptativa.

Palabras Clave. Detección de carriles, Asistencia al guiado, caminos rurales

1. Introducción

Si bien la existencia de caminos no demarcados o en condiciones precarias afecta a varios países, se ha concentrado en este trabajo el análisis de la estructura vial en la Argentina. En este sentido, el país tiene aproximadamente 500.000 km de carreteras y caminos, de los cuales 37.800 conforman la red primaria o nacional y 178.000 la red secundaria o provincial, restando 285.000 que forman la red terciaria dependiente de las municipalidades [1]. La tabla 1 muestra la distribución por zonas de esta red vial con amplia preeminencia de construcción en la provincia de Buenos Aires y, en particular, el conurbano bonaerense.

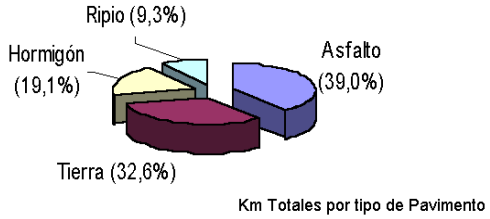
Solo 61.000 km de los 215.000 que forman las redes primarias y secundarias están pavimentados mientras que 37.000 tiene algún tipo de "mejora" (grava o estabilizado). Los restantes 117.000 km están compuestos por caminos de tierra, al igual que la gran mayoría de los que forman la red terciaria, lo que significa que Argentina tiene más de 400.000 km de caminos de tierra, (80% de su red vial).

Un dato importante es que, de los 28.000 km de red primaria pavimentada, 20.000 poseen carpeta asfáltica y casi 1.000 son de hormigón, existiendo alrededor de 8.000 km con tratamiento bituminoso. Considerando que Argentina posee una superficie continental de 2,8 millones de km^2, se puede concluir que el país tiene una estructura

vial subdesarrollada, lo cual se debe en parte a su escasa población, problema que se ve agravado por el hecho de la alta concentración demográfica en ciertas zonas.

Tabla 1: Red vial Argentina por zonas.

	Km. Lineales	Porcentaje
Buenos Aires	18.900	50%
Centro	7.560	20%
Norte	5.292	14%
Mesopotamia	2.268	6%
Cuyo	2.268	6%
Sur	1.512	4%
TOTALES	37800	



Distribución	Kms
Asfalto	170.395
Tierra	142.610
Hormigón	83.270
Ripio	40.550
TOTALES	436.825

Fig 1: Porcentajes de distribución de caminos en Argentina según su tipo de construcción.

Ya que cerca del 80% de la red vial es de tierra, grava u otro tipo de estabilización es necesario considerar el estudio de mecanismos de detección de estos caminos.

Estadísticamente, un 49% de los accidentes de tránsito ocurren en zonas rurales [2] siendo 12% el total de accidentes ocurridos en tipos de calzadas con condiciones anormales (ripio, tierra, pavimento húmedo/mojado, otros) [3]. Los accidentes de tránsito ocupan el lugar número 13 de las causas de mortalidad en Argentina [4] (2,17% en 2014) involucrando un total de 7268 víctimas fatales en 2016/17 [2].

Tecnológicamente hablando, hoy en día existen sistemas inteligentes aplicados en productos de la industria automotriz que tratan de evitar situaciones de riesgo a los conductores. Entre ellos se hallan sistemas de reconocimiento y detección de carriles. Este término refiere al proceso automático de localizar límites de una calzada sin conocimiento sobre la infraestructura vial. Estas técnicas permiten detectar curvas cerradas, obstáculos y la posición del vehículo, para mitigar posibles riesgos. Frecuentemente los conductores cambian intencionalmente de carril, o no son conscientes de esta acción por distracción o cansancio. Así, es importante conocer la posición del vehículo, y alertar al conductor de posibles situaciones de peligro.

En caminos correctamente estructurados y demarcados existen diversas técnicas para la detección de líneas basadas en los siguientes supuestos: 1) la textura del camino es consistente, 2) el carril está debidamente definido, 3) el ancho del carril es constante, 4) las señales siguen reglas estrictas para la aparición o colocación y 5) el camino es una superficie plana o sigue un estricto modelo de elevación.

Muchos algoritmos existentes tienden a utilizar al menos uno o más de los supuestos anteriores, obteniendo sensibles mejoras en los resultados globales y

disminuyendo el tiempo de proceso. Sin embargo, estas suposiciones pueden fallar o no existir. La hipótesis de una textura de carretera constante puede mejorar en gran medida los resultados de toda su superficie. Se puede utilizar como algo más que las marcas de carretera característica. En situaciones donde las marcas de carretera son escasas o ausentes, pueden proporcionar aún una estimación de la posición del carril.

La suposición de que el ancho del camino o de los carriles es constante puede mejorar el funcionamiento de los algoritmos de detección permitiendo la fusión de los límites laterales de la mano izquierda y de la mano derecha.

Las marcas de carreteras suelen ser claras y continuas en una superficie oscura. Frecuentemente, se supone que la superficie de la carretera es plana o sigue un modelo de elevación constante.

Existen, por otro lado, caminos no estructurados donde no hay carriles o límites obvios, y son vulnerables a sombras y cambios de iluminación. En estos casos, la textura del camino dependerá del tipo de material usado para su rellenado o, en muchos casos, el material natural existente: tierra, arcilla, grava, piedra, etc. Lo mismo ocurre con los límites del camino, determinados generalmente por los accidentes naturales de la geografía, como rocas, arboledas, límites de propiedades y otros. La señalización de cualquier tipo es inexistente o está limitada a cartelería lateral con indicaciones de algún accidente geográfico importante. Por último, este tipo de caminos no sigue ningún modelo estricto de elevación. Según [5] los métodos de detección de carreteras no estructuradas se pueden dividir en tres grandes grupos: basado en el modelo de carretera, basado en características de la carretera y basado en conocimiento previo adquirido mediante redes neuronales. El primer método asume una carretera con límites regulares y analiza ciertos puntos característicos para determinar la pertenencia a un modelo predefinido. El segundo, analiza las variaciones en color y textura para determinar los límites del camino. El último grupo se basa en la detección de los rasgos particulares de las rutas a partir de un gran número de muestras previas usadas en el entrenamiento de un sistema neuronal.

Los caminos rurales pertenecen al grupo de caminos no estructurados, fuertemente influenciados por el medio rural, con fuertes cambios por iluminación (debida fundamentalmente a la presencia de árboles), poca definición de límites laterales (usualmente el material del camino y la banquina son el mismo) y características morfológicas escasas o ausentes (ancho discontinuo y dependiente de las características del terreno), por lo que la detección es más difícil.

Algunos trabajos presentados tratan el tema de la detección de límites en este tipo de caminos. Lamentablemente no son tantos como aquellos que abordan la detección de carriles y otros objetos (peatones, carteles) en carreteras estructuradas.

En [5] se presenta un método de detección de carriles en un ambiente o entorno complicado. El trabajo propone el análisis de las imágenes obtenidas convertidas a espacio de colores HIS y análisis del canal H a fin de detectar los límites del camino.

Por otra parte, [6] propone un método algo diferente pero de similar inicio, que involucra un preproceso de la imagen con una conversión a escala de grises basado en una ecuación de ponderación de los colores rojo, verde y azul de la imagen original.

Kong *et. al* [7] hace un análisis de las imágenes de caminos rurales de una manera más estructurada analizando la textura de los pixeles mediante un filtro Gabor. El análisis se realiza sobre una transformación de los colores R, G, B de la imagen.

Otros métodos de la literatura analizan la imagen tratando de estimar tres áreas importantes, el área de horizonte (parte superior), el área de camino y el área de los laterales [8]. Dichas áreas se analizan en función de dos características: Su disposición en la imagen y su composición a partir de sus componentes de color.

Crisman [9] propone un método basado de reconocimiento de patrones, para navegar con éxito en una variedad de carreteras no estructuradas. El algoritmo utiliza una técnica de agrupación de píxeles con colores y ubicaciones similares, y se contrasta la imagen con modelos de carretera para localizar los límites del camino.

Otros métodos usan técnicas de visión estéreo, LIDAR [10,11] o la combinación de información de diferentes sensores [12] para reconocer de caminos rurales.

En este trabajo se propone un algoritmo alternativo que obtiene los límites laterales de camino rurales a partir de la captura de información de una cámara con la mayor certidumbre. En la sección 2 se presenta el desarrollo de la técnica propuesta, en la sección 3 los resultados obtenidos de la evaluación del algoritmo utilizando imágenes reales, y finalmente la sección 4 detalla las conclusiones del trabajo realizado.

2. Desarrollo

El método propuesto se basa en una serie de pasos para obtener los límites laterales (Figura 2). 1) transformación de espacio RGB a HSV. 2) histograma de cada canal (matiz, saturación, valor) para detectar cual contiene mejor información acerca del camino. 3) se suavizan los detalles con una máscara de promedio entre vecinos para disminuir imperfecciones del borde del camino, además de un filtro de umbral para eliminar las partes de más luminosidad de la imagen y un Laplaciano para destacar las regiones del camino con cambios bruscos de intensidad a fin de resaltar los bordes del mismo. La detección de bordes se realiza con un filtro de Roberts y, 4), una optimización basada en Hough permite detectar las líneas de los bordes.

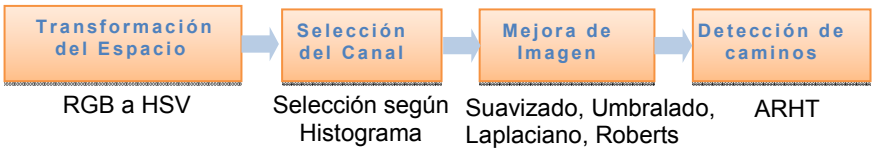


Fig 2: Método propuesto (Diagrama de bloques).

Paso 1 Transformación del espacio

Dado que en el espacio de color RGB las tres componentes están interrelacionadas y no aportan información útil en ninguno de los tres canales se aplicó a las imágenes una transformación al espacio HSV según las ecuaciones de conversión estandarizadas. Los resultados para tres caminos usados como ejemplo se muestran en la figura 3 a continuación.

Los resultados obtenidos fueron que el canal de matiz *H* presentó mejor definición del camino propiamente dicho respecto del entorno, por lo que se experimentó variantes de la ecuación de conversión inicial a fin de mejorar los resultados

El procesamiento de los bordes se realiza mediante la aplicación de un par de filtros de convolución: Laplaciano y Roberts. El primero ayuda a delinear los bordes de la imagen (incluido el camino) para que el filtro de Roberts pueda detectar los mismos con mayor continuidad, como se observa en la figura 7 b) y c).

Los mejores resultados en el cálculo del matiz se obtuvieron al modificar las ecuaciones de obtención del parámetro. De este modo, según el color básico predominante en cada pixel (*R*, *G*, *B*) se desfasa el matiz en 120, 240 o 360 grados.

Paso 2 Selección de canal por histograma

En distintos escenarios de caminos rurales las condiciones de tipo de terreno, material de construcción del camino, vegetación circundante y otros factores determinan la información específica acerca del camino que se obtendrá en los tres canales HSV. Se ha visto durante la realización de este trabajo que en ocasiones es el canal de matiz el que aporta mejor información sobre los límites reales del camino y que permitirá mejor detección de bordes del mismo. En otros escenarios la mejor información la aporta el canal de saturación. Más raramente es el canal de intensidad el que aporta información relevante.

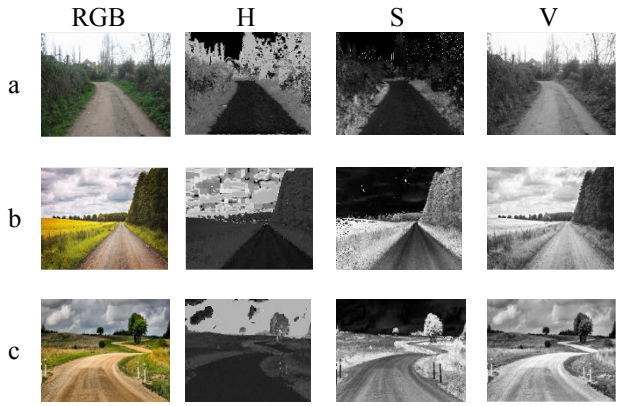


Fig 3: Resultante de conversión a HSV de tres caminos de tierra. En los tres el canal *H* muestra mejores resultados resaltando el camino de su entorno.

Por otro lado, en cualquier captura de escenario mediante una cámara en el frente del vehículo, la imagen del camino muestra al mismo aproximadamente en la zona inferior de la captura, quedando la zona superior relegada a reflejar información del horizonte o de la topografía lejana al vehículo como se observa en el esquema de la figura 4.

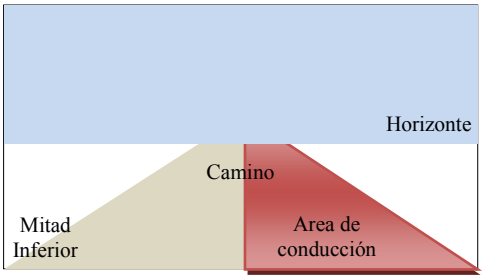


Fig 4: Regiones de interés de una imagen para detección de caminos.

En estas condiciones el histograma obtenido de la mitad inferior de la imagen para los tres canales permitirá determinar cuál de ellos aportará mayor información a los pasos posteriores de la determinación de límites.

La selección se realiza analizando la concentración de píxeles en la zona cercana al negro (aproximadamente los primeros 40 valores del histograma) y la frecuencia de los mismos (figura 5). Un histograma con picos en la zona de interés y baja frecuencia en otras zonas favorecerá mejores detecciones de bordes.

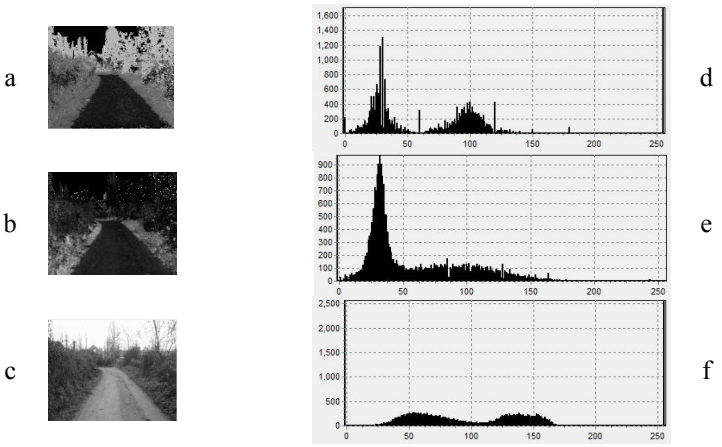


Fig 5: Histogramas de los tres canales H, S y V para el camino de la figura 2-a. d) histograma del canal H (a). e) histograma del canal S (b) y, f) histograma del canal V (c).

Para el ejemplo de la figura 5 se puede observar que el histograma del canal de intensidad (figura 5 f)) no aporta información útil para la determinación de bordes. Los canales de matiz y saturación, por el contrario, si aportarán información de borde. El canal H tiene gran concentración de píxeles en la zona del negro, con picos de hasta 1400 pixeles en valores de entre 30 y 40. El canal de saturación también presenta picos en la zona izquierda pero su concentración es inferior, por un lado, y la dispersión general del histograma es mayor que para el canal H. Por lo tanto, el canal H aportará mejor información para la determinación de bordes del camino.

Paso 3 Mejora de imagen: Promediado, umbralado y detección de bordes

A fin de suavizar la imagen se aplica un filtro de convolución con matriz unidad a los vecinos de cada píxel permitiendo la disminución de detalles y ruidos sobre todo en los límites del camino. Un umbralado adicional elimina la información más iluminada de la imagen dejando solamente el perfil del camino y algunas áreas adicionales (no deseadas) correspondientes a zonas de matiz similar al del camino (figura 6).

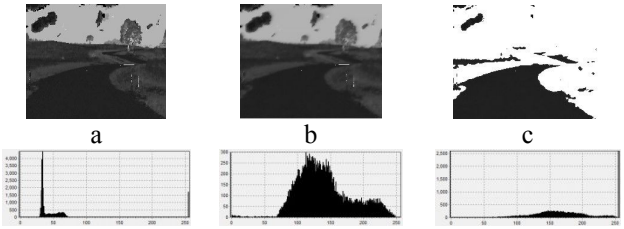


Fig 6: Suavizado (b) y umbralado (c) de la imagen de matiz H (a) para el camino de la figura 3 c).
Se muestran también los histogramas correspondientes a los tres canales donde se observa que el canal H (imagen a)) aporta mejor información de detección.



Fig 7: a) Filtro Laplaciano. b)Filtro de Roberts sin Laplaciano previo y, c) filtro de Roberts con Laplaciano previo.

Se analizaron distintas máscaras Laplacianas obteniéndose mejores resultados con una matriz de contraste general para todos los vecinos horizontales, verticales y diagonales al punto seleccionado del tipo

$$\begin{matrix} -1 & -1 & -1 \\ -1 & -8 & -1 \\ -1 & -1 & -1 \end{matrix} \quad (1)$$

A la imagen resultante (figura 7 a)) se le aplica un filtro de Roberts según la ecuación 2 dando como resultado una imagen como la de la figura 7 c).

$$\begin{bmatrix} p11 & p12 & p13 \\ p21 & p22 & p23 \\ p31 & p32 & p33 \end{bmatrix} \quad (2)$$

$$\eta p22 = \sqrt{(p22 - p33)^2 + (p23 - p32)^2}$$

Paso 4 Detección de bordes del camino mediante Hough adaptativo

Para la detección de las líneas de los bordes del camino se usó una variante del algoritmo de detección de Hough llamado ARHT (*Adaptive Random Hough Transform*) [13][14]. Esta versión mejorada permite no sólo la detección de rectas (como lo hace Hough) a partir de la localización de puntos que cumplan con la ecuación $y=m.x+n$, sino que es capaz de adaptar la trayectoria de búsqueda de nuevos puntos candidatos para el seguimiento de curvas que cumplan con:

$$|arctg(p0)-arctg(p1)|<thresold \quad (3)$$

El método evalúa la posible pertenencia de un nuevo punto *p1* a la luz de un punto de referencia *p0*. De este modo es posible calcular un conjunto de parámetros de curvatura y dirección de tangente tales que luego del muestreo de gran cantidad de píxeles se pueda elegir los parámetros más factibles.

Como resultado del algoritmo ARHT se obtienen muchos puntos candidatos y se puede encontrar el centro y radio más apropiados, que cumplan con el error límite (*threshold*) y que contenga a la mayoría de los puntos.

La aplicación del algoritmo a los casos de imágenes de caminos se muestra en la figura 8.

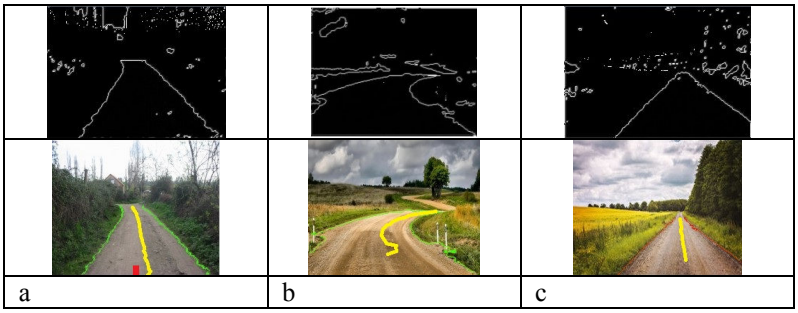


Fig 8: Resultado de la aplicación del filtro de Hough adaptativo para los tres caminos usados como ejemplos.

3. Resultados experimentales

El algoritmo para detección de caminos rurales no pavimentados ni demarcados fue testeado en dos grupos de imágenes. Un conjunto de imágenes de caminos rurales extraído de la Web, conformado por 156 imágenes, por un lado y 3000 imágenes (en grupos de 500) extraídas de seis videos capturados por una cámara fija montada en un vehículo -casos de estudio 1 a 6- recorriendo distintas travesías por senderos rurales. El algoritmo fue testeado sobre diferentes condiciones, como amanecer y atardecer, sombras, lluvia, entre otras. La resolución de las imágenes fue de 240x192 pixeles. El algoritmo fue ejecutado en un Intel i5 con 6Gb de RAM.

La estimación de acierto en la detección se realizó de forma semiautomática a partir de un algoritmo que promediaba las coordenadas horizontales de ambos bordes

(para algunas posiciones verticales dadas) para obtener una trayectoria central estimada (línea azul en la figura 8 a)). En cada análisis dicha trayectoria se compara con una referencia central (en rojo en la figura 8 a)) para la determinación de errores y aciertos. Se consideran aciertos los casos en que la trayectoria queda incluida en el *área de conducción* tal como se muestra en la figura 4.

La tabla 2 muestra los resultados obtenidos de la aplicación del algoritmo a las imágenes de los casos de estudio 1 a 6.

Resultados de entre 85 y 91% aparecen frecuentemente en la literatura para análisis de este tipo de caminos rurales [5][7][9], por lo que se concluye que el proceso algorítmico propuesto presenta buen comportamiento. Es de destacar el caso del video 3 con sólo 82% de aciertos. En un análisis más profundo el camino filmado en ese video fue un camino rural con variada arboleda en los laterales, lo que introdujo mucho ruido en el proceso de detección de bordes (figura 9)

4. Conclusiones

Este trabajo es parte de un proyecto mayor orientado al desarrollo de sistemas de asistencia a los conductores a fin de reducir accidentes de tránsito. Se propone un método de reconocimiento de caminos rurales no demarcados basado en filtrado de la imagen y aplicación de una transformación de Hough adaptativa. Primeramente la imagen se convierte de RGB a HSV y se analizan las tres componentes por separado para decidir cuál aporta mejor información acerca de los bordes del camino (en general la H o la S). Luego la imagen se filtra a partir de dos filtros Laplaciano y Roberts y en último lugar se realiza la detección propiamente dicha mediante la transformada de Hough adaptativa.

Tabla 2: Resultados experimentales.

Caso de estudio	Cantidad de imágenes	Número de aciertos	Aciertos (%)
1	500	441	88.2
2	500	459	91.8
3	500	410	82.0
4	500	432	86.4
5	500	452	90.4
6	500	444	88.8



Fig 9: Imagen tomada del video 3 con arboleda lateral al camino y mucho ruido resultante.

Los tiempos de cálculo permiten procesar del orden de 10 imágenes por segundo con lo cual en una conducción a 60 km/h se tiene información del camino cada 1.66 metros. El algoritmo mostró buenos resultados en caminos secos o mojados, con

variaciones de iluminación y tipo de material constructivo (arcilla, pedregullo, estabilizado). No se obtuvieron buenos resultados con condiciones de lluvia fuerte o en casos de caminos del mismo material que la banquina (casos en que la delimitación del camino no se da por color sino por construcción: cuneta pronunciada, altura diferente, etc). La técnica propuesta permite discriminar ambas líneas laterales, izquierda y derecha. Finalmente, el uso de datos históricos de imagines previas permite mejorar la exactitud de la detección.

Referencias

1. Red Vial Argentina, http://www.e-asfalto.com/redvialarg /redvial.htm
2. Asociación Civil: Luchemos por la Vida, Estadísticas, http://www.luchemos.org.ar/
3. CESVI Argentina, Seguridad Vial,: http://www.cesvi.com.ar/
4. World Life Expectancy, Health Rankings,: http://www.worldlifeexpectancy.com/
5. Zhang Wan-zhi1, 2 and Wang Zeng-cai.: Rural Road Detection of Color Image in Complicated Environment. International Journal of Signal Processing, Image Processing and Pattern Recognition. Vol.6, No.6, 161--168. http://dx.doi.org/10.14257/ijsip.2013.6.6.15 (2013)
6. Wenhong Zhu, Fuqiang Liu, Zhipeng Li, Xinhong Wang, Shanshan Zhang.: A Vision Based Lane Detection and Tracking Algorithm in Automatic Drive. IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application (2008)
7. Hui Kong, Jean-Yves Audibert, Jean Ponce, Willow Team.: General road detection from a single image. IEEE Transactions on Image Processing, Vol. 19, Issue 8 (2010)
8. Tobias Kuhnl, Jannik Fritsch.: Visio-spatial road boundary detection for unmarked urban and rural roads. IEEE Intelligent Vehicles Symposium Proceedings, 8-11 June 2014, Dearborn, MI, USA, 10.1109/IVS.2014.6856453 (2014)
9. Crisman, j., Thorpe, C.:UNSCARF-a color vision system for the detection of unstructured roads. IEEE Int. Conf. on Robotics and Automation, . (1991)
10. Manz, M., Himmelsbach, M., Luettel, T., & Wuensche, H. J.: Detection and tracking of road networks in rural terrain by fusing vision and LIDAR. IEEE International Conference on In Intelligent Robots and Systems (IROS), 4562--4568. (2011)
11. Bayerl, S. F., & Wuensche, H. J.: Detection and tracking of rural crossroads combining vision and LiDAR measurements. IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), 1274--1279. (2014).
12. S. F. X. Bayerl, T. Luettel, and H.-J. Wuensche.: Following Dirt Roads at Night Time: Sensors and Features for Lane Recognition and Tracking. Proceedings of 7th Workshop On Planning, Perception and Navigation for Intelligent Vehicles (PPNIV), IEEE/RSJ International Conference on Intelligent Robots and Systems, (2015)
13. Qing Li, Nanning Zheng, Hong Cheng.: Springrobot: A Prototype Autonomous Vehicle and Its Algorithms for Lane Detection. IEEE Transactions on Transportation System, Vol.5, No.4, 300--308. (2004)
14. Jing Lv, Yuanyuan Shang, Hui Ding.: Road Obstacle Detection Based on Randomized Hough Transform. 2nd International Conference on Teaching and Computational Science (ICTCS 2014), 72--74. (2014)

Software de segmento terreno de próxima generación

Pablo Soligo, Jorge Salvador Ierache

Universidad Nacional de La Matanza
Departamento de Ingeniería e Investigaciones Tecnológicas.
Florencio Varela 1903, B1754JEC San Justo, Buenos Aires.
psoligo@unlam.edu.ar
http://www.unlam.edu.ar/

Resumen Este trabajo presenta los resultados de las experiencias obtenidas en el desarrollo de la primera versión de un prototipo funcional de segmento terreno de próxima generación. El proyecto fue inicialmente desarrollado como parte de la materia "proyecto integrador" de la Maestría en Desarrollos Informáticos de Aplicación Espacial (MDIAE) y evolucionado por el grupo de investigación en desarrollo de software de aplicación aeroespacial de la Universidad Nacional de La Matanza (UNLaM) en el marco del proyecto PROINCE 211 "Sistemas de segmento terreno satelital de próxima generación". Las unidades de software desarrolladas permiten la ingestión y decodificación de telemetría de múltiples misiones con independencia del origen de datos y el comando y control de satélites. Se aplicaron técnicas y herramientas como camino alternativo a las tecnologías propietarias para desarrollar un entorno multimisión y multiplataforma. Como segmento de vuelo se ha utilizado el satélite de formación 2017 (FS2017) [1] y datos de misiones de terceros (Satnogs).

Keywords: Satelites Artificiales, Segmento Terreno, Diseño de Software, Tecnologia Espacial, Sistemas Distribuidos, Telemetría, Telecomandos

1. Introducción

El presente trabajo detalla las experiencias y resultados en el desarrollo de un prototipo de sistema de segmento terreno satelital costo-efectivo. El diseño e implementación del prototipo tiene como raíz las experiencias realizadas durante las maestrías con sistemas de segmentos terrenos actualmente en producción y su vinculación con los sistemas de software de propósito general de uso extendido en la industria. Se priorizo el desarrollo sobre herramientas de amplia aceptación y de probada robustez en contraste con soluciones comunes en la industria espacial, las cuales en algunos casos muestra un exceso de conservadurismo [2]. Parte del objetivo de este trabajo es confrontar dichas implementaciones, con predominancia en tres dominios:

- **Lenguajes de decodificación, comando y control:** Se propone el uso de lenguajes e intérpretes de propósito general para la decodificación de telemetría y el desarrollo de scripts de comandos.
- **Persistencia** [1] **de datos:** Se plantea una arquitectura donde la persistencia de datos se realiza sobre motores de base de datos relacionales y su acceso es siempre mediante un ORM(Object Relational Mapper).
- **Interfaces:** Las interfaces con sistemas internos y externos, tanto para la comunicación de módulos como para publicación e ingestión de datos se realiza mediante protocolos extendidos y aceptados.

Se marginan de la solución, estándares, formatos o protocolos *ad-hoc* o de escasa penetración en la industria de software de uso masivo y se prioriza las soluciones, bibliotecas, componentes y frameworks COTS(Commercial Off-The-Shelf). Los lineamientos de diseño adhieren a los manuales de buenas practicas del área espacial [3], casi todas ellas comunes y aplicables a las buenas practicas conocidas en la industria de software general.

2. Lenguajes de decodificación, comando y control

Una practica común, en el área espacial, cuando se trata de comandar un satélite, es el uso de secuencias de comandos sobre estructuras de control. Empresas y agencias espaciales en todo el mundo han desarrollados sus propios lenguajes e intérpretes para cumplir con este objetivo [4]:

- STOL: Satellite Test and Operation Language. Desarrollado por la Nasa y ampliamente utilizado en varias misiones.
- PLUTO: presente en algunas misiones de la ESA (Satellite Control and Operation System 2000).
- Otros: desarrollados o utilizados por diferentes compañias SOL(GMV), CCL (Harris), PIL (Astrium), SCL (ICS).

Estos lenguajes son típicamente propietarios e incompatibles entre sí [5], dificultando las migraciones entre sistemas y la posibilidad de compartir procedimientos entre distintas misiones.
El FS2017 [1] fue el primer segmento de vuelo utilizado como prueba experimental para el desarrollo de los sistemas planteados en este trabajo y se optó por usar un lenguaje de propósito general en lugar de crear un lenguaje especifico o utilizar los existentes en CONAE (Comisión Nacional de Actividades Espaciales). El cuadro 1 compara las ventajas y desventajas de cada tipo de implementación.

En el caso del proyecto integrador la opción desarrollar un lenguaje propio de la misión se descartó por no disponer del tiempo que implica desarrollar y validar un intérprete de propósito específico. La opción de utilizar los intérpretes de CONAE quedó relegada por sobre la opción Python, con una base de usuarios mucho más grande. El índice de popularidad TIOBE [6] ubica al lenguaje en la

[1] Se refiere a la propiedad de los datos para que estos sobrevivan de alguna manera.

Cuadro 1. Comparativo Lenguajes Propósito especifico vs Propósito General

Propósito específico.	**Propósito general.**
Ventajas	**Desventajas.**
Pueden ser más amigables para usuarios no programadores. Pueden presentar adaptaciones especificas a los problemas de operación del satélite.	Demasiado poderoso. Menos amigable a la lectura.
Desventajas	**Ventajas.**
Es una tecnología propietaria. Problemas de portabilidad. Requiere desarrollo propio. Requiere mantenimiento y las mejoras o ampliaciones son costosas. Muy baja base de usuarios.	Portable. Mas poderoso que las versiones de propósito específico. Gran cantidad de usuarios.

posición 5. Incluso en su versión más popular, las alternativas de propósito especifico, al ser parte de una única o un grupo de organizaciones siquiera figuran entre las primeras 50.
La utilización de Python sobre otra opciones de propósito general viene determinada por su popularidad (La variable probablemente mas importante) que lo ubica en el primer puesto entre los interpretados según el índice TIOBE, su utilización por organizaciones privadas dedicadas a desarrollo de software de segmento terreno, su carácter multiplataforma y el dominio del mismo entre estudiantes y científicos en general.
El uso de los interprete disponibles en CONAE sufre las restricciones de cualquier lenguaje de uso interno y de propósito especifico, documentación, base de usuarios y herramientas acotadas y el esfuerzo que implica portarlo a distintas nuevas plataformas como indica [4].

3. Persistencia de datos

Las definiciones de los datos (Alarmas, variables de telemetría, comandos disponibles, etc.) como los datos en sí, en cualquier nivel de procesamiento, están administrados por un motor de base de datos relacional. Para la telemetría satelital existen entidades para el almacenamiento de los datos crudos, independientemente del satélite al que pertenezcan, la fuente o el origen de la ingestión. A medida que estos datos son procesados pasan a tablas donde están las variables de ingeniería ya procesadas y disponibles para cualquier modulo que las necesite, incluyendo el procesador de comandos.
Previendo para futuras versiones la implantación de tablas que permitan el acceso eficiente a datos históricos, KPIs (Key Performance Indicators) y desnormalizaciones que permitan el almacenamiento de largo plazo.
Todo el acceso a datos se realiza exclusivamente a través de un ORM, el uso de esta capa intermedia ofrece además de una mejora en la productividad, una independencia del proveedor del motor de base de datos.

La solución propuesta en términos de persistencia, prácticamente un estándar de facto en la industria del software, aun no tiene presencia en muchos centros de control de misión. Su incorporación, exploración y dominio representan un valor agregado al trabajo.

4. Interfaces

En términos de interfaces tanto internas como externas en el área espacial abundan las soluciones *ad-hoc* o implementaciones sobre tecnologías de escasa penetración.
Un caso remarcable es el de CORBA [7] (Common Object Request Broker Architecture), que, si bien es un estándar propuesto por varios jugadores de la industria del software, no ha sido adoptado masivamente y se vio rápidamente relegado por implementaciones basadas en HTTP (Hypertext Transfer Protocol) [8].
Las interfaces del sistema propuesto se basan completamente en esta última tecnología, por ejemplo, toda ingestión de telemetría se realiza mediante un servicio REST que puede ser consumido tanto por módulos de terceros como por módulos internos que adapten formatos, tecnologías o características de antiguas implementaciones.
La figura 1 muestra como los adaptadores ajustan dos tecnologías diferentes (TCP/IP y FTP) a una entrada normalizada sobre servicios REST. Así mismo la evolución del sistema a una arquitectura distribuida prevé distribución, sincronización y comunicación de tareas mediante brokers basados también en http.

5. Arquitectura

La solución propuesta para el segmento terreno evoluciona desde su concepto cliente-servidor tradicional a una arquitectura de sistema distribuido multiprocesador. Un subconjunto finito de procesos se ejecuta de manera distribuida dentro de la infraestructura de hardware disponible. La tabla 2 enumera algunos de los servicios a ejecutar.

Esta arquitectura permite escalar horizontalmente en función de las misiones que se puedan ir sumando al centro de control. La figura 1 muestra un diagrama de alto nivel donde se pueden observar adaptadores que ajustan la telemetría provista por distintas tecnologías a un estándar REST (Representational state transfer). Los procesos 1, 2 ... n son los especificados en la tabla 2.

En términos de desarrollo el lenguaje utilizado es Python, normalizando en un único lenguaje el desarrollo de los módulos de software, los scripts de comandos y las funciones de calibración.
El framework utilizado es Django, lo que motiva un análisis y diseño orientado a objetos con un alta productividad tanto desde el framework como en el lenguaje [10].

Cuadro 2. Servicios distribuidos

Proceso	**Descripción**
Central.	Servicio central, responsable de ofrecer las interfaces, tanto sea de usuario como las API de comunicación.
Procesador de Telemetría (n).	Proceso responsables de la decodificación de la telemetría cruda.
Procesador de Telecomandos (n)	Procesos responsables de la ejecución de comandos
Servicios CODS	Responsable de generar y actualizar las efemérides, eclipses y oportunidades de contacto (Pasadas).
Procesos de datos	Desnormalizaciones, consolidados, estadísticas y tendencias[9].
Procesos de calibración	Verificar novedades en funciones de calibración.
Otros	Procesos adicionales a ser relevados.

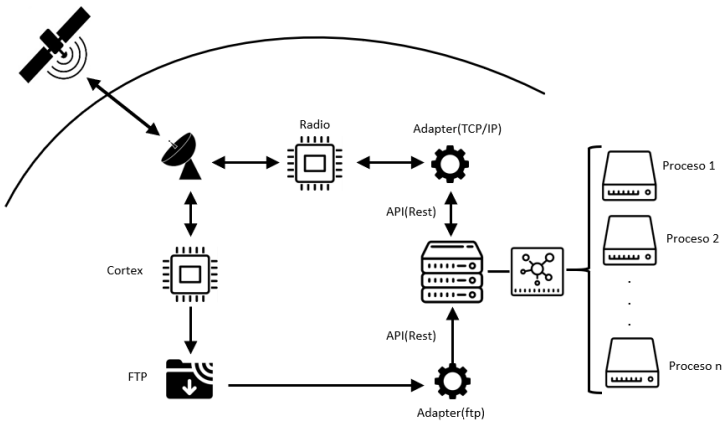


Figura 1. Arquitectura conceptual

A esto ultimo se debe agregar la capacidad de incorporar a los scripts de comandos y de calibración el acceso a herramientas de comunicación como Sockets, http, RPC (Remote Procedure Call), Web Services, email y a bibliotecas de manejo de archivos XML entre otros como explica [4].

6. Telemetría

Si bien la arquitectura del segmento terreno es agnóstica de la fuente de datos se han utilizado como segmento de vuelo el FS2017 y datos de la misión LITUANICASAT 2. Es común en este tipo de sistemas recibir los valores crudos o *raw*, para poder transformar esta telemetría cruda a valores de ingeniería es necesario someterlos a una función que realice la conversión, la cual también podría cambiar en función del desgaste o recalibración del sensor. Estos ajustes pueden ir desde un simple ajuste por función lineal hasta una discretización de valores por tablas de *look-up*. El sistema debe ser lo suficientemente flexible como para permitir aplicar cualquier función de transformación a toda variable de telemetría y permitir ajustarla en el tiempo.

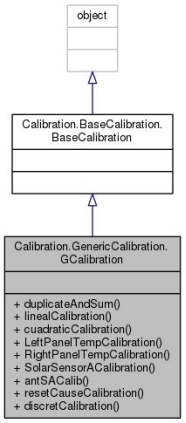


Figura 2. Calibración

En lugar de utilizar interpretes de desarrollo propio, la solución propuesta, utiliza técnicas de reflexión de software para cargar en tiempo de ejecución la función seleccionada de calibración y aplicarla al valor *raw* extraído de la trama de telemetría.
La función de calibración puede ser cualquier secuencia de comandos programable en python sin ningún tipo de restricción más allá del tiempo de procesamiento. Todas las bibliotecas y estructuras de datos están disponibles incluyendo funciones matemáticas e incluso el acceso a la base de datos completa mediante ORM. Al tener acceso a la base de datos se pueden obtener coeficientes actualizados de calibración permitiendo:

- **Reutilizar funciones:** Es posible crear una única función para ajustes típicos (lineal, cuadrático) y reutilizarla con distintos coeficientes según la variable de telemetría requiera
- **Calibración fina:** Creada una función es posible modificar los coeficientes para a medida que el sensor presenta desgaste.

Para que un método sea considerado de calibración o ajuste debe estar desarrollado como método de una clase heredada de *BaseCalibration*. El software realiza periódicamente una exploración de todos los métodos públicos de clases derivadas (Ver procesos tabla 2) de *BaseCalibration* y los disponibiliza para su

aplicación a las distintas variables de telemetría. La figura 2 muestra la clase *GCalibration*, heredada de *BaseCalibration*, donde se implementan algunos métodos de calibración incluido el mencionado *linealCalibration*.
El siguiente fragmento de código muestra parte de la implementación del método, encargado de llamar a la función de calibración para todo nuevo valor *raw*.
Calibración de variable raw

```
class TlmyVar(models.Model):
  def setValue( self,
raw,
saveifchange=False,
dt=datetime.now(utc)):

    value = self.tlmyVarType.getValue()
    if  (raw!=self.tlmyVarType.lastRawValue) or
        (self.tlmyVarType.lastUpdate==None):
      self.tlmyVarType.lastRawValue = raw
if self.tlmyVarType.calibrationMethod:
  if not self.tlmyVarType.calibrationLogic:
    klass =
      globals()
      [self.tlmyVarType.calibrationMethod.aClass]

    instance = klass()

    methodToCall =
      getattr(instance,
      self.tlmyVarType.calibrationMethod.aMethod)
    self.tlmyVarType.calibrationLogic = methodToCall
  else:
    pass #Calibracion ya cargada

          ...
```

(Extracción parcial del método setValue)

Si el valor *raw* para una variable determinada no cambió desde la última actualización y tampoco se recargo la función, entonces se guarda el registro de la recepción, pero no se aplican las calibraciones. Esto último hace ganar eficiencia ya que no se vuelve a transformar cuando se sabe que el resultado será el mismo. La carga del método de calibración a un atributo del tipo de variable de telemetría se realiza, si no ocurren cambios, una única vez durante la ejecución. Esto también permite mejorar los tiempos de ejecución ya que evita la carga por cada valor de telemetría recibido. Con una frecuencia configurable, el software analiza si alguna función de calibración fue actualizada, si así ocurriera se procede al limpiar el atributo que contiene la función para forzar su recarga.

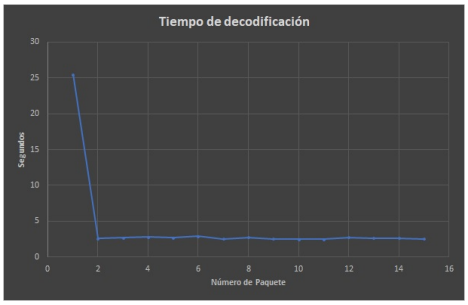


Figura 3. Tiempo de decodificación

La figura 3 muestra los tiempos de procesamiento para un conjunto de simulación de 15 paquetes con 5000 variables de telemetría cada uno, cantidad compatible con un satélite científico de envergadura (Ejemplo SABIAMAR `http://www.conae.gov.ar/index.php/espanol/introduccion-sace`).
La primera decodificación demora más que las siguientes dado que tiene que realizar la carga de las funciones de transformación. La primera telemetría recibida difiere de la anterior, ya que al ser la primera calibración la función no está precargada. Luego el tiempo se estabiliza entre 3 y 3.5 segundos para el conjunto de las 5000 variables con una probabilidad de cambio de 10 % entre muestra y muestra. Teniendo en cuenta que para este tipo de satélites un valor promedio de frecuencia entre paquetes es de 4 a 8 segundos entonces el hardware-software utilizado seria justo suficiente. Se debe tener en cuenta que el tiempo de procesamiento incluye además la verificación de que los valores de las variables de ingeniería estén dentro del rango de seguridad y el almacenamiento en el histórico. Las pruebas han sido realizadas en máquinas virtuales sobre equipos de escritorio.

7. Telecomandos

[2] En el caso del FS2017 los telecomandos deben ser enviados al segmento de vuelo por el mismo canal en donde se recibe la telemetría, TCP/IP puerto 3210. Los comandos deben ser codificados en una trama AX.25. Para permitir la creación de scripts de comandos, de la misma forma que con la telemetría (6) se utilizaran las capacidades de reflexión para analizar, en tiempo de ejecución los scripts a ejecutar. Los scripts de comandos pueden ejecutarse por acción explicita de un operador o porque fueron aplicados a una pasada. El operador tiene accesos a las bibliotecas de comandos y operación, las cuales están desarrolladas sobre el mismo lenguaje e intérprete de la aplicación. Las bibliotecas tienen pleno acceso al ORM de donde puede obtener el diccionario de comandos y los valores

[2] Los módulos que procesan comandos se están en proceso de desarrollo

de telemetría si necesitara aplicar condicionales que dependieran del estado del segmento de vuelo u otro valor disponible en el sistema.

8. Resultados

El lenguaje (python) es sencillo aprender, pero por sobre todo, está disponible una extensa comunidad con ayudas ante los problemas o desafíos que puedan aparecer. Para las posibles limitaciones[3] es fácil encontrar soluciones alternativas (workarounds). Esto presenta un fuerte contraste con las experiencias realizadas durante la maestría con lenguajes de propósito especifico que requirieron de la asistencia constante de un experto.
Las herramientas de depuración y análisis se mostraron poderosas, incluso superando las expectativas mas allá de las limitaciones en las capacidad de Code Insight/IntelliSense naturales de un lenguaje interpretado.
La solución fue ejecutada en ubuntu linux 15.10 como en windows 10.0 y windows 7.0. Se realizaron pruebas sobre dispositivos físicos y virtuales. En todos casos el software se ejecutó sin cambios. En este punto tanto el interprete como el framework mostraron uno de sus atributos más destacables.
El ORM de Django se mostró estable, se registraron pocos problemas o *bugs* y cuando ocurrieron se encontró información disponible que lo documentaba. Para la creación y modificación de modelos, su flujo divido en dos etapas denominadas *Make Migrations* y *Migrate* permitió el trabajo colaborativo y con múltiples servidores, demostración empírica de la madurez del framework.
El acceso a datos requiere de la práctica de convenciones propias (*Convention over configuration/coding by convention*) que, aunque muy bien documentadas, afectan la curva de aprendizaje. Por otro lado el modelo no acepta atributos privados, cuestión que afecta el encapsulamiento. La capacidad del ORM para trabajar con polimorfismo está limitada y se requiere añadir paquetes especiales, este último punto se pueden considerar como una limitación de relevancia.
Las capacidades de reflexión fueron suficientes para cumplir con los objetivos y los tiempos de respuesta no presentan una limitación a su implementación.La reutilización de métodos previamente cargados colabora en mantener los tiempos entre márgenes aceptables.

9. Conclusiones

Los lenguajes específicos del sector espacial fueron creados hace décadas. Durante los 70, si bien existían opciones interpretadas, están no eran de uso extendido. En la actualidad existen varias opciones (Python, Perl, VBScript, JavaScript) con una amplia cantidad de usuarios de base y múltiples proyectos

[3] Limitaciones que siempre un lenguaje, framework o biblioteca pueden tener. En el caso del proyecto las mas salientes fueron los problemas explicados en el documento relacionados con el ORM y polimorfismo, problemas de administración ineficiente de memoria del interprete o manejo de tipos no contemplados como el Unix Timestamp)

que avalan su robustez lo que plantea un nuevo escenario a evaluar en lo que respecta a scripts de comandos o calibración de telemetría. Las pruebas y simulaciones con telemetría (6) demostraron que es posible decodificar variables de telemetría para una misión de tamaño medio utilizando hardware COTS y lenguajes de propósito general. Las opciones ampliamente probadas no justifican el desarrollo de un lenguaje propietario, no tener una base de usuarios implica ausencia de documentación, soporte, herramientas y por sobre todo recursos como explica [11] refiriéndose a ADA.
En términos de persistencia de datos, la vinculación entre el ORM y el motor relacional ha favorecido la productividad dado que no fue necesario escribir ninguna consulta del tipo Structured Query Language (SQL), delegando tanto las búsquedas como las inserciones, actualizaciones. y borrados. Tampoco fue necesario capacitar a todos los recursos en estas tecnologías. Se ha cuidado de eliminar cualquier consulta directa al motor, para favorecer la compatibilidad con múltiples motores, factor que también ofrece un adicional en términos de productividad evitando la necesidad de crear un selector de sintaxis según motor.
Las interfaces aun no se han explotado en su totalidad. Se espera su aplicación efectiva durante las integraciones con diferentes actores o socios del proyecto.
En el estado actual, la solución permite cumplir los objetivos haciendo uso únicamente de componentes COTS y sobre herramientas de propósito general y uso extendido en la industria del software.

Referencias

1. Pablo Soligo, Ezequiel González, Eduardo Sufán, Emmanuel Arias, Ricardo Barbieri, Pablo Estrada, Alfonso Montilla, José Robin, Javier Uranga, M Cecilia Valenti, et al. Misión cubesat fs2017: Desarrollo de software para una misión satelital universitaria. *WICC 2017*, page 843.
2. José J Ramos Pérez. A design for an advanced architecture of satellite ground segments.
3. Ken Galal and Roger P Hogan. Satellite mission operations best practices. 2001.
4. Gonzalo Garcia. Use of python as a satellite operations and testing automation language. In *GSAW2008 Conference, Redondo Beach, California*, 2008.
5. Geraldine Chaudhri, Jim Cater, and Brad Kizzort. A model for a spacecraft operations language. In *SpaceOps 2006 Conference*, page 5708, 2006.
6. TIOBE Software. TIOBE programming community index, september 2017, 2017. [Online; accessed 26-September-2017].
7. L Foti. Corba technology for ground segment system software development. In *DASIA 98-Data Systems in Aerospace*, volume 422, page 303, 1998.
8. Michi Henning. The rise and fall of corba. *Queue*, 4(5):28–34, 2006.
9. Thomas Morel, Gonzalo Garcia, Mike Palsson, and Juan Carlos Gil. High performance telemetry archiving and trending for satellite control centers. In *SpaceOps 2010 Conference Delivering on the Dream Hosted by NASA Marshall Space Flight Center and Organized by AIAA*, page 2111, 2010.
10. Lutz Prechelt. An empirical comparison of seven programming languages. *Computer*, 33(10):23–29, 2000.
11. II Smith et al. What about ada? the state of the technology in 2003. 2003.

Interleaving Scheduling Algorithm for SLM Transactions in Mode S Surveillance Radar

Oscar Bria[1✉], Javier Giacomantone[1], and Horacio Villagarcía Wanza[1,2]

[1] Research Institute in Computer Science (III-LIDI) - School of Computer Science
National University of La Plata - Argentina
[2] Scientific Research Commission (CIC) - Province of Buenos Aires - Argentina
onb@info.unlp.edu.ar

Abstract. Mode S Secondary Surveillance Radar establishes selective and univocally addressed transactions with aircraft within its coverage while using efficiently the available budgets of time and energy. This benefits are key to supporting high-traffic density. A preliminary interleaving algorithm for the scheduling of Short Length Message transactions is presented and tested under a heavy load simulated scenario.

1 Mode S SSR and Resource Management

Mode S (Selective) is a Secondary Surveillance Radar (SSR) process that allows selective interrogation of aircraft according to the unique 24-bit address assigned to each aircraft. Such selective interrogation improves the quality and integrity of the detection, identification and altitude reporting with the addition of new reports from the aircraft and data-link capabilities. These improvements translate into benefits in terms of safety, capacity and efficiency, benefits which are key to supporting high-traffic density scenarios [1] [2].

The radar resource management function plays a critical role to maximize the radar resource usage for improving performance. In addition to the tracking tasks, the system also includes search and target confirmation tasks. A search task involves looking for new targets in the sky and a target confirmation task confirms the target after it is detected by the search task. Due to the multidimensional nature of radar resource allocation, the problem of optimally determining the process of resource allocations to maximize total system utility is NP-hard [3].

The radar resource management function includes a specific scheduling algorithm for the several transactions of the tracking task [4] [5]. The scheduling algorithm considers waveform, beam shape, type of coding, dwell time, pulse repetition frequency, energy level, the time characteristics of the transactions and aircraft predicted positions. Since the targets move continually, and sometimes evasively, the resource allocation and scheduling decisions must be made frequently and in real-time.

Dwell time in a radar is the time that an antenna beam spends on a target. The beam dwell time of a 2D surveillance radar is derived from the antenna

horizontal beam width and the turn speed of the antenna [6]. In Mode S SSR, during the beam dwell time, there is an alternation of two basic types of tasks, broadcasting and selective[3]. Broadcasting periods include Mode A/C searches and transactions (for compatibility with heritage radars and transponders), and Mode S searches. Selective periods include Mode S selective tracking transactions. Commonly, one-third of the beam dwell time is for broadcasting tasks.

Some characteristics of a Mode S SSR radar system of particular interest in resource management are [7] [8]:

- Each Mode S selective transaction is compose of three phases in sequence: a transmission phase, a waiting phase and a receiving phase.
- Once a transmission or a reception starts, it cannot be preempted.
- The waiting phase is a wasting of time for a transaction.
- Transactions overlapping in the same beam can be interleaved by scheduling the transmit and/or receive phase of one transaction in the wait phase of another transaction.
- The longer the distance between the target and the radar, the higher the energy requirement.

Due to the selective nature of the Mode S transactions, interleaving is mandatory for high-traffic density. The waiting time may change from one dwell to the next depending on the velocity vector of the aircraft relative to the radar. Therefore, the radar should be able to predict the approximate waiting time based on the previous tracking information about the aircraft.

The energy of the transmissions may be modulated by the distance, but this could result non-viable for some radar electronic implementations.

Section 2 shows the characteristics of Mode S transactions. Section 3 present an interleaving algorithm for SLM Mode S transactions. In Section 4 simulation results of the use of the algorithm are presented. Finally, section 5 draws some conclusions and sketches future work.

2 Mode S SSR Transactions

Figure 1 shows the phases of a transaction as part of a Mode S tracking task. The transmission interval is t_{X} while the reception period t_{r} begins after a waiting time t_{W}. The cool-down interval, t_{C}, precedes the transmission interval of any transaction [3]. During this interval, there is no transference, and therefore it contributes to the evacuation of the heat from the active components of the radar. In the algorithm presented here the cool-down is a non conditioned variable. The interval time t_{V} represents the remaining time to the next transactions programmed for the current aircraft task. Even when the next transaction is normally programmed for the next antenna azimuth scan in a rotating surveillance radar, current transaction could be repeated in the present beam dwell for particular reasons.

[3] In the literature [7] these types of periods are called SSR/all-call period and Mode S roll-call period, respectively.

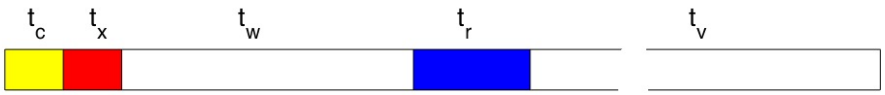


Fig. 1: Transaction Time Intervals.

Transactions are characterized by the duration and coded features of the pairs transmission and response. From an scheduling point of view only the time intervals are relevant. Table 1 shows the possible pair combination encountered for Short Length Messages (SLM)[9].

#	Transaction	t_{X}	t_{r}
1	Surveillance/Surveillance	20μs	64μs
2	CommA/Surveillance	34μs	64μs
3	Surveillance/CommB	20μs	120μs
4	CommA/CommB	34μs	120μs

Table 1: SLM Transaction Types

3 Interleaving Algorithm for SLM Mode S

The phases t_{X} and t_{r} are non-preemptive, since a radar can only perform a single transmission or a single reception at a time. However, t_{c} of one task can be overlapped with t_{r} or t_{w} of another task, since the radar can cool down during the waiting and the receiving period. Allowing the entire duration of a transaction (from transmission start to reception end) to be a non-preemptive job wastes resources and decreases the schedulability of the system [3]. Transactions can be interleaved to improve schedulability. The constructed interleaving may not be optimal in some restricted sense [10] [11] [12], but it must be effective for the target application, and preferably simple and with well-known properties.

Transactions can be interleaved in two ways: (a) properly nested interleaving and (b) improperly nested interleaving. Two transactions are said to be properly nested if one transaction fits inside the waiting time (t_{w}) of another, as in the left transactions in Figure 2. Two transactions are said to be improperly nested when one transaction only partially overlaps with another as illustrated by the right transactions in Figure 2.

Suppose that transaction T1 is improperly interleaved with transaction T2, where T1 starts first. Transaction T1 is called the leading transaction and transaction T2 is called the trailing transaction. Based on the phasing illustrated

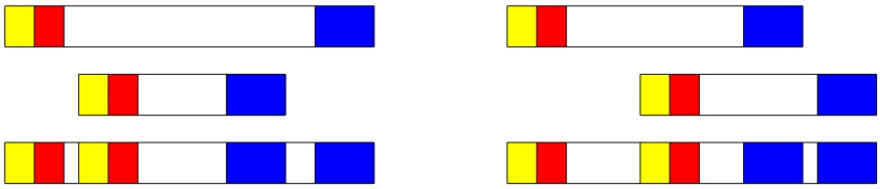


Fig. 2: Properly (left) and Improperly (right) Interleaving.

in Figure 2, the necessary conditions for the interleaving to occur are given by Equations (1) and (2),

$$t_{\mathrm{W1}} \geq t_{\mathrm{C2}} + t_{\mathrm{X2}}, \tag{1}$$

$$t_{\mathrm{C2}} + t_{\mathrm{X2}} + t_{\mathrm{W2}} \geq t_{\mathrm{W1}} + t_{\mathrm{R1}}. \tag{2}$$

A phase offset for a improper interleaving is defined in (3) [13]. The value of the phase offset determines how tightly two nested tasks fit together. The aim is to minimize this offset in reception because it is a useless wasting of time,

$$o_i = t_{\mathrm{C2}} + t_{\mathrm{X2}} + t_{\mathrm{W2}} - (t_{\mathrm{W1}} + t_{\mathrm{R1}}). \tag{3}$$

Cool-down time t_{C2} could have any positive value larger than a prescribed or derived minimun. For SLM $t_{\mathrm{R}} > t_{\mathrm{X}}$; if $t_{\mathrm{W1}} \geq t_{\mathrm{W2}}$ is taken, then o_i can be fixed to 0 and the following is derived,

$$t_{\mathrm{C2}} = (t_{\mathrm{W1}} - t_{\mathrm{W2}}) + (t_{\mathrm{R1}} - t_{\mathrm{X2}}) > 0, \tag{4}$$

$$t_{\mathrm{C2}} \geq \min(t_{\mathrm{C2}}) = t_{\mathrm{R1}} - t_{\mathrm{X2}} > 0. \tag{5}$$

It means that for SLM, when interleaving condition are satisfied, the algorithm can forces offset $o_i = 0$ because the value of t_{C2} remains > 0 for any other condition.

The implemented algorithm starts with the transaction of the largest waiting time t_{W1}, and attempts to interleave it, avoiding overlapping, with the transaction with the largest possible t_{W2} smaller than that of the leading transaction based on the stated conditions (1) and (2) with $o_i = 0$. Cool-down time t_{C2} is a non conditioned variable derived in the process.

The algorithm repeats the process[4] taken the actual trailing transaction as the next leading transaction until it reaches the transaction with the smallest t_{W} that can no longer be interleaved, or all transactions are interleaved to form a single virtual transaction called a cycle [7].

If there is a backlog of transactions, the process is repeated until all transactions are included in as many cycles as necessary.

[4] The conditions of equations (1) and (2) have to be suitably modified.

The improperly nested algorithm ensures that in any cycle the transmissions and receptions are equally sequenced. That is not the case for the properly nested alternative.

As was mentioned, for any type of SLM transaction t_{x} is shorter than t_{r}. As a consequence of that condition, it was demonstrated that the sequence of receptions of any cycle does not have gaps[5]. Meanwhile the sequence of transmissions of any cycle has transmission silences that contribute to cool down the active components of the radar[6].

4 Algorithm Test for Heavy Load Simulated Scenario

SLM are used in level 1 (CommA) and level 2 (CommB) data-link services in Mode-S SSR, particularly in the GICB (Ground Initiated CommB) protocol as used in the Mode S Enhanced Surveillance (EHS)[14].

Even when SLM includes different types of transaction, entirely GICB transactions are used for testing, as recomended by Eurocontrol; this corresponds to transaction #4 in Table 1. The distribution of aircraft in the high-traffic density scenario follows the non uniform histogram of Table 2 [15]. A uniform random distribution is applicable in each range band.

Range NM	5-10	10-20	20-40	40-60	60-80	80-90	90-130	130-150
Distribution	1	3	12	7	7	2	6	10

Table 2: Aircraft Distribution in a Beam Dwell

Complementary parameters of the scenario are:

1. Beamwidth: the above 48 aircraft are distributed in a 3.5° sector.
2. GICB rate: 1 GICB per aircraft.
3. Minimum range: 5 NM.
4. Maximum range: 150 NM.
5. Scan rate: 4 seconds.

To relax the demand on the tracker accuracy, a guard of 12 μs is added between each response. The guard allows almost 1 NM error in the tracker estimation, which is loose.

The output of the algorithm for one realization of the scenario described is shown in Figure 3. The bars from below are the transmissions while the bar from above are the receptions. In this case, the schedule is composed by as much as 9 cycles. Relevant here is the elapsed time consumed for the entire schedule that is less than 14 ms.

The left peak in Figure 4 shows the normalized histogram of the elapsed time consumed by 5,000 random runs of the algorithm simulating 48 original

[5] Except for the guards to cover the estimates of the trackers.

[6] The consideration of the constraints over the working cycle of the active components of the radar is beyond the scope of this work.

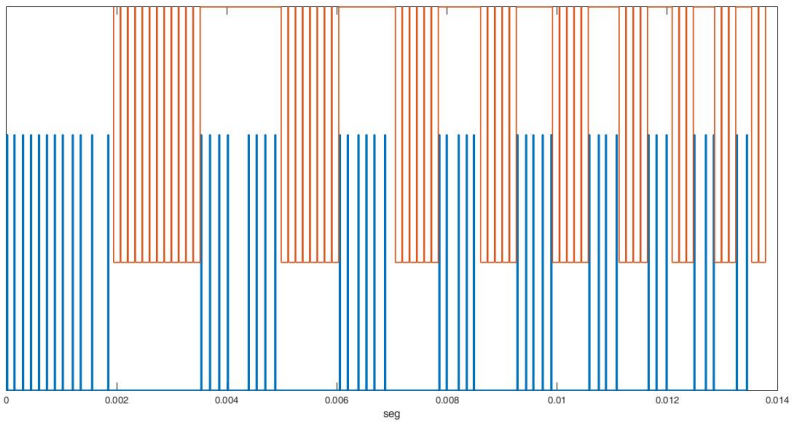


Fig. 3: Improperly nested scheduling for a 48 transactions scenario.

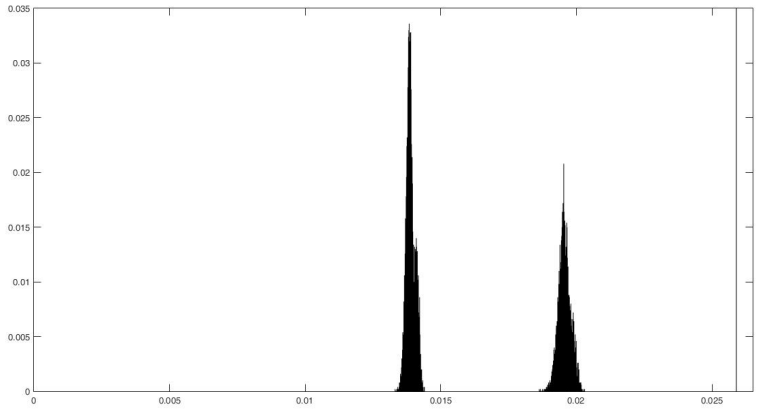


Fig. 4: Schedule/s elapsed time normalized histograms for 5,000 random run.

transactions. The mean value of this data is 13.9 ms, the standard deviation is 158.3 μs, the maximum value is 14.4 ms, and the minimum is 13.3 ms.

Given: beamwidth, $\theta = 3.5°$; scan period, $\tau = 4$s; fraction of dwell for Mode S transactions, $f = 2/3$. The available time for Mode S transactions in a dwell, A is,

$$A = f\tau\frac{\theta}{360°} = \frac{2}{3}\ 4\ \frac{3.5}{360}\ s = 25.9\ ms, \tag{6}$$

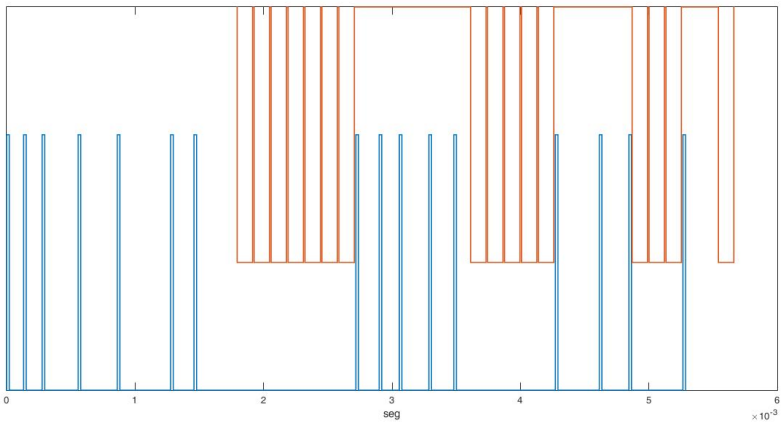


Fig. 5: Improperly nested 16 re-scheduled transactions from original scenario.

Rougly, it can be said that many times the schedule consumes more than 50 % of the available time (the continuous vertical line in the right side of Figure 4)[7]. As a consequence, two complete schedules are not allowed to be included in the same dwell, in several occasions.

Actually, the repetition of a transaction within the dwell is mandatory when not answer is received or any reception is pointed wrong as a consequence of a coding error detection [7].

Suppose that the initial probability of right reception of a transactions is $p_i = 68.38\%$. That means that 16 out of 48 receptions are misleading or wrong and the corresponding transactions have to be repeated during the present dwell in a new schedule. If that is possible, the final probability p_f for the transaction load in the dwell would be the prescribed [15],

$$p_f = p_i + (1 - p_i)p_i = 2p_i - {p_i}^2 = 90\% \tag{7}$$

The right peak of Figure 4 shows the elapsed time consumed by 5,000 random runs of the algorithm for two consecutive schedules, i.e., a first schedule of original 48 transactions and a second schedule of 16 transactions peaked out randomly from the original. The mean value of this data is 19.5 ms, the standard deviation is 227.4 μs, the maximum value is 20.2 ms, and the minimum value is 18.6 ms. As can be seen in Figure 4, the simulated data is far to the left of the available dwell time. Figure 5 is sample of a 16 transactions schedule consisting of 4 cycles and elapsing less than 6 ms.

[7] A more precise probability evaluation, possibly including inflation of Gaussian distribution models [16], is beyond the goals of this article.

5 Conclusion and Future Work

A preliminary improperly nested interleaving algorithm has been presented for the scheduling of Short Length Message transactions in the context of Mode S Secondary Surveillance Radar. The main characteristics of the algorithm are:

- Responses are in the same order as transmissions, for any schedule.
- Responses are linked one after the other without gaps, for any cycle.
- Cool-down time is always greater than zero, for any transaction.

The algorithm has been tested under a heavy load simulated scenario for 5,000 random runs. Elapsed time for the original 48-transaction schedule is barely greater than 50% of the available dwell time, making it possible to add a repetition schedule just shorter than the original. Then, for a probability of right reception as low as 68.38%, the re-scheduling of 16 out of 48 transactions is not only possible but also enough to accomplish a compound probability of 90% as prescribed [15].

Future work may include:

- Analysis of the requirements that the algorithm imposes under heavy load to the radar active components used during transmission.
- Modification of the algorithm to incorporate Long Length Messages as well.
- Study of the functional relation among cool-down, energy and distance between the radar and the aircraft.
- Determination of Gaussian bounding distributions from sample distributions of elapsed times.

References

1. ICAO (International Civil Aviation Organization). *Annex 10, Third Edition of Volume IV.* (2014)
2. ICAO (International Civil Aviation Organization). Manual on the Secondary Surveillance Radar (SSR) Sytems. In: Doc 9684 AN/951 (2004)
3. Ghosh, S., Hansen, J., Rajkumar, R., Lehoczky, J.: Integrated resource management and scheduling with multi-resource constraints. In: IEEE International Real-Time Systems Symposium, Lisbon, Portugal (2004)
4. Ding, Z.: A Survey of Radar Resource Management Algorithms. In:IEEE Canadian Conference on Electrical and Computer Engineering - CCECE - Niagara Falls, ON, Canada (2008)
5. Mir, H., Wilkinson, J.: Task Scheduling Algorithm for an Air and Missile Defense Radar. In: IEEE Radar Conference, Rome, Italy (2008)
6. Richards, M., P., Scheer, J., Holm, W.: *Principles of Modern Radar, Basic Principles.* Scitech Publisher (2015)
7. Orlando, V., Drouilhet, P.: Functional Description of Mode S Beacon System. In: Project Report ATC-42 Revision B, Lincoln Laboratory, MIT (1982)
8. Moo, P., Ding, Z.: *Adaptive Radar Resource Management.* Elsevier Academic Press (2015)

9. Stevens, M., Ding, Z.: *Secondary Surveillance Radar.* Artech House (1988)
10. Shih, C., Gopalakrishnan, S., Caccamo, M., Sha, L.: Template-Based Real-Time Dwell Scheduling with Energy Constraints. In: Proceedings of the IEEE Real-Time and Embedded Technology and Applications Symposium, Toronto, Canada (2003)
11. Charlish, A., Nadjiasngar, R.: Quality of Service Management for a Multi-Mission Radar Network. In: IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Cancun, Mexico (2015)
12. Sgambato, P., Celentano, S., Di Dio, C., Petrillo, C.: A flexible on-line scheduling algorithm for multifunctional radar. In: IEEE Radar Conference, Philadelphia, PA, USA (2016)
13. Ghosh, S., Rajkumar, R., Hansen, J., Lehoczky, J.: Integrated QoS-Aware Resource Management ans Scheduling with Multiple-Resource Constraints. In: Real-Time Systems Journal, 33:7-46, Springer (2006)
14. ICAO (International Civil Aviation Organiztion). Manual on Mode S Specific Services. In: Doc 9688 AN/952 (2003)
15. EUROCONTROL (European Organisation for the Safety of Air Navigation): European Mode S Station Functional Specification. In: SUR/MODES/EMS/SPE-01 (2005)
16. Blanch, J., Walter, T., Enge, P.: A MATLAB Toolset to Determine Strict Gaussian Bounding Distributions of a Sample Distribution. In: Proceedings of the 30th International Technical Meeting of The Satellite Division of the Institute of Navigation (2017)

Reconstrucción topográfica a partir de iluminación estructurada. Su aplicación en Radioterapia Externa

L. Garavaglia[1,2], L. Mairal[2,4], J.Runco[2,3]

[1] INIFTA Inst.de Investig.Fisicoquímicas Teóricas y Aplicadas UNLP – CONICET
[2] Depto.de Física – Facultad de Cs.Exactas – UNLP
[3] IFLP – Instituto de Física La Plata – CONICET
[4] Mevaterapia Centro Médico, Buenos Aires, Argentina
{runco@fisica.unlp.edu.ar}

Resumen: En tratamientos de radioterapia el objetivo es entregar al tumor la dosis de radiación prescripta por el médico, minimizando la irradiación de los tejidos sanos. Para ello el paciente debe ser ubicado en la camilla en la posición respecto del equipo establecida en la etapa de planificación. En tratamiento de cabeza y cuello suele usarse una máscara termoplástica personalizada como accesorio de inmovilización y posicionamiento. La iluminación estructurada permite obtener las dimensiones espaciales de un objeto iluminándolo utilizando un patrón de franjas y fotografiándolo desde un lugar predeterminado. Si se conocen las dimensiones del objeto, puede calcularse la forma de las franjas que debe proyectarse para que solo cuando el mismo ocupe su posición conocida inicialmente, se vea en la fotografía el patrón original de franjas. En tratamientos de radioterapia, esta técnica puede usarse para mejorar la repetitividad del posicionamiento del paciente en la camilla de tratamiento ya que durante el proceso la cara del paciente sufre modificaciones. El presente trabajo muestra el desarrollo de una herramienta de software que permite verificar cuantitativamente el correcto posicionamiento de un objeto en las condiciones experimentales correspondientes a la geometría de tratamientos de radioterapia externa y detectar simultáneamente eventuales cambios morfológicos de dicho objeto. El monitoreo del posicionamiento del paciente mediante esta técnica tiene tres ventajas importantes: en términos de protección radiológica no agrega dosis de radiación frente a las placas verificadoras, podría reemplazar a las máscaras de inmovilización y facilita la verificación durante el tratamiento.

Palabras clave: Procesamiento de señales, sensores, radioterapia, iluminación estructurada.

1. Introducción

Radioterapia es el uso de las radiaciones ionizantes para tratar el cáncer, que junto con la Cirugía y la Quimioterapia constituyen las tres principales modalidades médicas que tienen esta finalidad [1]. La Radioterapia Externa se basa

en irradiar el tejido maligno desde el exterior del paciente, siendo el acelerador lineal de electrones (LINAC) el equipo más usado para ello. La Braquiterapia en cambio, es la rama de la Radioterapia que entrega la dosis de radiación en contacto o estrecha proximidad con el tejido enfermo, haciendo uso de fuentes radiactivas [2].

Las radiaciones ionizantes alteran el normal funcionamiento de los tejidos vivientes, pudiendo impedir la proliferación de las células y llegando incluso hasta provocar la muerte de las mismas [3]. Es por esto que uno de los principios fundamentales de la Radioterapia es entregar la dosis prescripta al tumor, minimizando la irradiación de los tejidos sanos [4]. Para cumplir este requerimiento básico, se debe reproducir la posición del paciente, en la camilla de tratamiento, establecida en la adquisición de las imágenes tomográficas indispensables para la planificación del tratamiento radiante. Es de fundamental importancia que esta posición se repita en cada fracción de tratamiento con la precisión y exactitud según las tolerancias (en general del orden de algunos mm) correspondientes a la técnica de irradiación aplicada [5].

La iluminación estructurada permite obtener las dimensiones espaciales de un objeto iluminándolo con un patrón de franjas y fotografiándolo desde un lugar predeterminado [6]. Pero si se conocen las dimensiones del objeto, como en el caso de un paciente de Radioterapia, esta técnica puede usarse para mejorar la reproducibilidad y repetitividad del posicionamiento del paciente. Cambios en los contornos externos del paciente, como los producidos por una Quimioterapia coadyuvante o el uso de corticoides, podrían detectarse antes del inicio de una fracción, lo mismo que los desplazamientos de la posición correcta de tratamiento.

Como ya se ha mencionado, la posición del paciente sobre la camilla de tratamiento debe reproducir lo más fielmente posible la posición de referencia que se tenía durante la adquisición de imágenes, obtenidas típicamente mediante tomografía computada de rayos X. Con estas imágenes se inicia el proceso de planificación, donde se establece la posición de tratamiento en la que el campo de radiación atraviesa la región donde se encuentra la lesión, y llevando a cabo los cálculos de los tiempos de irradiación.

Los técnicos operadores de equipos de Radioterapia se encargan de posicionar al paciente en la unidad de tratamiento haciendo coincidir el isocentro del Linac con el isocentro correspondiente a la planificación de su tratamiento. Para ello se utiliza la proyección de láseres visibles con los tres puntos de referencia tatuados sobre la piel del paciente, dos laterales y uno anterior. El cruce de las líneas rectas luminosas de los láseres del bunker de tratamiento permite visualizar el isocentro del Linac que se hace coincidir en el paciente, con la ayuda de los tatuajes de referencia y según planificación (figura 1). Toda inexactitud en el posicionamiento conlleva una inadecuada distribución de dosis y por lo tanto es crucial alcanzar una precisa repetitividad de la posición correcta a lo largo de las fracciones. En tratamientos de cabeza y cuello suele usarse una máscara inmovilizadora termoplástica personalizada (figura 2). Las máscaras termoplásticas deben ajustarse perfectamente a la superficie del paciente. El plan de tratamiento



Figura 1. Láseres rojos de posicionamiento cruzándose en el isocentro con paciente sobre la camilla de tratamiento

tiene en cuenta la influencia dosimétrica de dicho elemento externo, y si aparecen "gaps" de aire entre este y la piel del paciente, la distribución de dosis que efectivamente recibe el paciente se aleja de la prescripción [7]. Algunos pacientes pueden inflamarse o adelgazar notablemente durante el curso de la terapia, y en ciertos casos, la máscara puede hasta dejar de cumplir su función inmovilizadora. Siendo la proyección de franjas una técnica segura desde el punto de vista de la

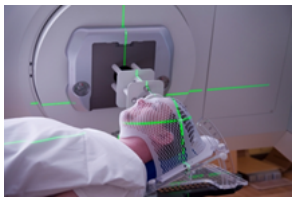


Figura 2. Paciente con máscara termoplástica en posición de tratamiento

radioprotección, es decir que no existe radiación ionizante involucrada, su uso puede repetirse tantas veces sea necesario. Incluso el monitoreo continuo de la posición del paciente es una potencial aplicación de la técnica a fin de evitar movimientos "intra-fracción" fuera de la tolerancia. Esto último es particularmente útil en el supuesto caso de usarla en reemplazo de la máscara inmovilizadora.

2. Descripción de la técnica y sistema de medida

La obtención de la topografía de un objeto mediante iluminación estructurada comienza por posicionar a este último sobre una superficie plana, para luego iluminar ambos plano y objeto con un patrón luminoso dado[8]. Si el plano y el objeto iluminados se observan siempre desde un punto fijo en el espacio, el patrón luminoso se modifica de manera particular según sea la forma del objeto. De esta manera, el patrón así modificado contiene la información topográfica del objeto (figura 3). La virtud de esta técnica reside en obtener las dimensiones espaciales de objetos tridimensionales sin necesidad de hacer contacto con ellos, sino simplemente observándolos. El patrón luminoso puede generarse de distin-

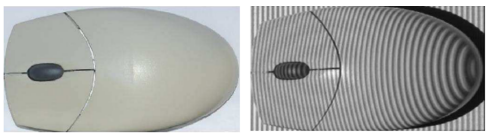


Figura 3. a). Fotografía de un mouse sobre una superficie plana de referencia. b). El mouse y el plano de referencia de la foto a) iluminados con un patrón de franjas de perfil sinusoidal.

tas maneras, por ejemplo iluminando adecuadamente una diapositiva de franjas con variación de niveles de gris siguiendo un perfil sinusoidal, dientes de sierra o rectangular. También se puede obtener un patrón de franjas con perfil sinusoidal por interferencia, usando láser. Otra forma comúnmente usada es mediante un proyector (cañón). Con un dispositivo de este tipo se tiene la posibilidad de generar por computadora diferentes tipos de patrones, sólo limitados por el tamaño de pixel y la profundidad de campo (la distancia dentro de la cual está en foco la proyección), ya que la iluminación no se restringe a un plano, sino que justamente se realiza sobre un objeto tridimensional.

Uno de los patrones más simples es la grilla de Ronchi, que consiste en franjas uniformemente espaciadas, con una distribución de intensidad luminosa tipo escalón, es decir que en una grilla blanco y negro, se alternan las franjas blancas y negras sin aparición de tonos grises (figura 4). El punto de observación se resuelve mediante el uso de un proyector y una cámara fotográfica ubicados de manera que sus ejes ópticos se encuentren perpendiculares al plano de proyección de la grilla de Ronchi.

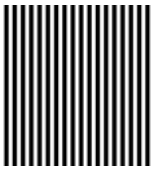


Figura 4. Grilla de Ronchi en blanco y negro.

La figura 5.a) muestra una vista superior, para el caso en el que se observa con la cámara un punto localizado en s, que se proyecta sobre una de las franjas. P y C simbolizan respectivamente al proyector y la cámara, P' y C' los puntos sobre el plano de referencia a donde llegan los rayos que representan los ejes ópticos que también se ilustran. Luego, al colocar un objeto por sobre el plano de referencia y según el punto de vista de la cámara, el punto s parece situarse en b (figura 5.b).

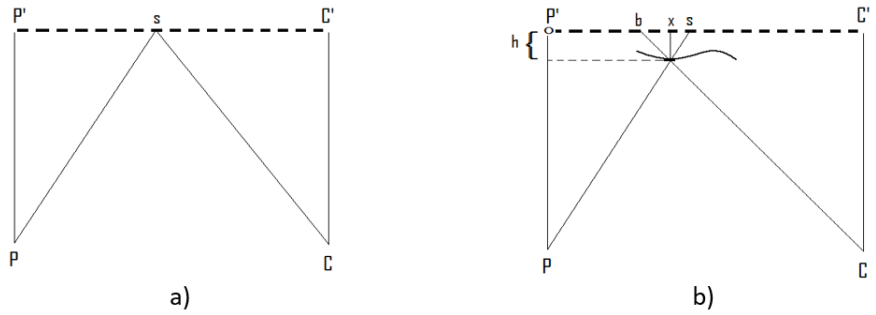


Figura 5. Vista superior del punto s proyectado sobre una de las franjas del patrón. P y C simbolizan las ubicaciones del proyector y la cámara respectivamente.

Finalmente, la altura h del objeto se calcula de la siguiente forma considerando al punto P' el origen de coordenadas:

$$h = ||s - b|| \left[\frac{||C' - b||}{||C - C'||} + \frac{||s||}{||P - P'||} \right]^{-1} \quad (1)$$

donde $||.||$ es la función módulo de un vector. Por lo tanto, $||s - b||$ representa el desplazamiento que sufre el punto localizado en s luego de haberse colocado el objeto. Denominando $||P - P'|| = ||C - C'|| = d$, la ecuación (2) muestra la forma en la que este desplazamiento sólo depende de la altura del objeto a medir, dado que C' y d están fijados por la elección de la geometría del experimento. La deducción completa de las ecuaciones puede encontrarse en [9] [10].

$$||s - b|| = \frac{||C'||h}{abs(d - h)} \quad (2)$$

En base a lo anterior y haciendo el paso inverso, si se conoce el valor de h es posible calcular los valores s y b conociendo la ubicación del proyector y la cámara. Luego, la iluminación estructurada puede usarse como técnica de posicionamiento de objetos de dimensiones conocidas.

Cuando se obtiene la topografía de un objeto mediante iluminación estructurada, el patrón de franjas se deforma, y así se obtiene la información tridimensional. Luego existe una sola proyección de franjas ya deformadas, que cuando iluminen el objeto para el cual han sido deformadas (o formadas), las franjas se vean perfectamente verticales, paralelas y equidistantes en una imagen. A partir de un objeto cuyas dimensiones espaciales son conocidas, se puede hallar una imagen de franjas específicamente moldeadas para que cuando dicho objeto se posicione en un lugar elegido de antemano, sobre el plano de referencia, todas las franjas que se observen con la cámara fotográfica se ajusten a un patrón como el de Ronchi, por ejemplo. Podría decirse que es la operación inversa para la medición de un objeto mediante proyección de franjas. Con esta inversa, un objeto ya conocido es quien dicta la forma que deben tener las franjas para que al ser

proyectadas sobre él, se obtenga un patrón de franjas dado. Por lo tanto, hay un solo lugar sobre el plano de referencia donde puede situarse el objeto para que las franjas observadas sean un patrón específico. Además, si el objeto conocido sufre cualquier suerte de cambio en su forma, será evidente que ya no pueda volver a observarse el patrón cuando se le proyecten las franjas modificadas para posicionamiento.

3. Resultados obtenidos

Para el desarrollo de este trabajo se utilizó un Proyector SONY VPL-ES5 con sistema de proyección de 3 paneles LCD y 1 objetivo con zoom de 1,2 aumentos. Un panel SVGA de 800 x 600 pixels (0,48 Megapixels) y un panel por cada componente del RGB. La cámara fotográfica utilizada fue una Canon PowerShot A590 IS con una resolución de 3264 x 2448 pixels (8 Mp), zoom óptico de 4 aumentos y conexión mini USB sólo para trasmisión de datos.

El tipo de patrón usado fue una grilla de Ronchi de 600 x 800 pixeles pudiendo seleccionarse la cantidad, el ancho y el color de cada franja así como la separación entre ellas. El color fue representado en RGB y resultó de suma utilidad para determinar la correspondencia entre la posición del objeto y el patrón generado.

Para medir el desempeño de la iluminación estructurada en el relevamiento de topografías simples se utilizaron los objetos de la figura 6, y son: i) Un semicilindro de 17,5 cm de alto y radio 5 cm, simula el cuello de un paciente. ii) Una semiesfera de 7 cm de radio y una esfera de 10.7 cm de radio, simulan la cabeza de un paciente.

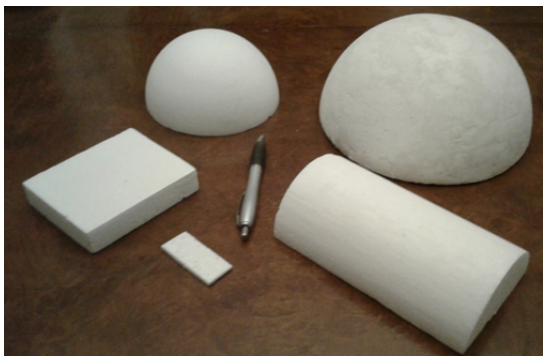


Figura 6. Figuras con dimensiones equivalentes a la cabeza y cuello de pacientes sobre las que se realizó un relevamiento topográfico utilizando iluminación estructurada.

En las figuras 7 y 8 pueden verse los resultados del relevamiento topográfico del semicilindro y las semiesferas realizados con el software desarrollado. Se pudo comprobar que la diferencia entre las dimensiones medidas del semicilindro hechas con iluminación estructurada y las dimensiones reales fueron inferiores

a 1 mm para todos los pixels. Para las semiesferas, estas discrepancias fueron todas inferiores a 2 mm.

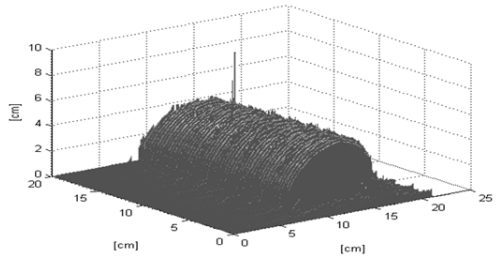


Figura 7. Reconstrucción del semicilindro utilizando el resultado del relevamiento topográfico realizado con el software desarrollado.

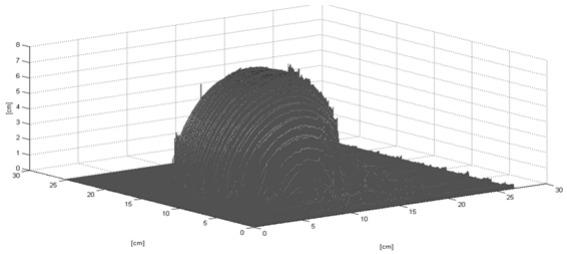


Figura 8. Reconstrucción de una de las semi esfereas a partir del relevamiento topográfico correspondiente.

La figura 9 muestra una sustracción de imágenes entre una semi esfera generada computacionalmente y la reconstrucción realizada utilizando el relevamiento topográfico. En ella puede observarse que las mayores discrepancias se encuentran en los bordes (zonas claras) mientras que en la zona central es la que presenta la mejor correspondencia.

Este relevamiento topográfico resulta sumamente útil para determinar la ubicación correcta en la debe posicionarse el paciente. Es un enfoque diferente al utilizado habitualmente cuando se trabaja con iluminación estructurada ya que en este artículo, el énfasis no está puesto en determinar la topografía de un objeto sino su ubicación según un patrón dado.

En el caso particular de la radioterapia, el objeto cuya topografía interesa analizar es la cabeza o el cuello del paciente. La posibilidad de calcular esta ubicación reiteradas veces, es fundamental en estos casos, dado que se trata de procedimientos de varias sesiones a través del cual, la zona irradiada modifica

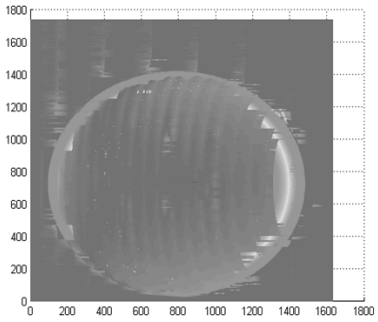


Figura 9. Sustracción entre una semi esfera generada computacionalmente y el resultado obtenido de la reconstrucción basada en el relevamiento topográfico. La zona central (más oscura) es la que posee mayor precisión.

su forma sustancialmente. Actualmente este proceso requiere de una tomografía de rayos X, la cual afecta el cuerpo del paciente al irradiarlo innecesariamente.

Utilizando nuevamente una de las semi esfera para simular la cabeza del paciente, se proyecta sobre ella el patrón de franjas de la figura 10.a) obteniendo como resultado lo que se obseva en al figura 10.b). Luego, este patrón deformado será utilizado como el patrón de referencia, para iluminar al objeto que se desea posicionar. Sólo cuando dicho olbjeto se encuentre en el lugar esperado, el patrón volverá a verse como en la figura 10.a)

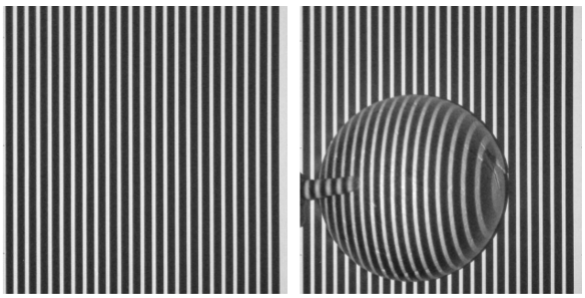


Figura 10. a) Ilustración del patrón de referencia; b) Objeto ubicado arbitrariamente en un lugar diferente al esperado.

La secuencia de imágenes que se muestra en la figura 11 ilustra el posicionamiento en sí, y la forma en la que gradualmente las franjas ubicadas sobre el objeto toman la forma del patrón generado como se indicó para la figura 10.b). A medida que la semiesfera se acerca al lugar correcto, las franjas que se proyectan sobre ella van tomando la forma vertical esperada. La figura 11.d) muestra

la mejor alineación conseguida. En dicha imagen, la posición de la semiesfera se encuentra a menos de 4 mm de la posición correcta, entendiendo por tal la que la alinea con el patrón. Puede verse que el mejor posicionamiento al que

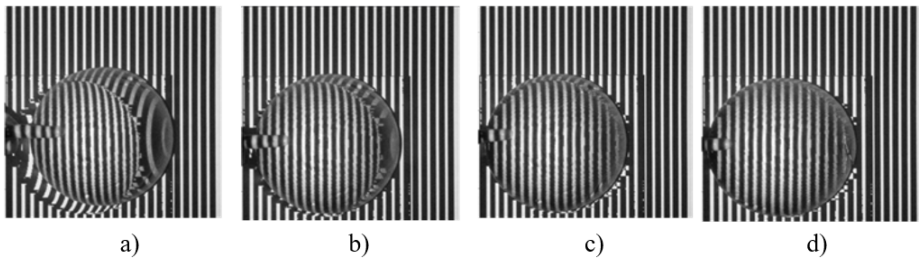


Figura 11. Secuencia de posicionamiento de la semi esfera utilizando el patrón generado a partir del relevamiento topográfico

se llegó en esta prueba (Figura 11.d)) exhibe un buen acuerdo con lo esperado, es decir, llegar a ver sólo el patrón constituido por franjas iguales, paralelas y equidistantes.

4. Conclusiones

Se ha desarrollado un software que permite relevar la topografía de objetos utilizando la iluminación estructurada llegando a alcanzar una exactitud del orden de los 2 mm. La iluminación estructurada es una técnica que permite obtener información espacial tridimensional a partir de información bidimensional, el conocimiento de la geometría del experimento y de las características de la proyección y la adquisición de las imágenes.

Este artículo propone un nuevo uso de la iluminación estructurada ya que utiliza las dimensiones del objeto, independientemente de la manera en que hayan sido obtenidas, para posicionarlo en una ubicación predeterminada. Esto se realiza a través de un nuevo patrón de franjas generado a partir de la información topográfica que al ser proyectado sobre el objeto estará conformado por franjas verticales cuando el mismo se encuentre en la posición correcta. Es importante considerar que se trata de un proceso que puede verificarse de manera inmediata con la sola inspección ocular por parte del operador. Independientemente de esto, se ha comprobado que el posicionamiento más adecuado sólo presenta una diferencia máxima de 4 mm siendo éste un valor aceptable según el criterio médico [11].

Con respecto a la potencial aplicación clínica de la técnica en radioterapia externa se propone utilizar la tomografía computada para planificación del paciente en posición de tratamiento y generar con ella la imagen de posicionamiento para ser proyectada en el bunker de tratamiento. Así la iluminación estructurada constituiría una herramienta poderosa para garantizar la reproducibilidad y

repetitividad de la posición del tratamiento. Nótese que se trata de un procedimiento no invasivo que puede reproducirse a lo largo del tratamiento

Especial interés podría tener su uso como sustituto de la máscara inmovilizadora ya que la misma tiene un costo elevado y en a lo largo del tratamiento no se ajusta adecuadamente a los cambios en la fisonomía del paciente.

Referencias

1. E.B. Podgorsak. *Radiation Oncology Physics*. International Atomic Energy Agency, 2005.
2. F.M. Khan and J.P. Gibbons. *Khan's The Physics of Radiation Therapy*. Wolters Kluwer Health, 2014.
3. M. Tubiana. *Radiobiologia: Radioterapia y Radioproteccion. Bases Fundamentales*. Hermann, medecina, 2012.
4. S. Webb. *The Physics of Three Dimensional Radiation Therapy - Reprinted with corrections*. Series in Medical Physics and Biomedical Engineering. IOP Publishing, 2001.
5. ICRU. Prescripción, registro y elaboración de informes en la terapia con haces de fotones. reporte técnico 62. Technical report, ICRU, 1999.
6. Sai Siva Gorthi and Pramod Rastogi. Fringe projection techniques: Whither we are? *Optics and Lasers in Engineering*, 48(2):133 – 140, 2010. Fringe Projection Techniques.
7. P. Mayles, A. Nahum, and J.C. Rosenwald. *Handbook of Radiotherapy Physics: Theory and Practice*. CRC Press, 2007.
8. R. Windecker and H. J. Tiziani. Topometry of technical and biological objects by fringe projection. *Appl. Opt.*, 34(19):3644–3650, Jul 1995.
9. M. Wördemann. *Structured Light Fields: Applications in Optical Trapping, Manipulation, and Organisation*. Springer Theses. Springer Berlin Heidelberg, 2012.
10. D.L. Andrews. *Structured Light and Its Applications: An Introduction to Phase-Structured Beams and Nanoscale Optical Forces*. Elsevier Science, 2011.
11. S. Lahrmann, O. Jäkel, and C. P. Karger. Investigation of acceptance criteria for the gamma-index in scanned carbon ion radiotherapy. In Olaf Dössel and Wolfgang C. Schlegel, editors, *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, pages 231–233, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

Control Activo de Ruido Basado en DSP para un Soplante Industrial

Guillermo Friedrich, Guillermo Reggiani, Pablo Girón y Adrián Azzurro,

Universidad Tecnológica Nacional Facultad Regional Bahía Blanca,
Bahía Blanca, Argentina
{gfried, ghreggiani, pgiron, azzurro}@frbb.utn.edu.ar

Abstract. Los soplantes rotativos por lo general emiten elevados niveles de ruido, de frecuencias discretas relacionadas con su velocidad de rotación. Los elevados niveles de sonido pueden causar daños más o menos graves al oído, dependiendo de su intensidad y del tiempo de exposición. En ciertos casos es obligatorio usar protección auditiva y/o reducir el tiempo de exposición. En otros casos, el ruido afecta al confort de los trabajadores y reduce su productividad. Por lo tanto se hace necesario utilizar elementos de protección auditiva y, de manera complementaria, aplicar técnicas para atenuar el ruido. Si las frecuencias de ruido se encuentran en el extremo inferior del espectro pueden ser difíciles de atenuar mediante técnicas pasivas. En el presente trabajo se presenta el desarrollo de un sistema de control de ruido activo cuyo objetivo es atenuar significativamente el ruido emitido por un soplante de alta potencia de una planta petroquímica. Esta solución se ha implementado sobre un procesador de señales digitales y se basa en un esquema de tipo realimentado (feedback ANC).

Keywords: control activo de ruido, ruido de soplante, feedback ANC.

1 Introducción

El ruido en los ambientes industriales es uno de los factores de riesgo más comunes para la salud de los trabajadores. En muchos casos es técnicamente factible controlar el impacto de su nivel excesivo utilizando elementos de protección, pero también es posible aplicar tecnología a las máquinas ruidosas. Los altos niveles de ruido pueden causar daños más o menos graves en el oído, dependiendo de su intensidad y del tiempo de exposición. En algunos casos es obligatorio usar protección auditiva y además reducir el tiempo de exposición. Otras veces el ruido afecta al confort de los trabajadores y reduce su productividad. Por lo tanto se hace necesario utilizar elementos de protección auditiva y, de manera complementaria, aplicar técnicas para atenuar el ruido.

Por lo general, los soplantes rotativos de uso industrial emiten elevados niveles de ruido de frecuencias discretas, relacionadas con su velocidad de rotación. La Fig. 1 muestra un soplante rotativo típico montado en su gabinete.

Si las frecuencias de ruido se encuentran en el extremo inferior del espectro son difíciles de atenuar mediante técnicas pasivas, lo que motiva la aplicación de técnicas de control activo de ruido (ANC).



Fig. 1. Un soplante típico montado en su gabinete.

El soplante tomado como base para el presente trabajo es uno de tres lóbulos y 110 kW de potencia; y produce una intensidad de ruido de 105 dB (A) frente a la puerta del gabinete. Este ruido tiene un componente tonal principal de 240 Hz y varios armónicos de amplitudes significativas de hasta 960 Hz y también 1200 Hz. La Fig. 2 muestra un diagrama espectral obtenido por medio de un micrófono de medición Presonus PRM1 [1], un dispositivo AudioBox USB [2] y el software StudioOne [3].

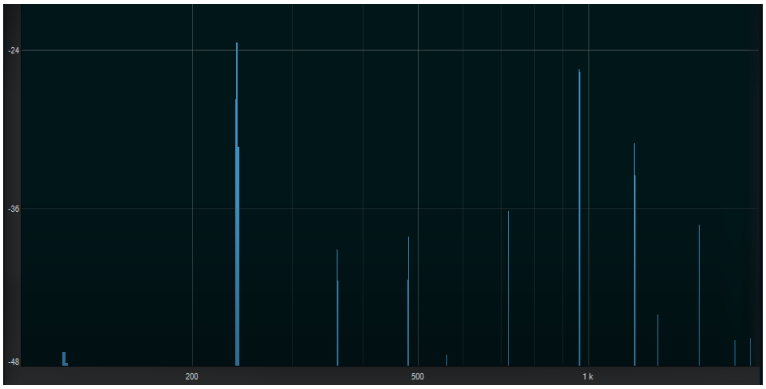


Fig. 2. Diagrama espectral del ruido producido por el soplante, medido frente a la puerta del gabinete (Obs.: la escala izquierda está indicada en dB pero no está relacionada con ninguna magnitud acústica).

Frente al soplante, a unos 30 m de distancia, hay un edificio de oficinas a través de cuyas ventanas ingresa el ruido que proviene de dicho equipo. Debido a que las componentes de frecuencias más altas, 960 y 1200 Hz, podrían atenuarse mediante

técnicas pasivas, y que también se atenúan más por la distancia, se propuso aplicar técnicas de ANC para reducir la componente de 240 Hz.

Aunque las bases teóricas de ANC ya se encuentran maduras, actualmente hay una producción importante en términos de su aplicación a diferentes entornos y problemas [4]. Cabe mencionar, entre otras, las aplicaciones de ANC para reducir el ruido de los transformadores de potencia en entornos urbanos [5]; atenuación del ruido en ventanas de viviendas [6]; atenuación de ruido en conductos de aire acondicionado [7]; control activo del ruido en cabinas de vehículos [8]; silenciadores activos / pasivos [9]; sistemas de ruido activo para reducir el ruido exterior [10].

La solución fue desarrollada sobre un kit ADSP-21061 EZ-KIT Lite [11] de Analog Devices, que está basado en el procesador ADSP-21061 [12] funcionando a 40 MHz, con un códec AD1847 de 16 bits estéreo [13], configurado para una frecuencia de muestreo de 8 kHz. El código de programa fue desarrollado en C.

Los ensayos preliminares de laboratorio se llevaron a cabo usando un ducto típico para este tipo de ensayos [14], con el fin de ajustar detalles y parámetros de los algoritmos. Luego se llevaron a cabo otros ensayos en espacio abierto, emulando las condiciones de operación del soplante junto con la solución activa. La Fig. 3 muestra ambos parlantes, uno utilizado para reproducir el ruido grabado del soplante, y el otro utilizado para emitir el antirruido generado por el sistema ANC. Además, la misma figura muestra la cápsula del micrófono de error y el micrófono utilizado para hacer las mediciones, ubicados en frente de los parlantes.



Fig. 3. Parlantes y micrófonos utilizados en los ensayos.

El esquema ANC adoptado fue el de retroalimentación (feedback ANC) porque, al usar un solo micrófono, es más simple y también evita la realimentación acústica que aparece en el esquema denominado feed-forward. Sin embargo, se utilizaron con fines comparativos los resultados presentados en [7], obtenidos en base a una configuración de feed-forward.

2 Sistema de Control Activo de Ruido

2.1 Tipos de Sistemas de Control Activo de Ruido

Los sistemas de control activo de ruido se basan en dos métodos: feed-forward y feedback. En el método feed-forward se sensa el ruido en la fuente, y se lo toma como referencia. En el método feedback el controlador intenta cancelar el ruido sin tener una referencia del ruido de entrada, por lo que debe estimarla (Fig. 4). Se clasifican además en dos categorías: control adaptativo feed-forward de banda ancha, que utiliza un sensor de entrada acústico (figura 4) y control adaptivo feed-forward de banda estrecha, que utiliza un sensor de entrada no acústico (figura 5).

Figs. 4, 5 y 6 muestran los tres esquemas de ANC sobre un ducto, que es la típica estructura utilizada para ensayar los sistemas y técnicas de ANC.

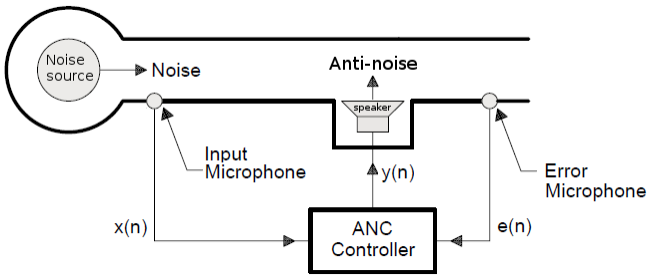


Fig. 4. Sistema ANC feed-forward de banda ancha

Un sistema ANC feed-forward de banda ancha tiene la desventaja de presentar realimentación acústica, que debe ser cancelada. El anti-ruido emitido por el parlante no solo cancela el ruido acústico que va hacia el final de la tubería, sino que también irradia hacia el micrófono de entrada, contaminando la entrada de referencia *x(n)*. Esta realimentación acústica es una fuente potencial de inestabilidad.

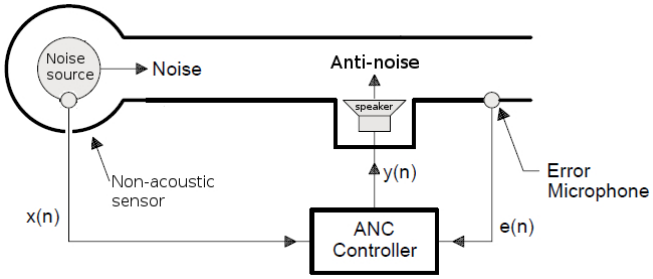


Fig. 5. Sistema ANC feed-forward de banda estrecha

Cuando el ruido primario es periódico (o casi periódico), como es el caso del ruido producido por los soplantes rotativos, el micrófono de entrada puede reemplazarse por algún tipo de transductor no acústico (por ej.: un tacómetro, un acelerómetro o algún sensor óptico), que no se ve afectado por el fenómeno de la realimentación acústica.

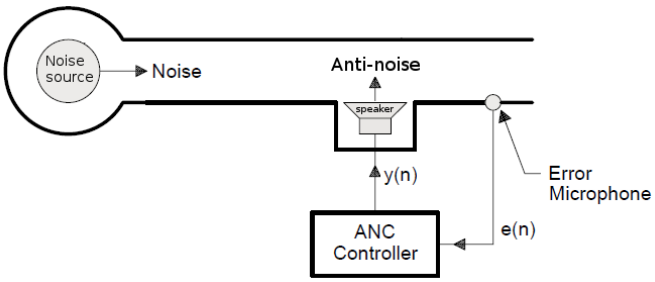


Fig. 6. Sistema ANC feedback.

En el esquema ANC feedback, se usa un micrófono como sensor de error para detectar el ruido no deseado. A partir de esta señal de error se extrae una estimación de la señal de ruido original, que luego es filtrada y amplificada para producir la señal de anti-ruido, con la magnitud y fase necesarias para cancelar el ruido.

Esta configuración proporciona una menor atenuación que la feed-forward, y podría sufrir inestabilidad debido al lazo de realimentación. Sin embargo, en base a este esquema es posible obtener un sistema robusto para cancelar ruido de banda estrecha.

En el caso de la máquina que es objeto de este trabajo, el esquema feedback tiene la ventaja de no ser invasivo, ya que no es necesario montar un sensor en dicha máquina para obtener la señal de ruido.

En caso de obtenerse resultados favorables, será posible justificar la necesidad de trabajar sobre la máquina para mejorar el sistema, convirtiéndolo en feed-forward de banda estrecha. Debe tenerse en cuenta que este tipo de máquina es parte de un proceso continuo de muy alto costo, cuya operación y mantenimiento se controla mediante procedimientos muy rigurosos y estandarizados. Cualquier modificación experimental que uno desee hacer debe estar muy bien justificada.

2.2 Sistema ANC feedback

Un sistema ANC feedback, como el que se muestra en la Fig. 7, se basa en un predictor adaptivo. Debido a que este sistema requiere solo un micrófono de error, evita el problema de realimentación acústica inherente al método feed-forward.

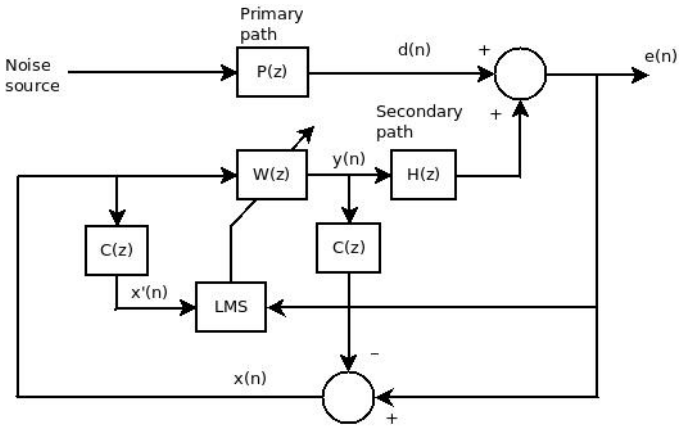


Fig. 7. Diagrama en bloque de un sistema ANC feedback

La idea básica de este algoritmo es estimar el ruido primario *d(n)* y utilizar este valor como la entrada de referencia para el filtro adaptativo. *H(z)* representa la función transferencia del camino secundario, mientras que *C(z)* es una estimación de *H(z)* obtenida por medio de un proceso de identificación fuera de línea FxLMS. El ruido primario se estima como:

$$x(n)=e(n)-\sum_{i=0}^{M-1}\left(e(i)-c_i\,y(n-i)\right) \tag{1}$$

donde c_i es el coeficiente i-th de *C(z)*, el filtro de estimación del camino secundario, y M el orden del filtro. De la Fig. 7 se tiene:

$$D(z) = E(z) - H(z)\,Y(z) \tag{2}$$

Si *C(z)* es un modelo de *H(z)*, entonces:

$$D(z) \cong X(z) = E(z) - C(z)\,Y(z) \tag{3}$$

Y la señal de error puede ser calculada así:

$$E(z) = D(z) + W(z)\,H(z)\,X(z) \tag{4}$$

La señal de error es cero cuando:

$$-W(z)\,H(z)\,X(z) = D(z) \tag{5}$$

Donde *W(z) H(z)* es igual a un retardo equivalente a un múltiplo del período de señal, y el signo menos indica el desplazamiento de fase de 180° que es necesario aplicar para cancelar *D (z).*

Los coeficientes de *W (z)* se ajustan en línea por medio de un algoritmo adaptativo, tal como LMS, basado en el valor de la señal de error, *e(n)*.

El valor anti-ruido se calcula a partir del valor de *x (n)* obtenido en (1):

$$y(n)=\sum_{i=0}^{N-1} w_i x(n-i) \tag{6}$$

donde N es el orden del filtro.

También se computa $x_i(n)$, la versión filtrada X de *x(n)*:

$$x'(n)=\sum_{i=0}^{M-1} c_i x(n-i) \tag{7}$$

A continuación, *x'(n)* es utilizado por el algoritmo LMS para actualizar los coeficientes w_i para la próxima iteración:

$$w_i(n+1)=w_i(n)-\beta e(n)x'(n-i), \qquad i=0,1,\ldots,N-1 \tag{8}$$

Donde β es el tamaño del paso del algoritmo LMS, del cual depende la velocidad de convergencia y la estabilidad del sistema.

Existe otra versión del algoritmo FXLMS, llamada Leaky FXLMS, que introduce un coeficiente de perdida que evita el desbordamiento numérico que podría aparecer en las arquitecturas de punto fijo, debido a su precisión limitada. Aunque este proyecto se basa en un punto flotante de 32 bits DSP, Leaky FXLMS mejora la estabilidad del sistema a expensas de una menor reducción en la cancelación de ruido. La actualización de los coeficientes utilizando Leaky FXLMS se calcula de la siguiente manera:

$$w_i(n+1)=\lambda w_i(n)-\beta e(n)x'(n-i) \tag{9}$$

Donde λ es el coeficiente de pérdida, cuyo valor es ligeramente menor que 1.

A continuación, se presenta la sección del código C que implementa el ajuste adaptativo de los coeficientes de filtro *W (z)* a través de FXLMS:

```
eb = error * beta;

for(i=N-1; i>=0; --i)
{
  w[i] = leaky * w[i] - eb * xw2[i];
  if(i)
    xw2[i] = xw2[i-1];
}
```

Donde *error* contiene el valor de *e(n)*, *beta* el valor de *β*, λ el coeficiente de pérdida (leaky), y $xw_2[]$ es una matriz que contiene los valores de *x'(n-i)*.

2.3 Identificación adaptiva del camino secundario

Como ocurre en aplicaciones prácticas, *H(z)* es desconocida y debe ser estimada por el filtro *C(z)*. Esta estimación se puede hacer por medio de una técnica de adaptación fuera de línea o en línea.

Suponiendo que las características de *H(z)* son desconocidas pero invariantes en el tiempo, se puede usar una técnica de modelado fuera de línea para estimar *H(z)* durante una etapa de identificación. Después de esta identificación, el modelo estimado *C(z)* se fija y se utiliza para el control del ruido activo. La Fig. 8 muestra el diagrama de bloques de la configuración experimental para la identificación directa del camino secundario fuera de línea. El ruido blanco no correlacionado es generado internamente por el DSP y enviado simultáneamente al camino secundario a través del parlante y al filtro *C(z)*. La diferencia entre *e(n)* y *r(n)*, llamada *e'(n)*, que debería tender a cero durante la etapa de identificación, se utiliza para alimentar el algoritmo LMS, que ajusta los coeficientes del filtro *C(z)*.

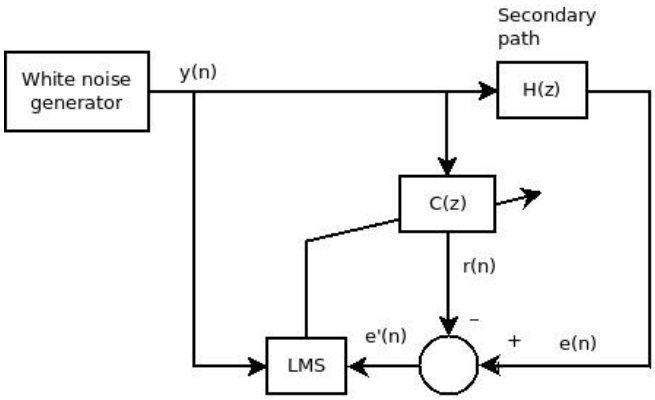


Fig. 8. Diagrama en bloque del proceso de identificación del camino secundario.

La respuesta del modelo adaptativo se calcula mediante:

$$r(n)=\sum_{i=0}^{M-1} c_i\, y(n-i) \tag{10}$$

Entonces, usando el valor *e'(n) = e(n) - r(n),* los valores c_i son actualizados adaptivamente mediante el algoritmo LMS de la siguiente manera:

$$c_i(n+1)=c_i(n)+\mu\, e'(n)\, y(n), \qquad i=0,1,\ldots,M-1 \tag{11}$$

Donde μ es el tamaño del paso, cuyo valor debe ajustarse entre 0 y $(MP_y)^{-1}$, siendo P_y la potencia del ruido blanco.

La estrategia utilizada para encontrar el mejor ajuste de los coeficientes c_i en función de los valores de M y μ consistió en enviar las muestras del error residual *e'(n)* al otro canal de salida del códec, y leer el valor del voltaje con un voltímetro de

CA. El objetivo es obtener la lectura más baja en el instrumento. De esta forma, se determinó que los valores de $M = 384$ y $\mu = 5.10^{-3}$ fueron los que proporcionaron el mejor ajuste, tanto para las pruebas en la tubería como en el espacio abierto.

Luego, con los valores apropiados de M y μ, podría determinarse que para fines prácticos, solo es necesario ejecutar el proceso de identificación hasta que el valor RMS de *e(n)* alcance el 20% del valor inicial (para fines prácticos y según a las observaciones hechas durante las pruebas, sería suficiente mantener el proceso de identificación durante aproximadamente diez segundos).

2.4 Configuración Experimental

En la Fig. 9 se muestra el diagrama en bloques del ensayo implementado para realizar las pruebas.

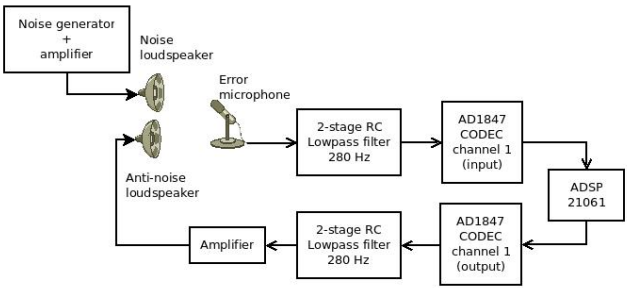


Fig. 9. Diagrama en bloque de la configuración usada para los ensayos.

El filtro pasa-bajo RC de entrada ha sido insertado para atenuar las frecuencias de orden superior que no son de interés para el sistema ANC. El filtro pasa-bajo RC de salida tiene el objetivo de eliminar el ruido de alta frecuencia (ripple) causado por el procesamiento adaptativo, y que corrompe la señal anti-ruido. Ambos filtros mejoran la estabilidad y el rendimiento del sistema.

3 Ensayos y Resultados

En las primeras etapas se realizaron varios ensayos sobre un ducto instalado en el laboratorio, a fin de verificar y ajustar el código de programa, los algoritmos, el hardware y otros detalles. Luego, en una sala multiusos sin acondicionamiento acústico, se montó la configuración para realizar ensayos en espacio abierto (Fig. 3).

Se utilizaron dos señales de ruido para las pruebas: un tono puro de 240 Hz y una grabación del sonido emitido por soplante, cuyo diagrama espectral se muestra en la Fig. 2.

Las mediciones de ruido se realizaron utilizando dos dispositivos / métodos: (i) software Studio One alimentado por un micrófono de medición Presonus PRM1 a través de un USB AudioBox; (ii) Sonómetro Cesva. La atenuación del tono de 240 Hz se midió con el primer método, mientras que la atenuación del nivel de sonido global se midió con el sonómetro.

La Tabla I muestra los resultados obtenidos para un tono puro de 240 Hz. Se puede observar que cuando el micrófono de error estuvo ubicado más cerca de los parlantes, obtienendo mayor intensidad de sonido, se logró una mayor atenuación. Esto se debe a que la relación señal/ruido es 1,5 dB mejor y permite un control adaptativo más preciso.

También se obtuvo una mayor reducción en la escala global con respecto al tono único de 240 Hz, 23,8 dB frente a 19,4 dB, debido a la curva de compensación de dB (A) utilizada por el sonómetro en dicha escala global.

En la Tabla II se presentan los resultados obtenidos usando una grabación del ruido emitido por el soplante, cuyo diagrama espectral se muestra en la Fig. 2. Se puede observar que la atenuación del tono de 240 Hz es aproximadamente 3 dB menor que para el caso del ruido monotonal (Tabla I). La presencia de otras componentes tonales significa una menor relación señal a ruido con respecto a la componente de interés, lo que degrada la calidad de la estimación de la señal a cancelar. Asimismo se observa que el ruido global casi no tiene reducción, porque el resto de las componentes tonales tiene una gran influencia en el valor de la suma ponderada con la que se calcula el valor expresado en dB (A).

Tabla I. Atenuación para un tono puro de 240 Hz

Atenuación para un tono puro de 240 Hz LMS: coeficiente de pérdidas (leaky) = 0.99995 β = 10-3		
Distancia entre parlantes y micrófono de error	40 cm	30 cm
Intensidad inicial del ruido (medida junto al micrófono de error microphone con un sonómetro Cesva usando la escala global)	82,6 dB(A)	84,3 dB(A)
Atenuación medida con el micrófono Presonus + AudioBox + software Studio One	15,2 dB	19,4 dB
Atenuación medida con sonómetro Cesva en escala global (dB(A))	18,1 dB	23,8 dB

Tabla II. Atenuación de la componente de 240 Hz del ruido del soplante

Atenuación de la componente de 240 Hz del ruido del soplante LMS: coeficiente de pérdidas (leaky) = 0.99995 β = 10-3	
Distancia entre parlantes y micrófono de error	30 cm
Intensidad inicial del ruido (medida junto al micrófono de error microphone con un sonómetro Cesva usando la escala global)	85,7 dB(A)
Atenuación medida con el micrófono Presonus + AudioBox + software Studio One	16,2 dB
Atenuación medida con sonómetro Cesva en escala global (dB(A))	0,2 dB
Atenuación medida con sonómetro Cesva en el modo espectral (división de 250 Hz)	16 dB

A fin de comparar los resultados obtenidos con los de otro sistema, en la Tabla III se muestran los resultados obtenidos con un esquema FXLMS feed-forward con coeficiente de pérdida, presentado en [7].

A partir de la comparación entre los valores de las Tablas I y II, es posible concluir que la presencia de una combinación de tonos reduce el rendimiento del sistema ANC: 19,4 dB frente a 16,2 dB. Esto también se puede ver en la Tabla III para un esquema de feed-forward: 28,3 dB para un solo tono de 200 Hz frente a 19,8 dB para tonos múltiples. Esto se debe a que, considerando solo la componente de 240 Hz, la relación señal a ruido es peor en el segundo caso.

Tabla III. Atenuación usando un esquema FXLMS feed-forward con pérdida

Atenuación reportada por Biaggini et al [7] usando ANC de tipo feed-forward FXLMS con pérdida, para reducir ruido en el exterior		
Características de la señal	Tono puro de 200 Hz	Múltiples tonos 200-300 Hz (every 20 Hz)
Intensidad de sonido original	75 dB	70 dB
Atenuación	28,3 dB	19,8 dB

El orden de atenuación logrado mejora las condiciones de salud y confort en el área cercana al soplante, lo que permitiría reducir la protección auditiva y/o aumentar el tiempo de exposición. De todos modos, estas modificaciones deben ser analizadas y autorizadas por los especialistas en seguridad e higiene industrial.

4 Conclusiones

El ruido producido por los equipos industriales es un grave problema que afecta la salud y el confort de trabajadores y vecinos. Por esta razón es interesante encontrar soluciones innovadoras, como es el caso del uso de ANC para reducir los ruidos de baja frecuencia generados por este tipo de máquinas.

En este trabajo se presentó el desarrollo de un sistema ANC de tipo realimentado (feedback), diseñado para su aplicación en la reducción del ruido emitido por un soplante de alta potencia de una planta petroquímica, como así también se presentaron los ensayos realizados al mismo.

El sistema fue desarrollado sobre un procesador de señales digitales ADSP-21061 de Analog Devices operando a 40 MHz, usando un códec AD-1847 de dos canales de 16 bits, configurado a una frecuencia de muestreo de 8 kHz.

Se realizaron ensayos en laboratorio, primero sobre un típico ducto y luego en espacio abierto, tratando de emular las condiciones en que el sistema debería operar junto al soplante.

Los ensayos dieron resultados satisfactorios, lográndose más de 19 dB de atenuación para un ruido tonal de 240 Hz (y más de 23 dB cuando se efectuaron mediciones en la escala de dB (A)). La comparación con soluciones basadas en un esquema de tipo feed-forward permite suponer apriori que es posible intentar mejorar el sistema actual, ajustando ciertos parámetros del mismo, tales como: orden de los filtros, β, μ, coeficiente de pérdida y frecuencia de muestreo.

Además, a partir de la mencionada comparación se puede resultar de interés evaluar una solución tipo feed-forward debido a su mayor rendimiento, aunque tiene algo más de complejidad.

Sin embargo, debido a su potencial mayor rendimiento también se considerará la posibilidad de implementar y emsayar una solución de tipo feed-forward, utilizando un micrófono o bien un sensor no acústico montado sobre la máquina.

Referencias

1. "PreSonus PRM1: Precision Reference Microphone". https://www.presonus.com/productos/es/PRM1/especificaciones [Jul. 04, 2018].
2. "AudioBox USB: Sistema de grabación USB 2x2". https://www.presonus.com/productos/es/AudioBox-USB/especificaciones [Jul. 04, 2018].
3. "StudioOne: software para masterización de sonido". https://www.presonus.com/productos/es/Studio-One/especificaciones [Jul. 04, 2018].
4. S. M. Kuo and D. R. Morgan, "Active noise control: a tutorial review", Proc. IEEE, vol. 87, no. 6, pp. 943-973, Jun. 1999.
5. J. Liang, T. Zhao, L. Zou, L. Zang and Z. Li, "Adaptive active noise control system of a power transformer", 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communications and Control (IMCCC), 18-20 September 2015, Qinhuangdao, China, pp. 1394-1397.
6. H. Huang, X. Qiu and J. Kang, "Active noise attenuation in ventilation windows", J. Acoust. Soc. Am. 2011, 130(1), pp. 1-13.
7. I. Doré Landau, R. Meléndez, L. Dugard and G. Buche "Robust and adaptive feedback noise attenuation in ducts", IEEE Transactions on Control Systems Technology (early access), December 2017, pp 1-8.
8. J. Pan, Y. Liu, "Active control of noise in vehicle cabins", Journal of the Acoustical Society of America, vol. 131, no. 4, pp. 3379–3385, 2012.
9. Health and Safety Executive (HSE), "Active/absorptive silencer for a rotary blower", http://www.hse.gov.uk/Noise/casestudies/soundsolutions/rotaryblower.htm, last access: may 10, 2018.
10. M. Biagini, F. Borchi, M. Carfagni, L. Fibucchi, A. Lapini and F. Argenti, "Active Noise Systems for Reducing Outdoor Noise", 24th International Congress on Sound and Vibration (ICSV24), London, July 2017.
11. "ADSP-21061 EZ-KIT Lite Evaluation System Manual". http://www.analog.com/media/en/dsp-documentation/legacy-evaluation-kit-manuals/39812179506535ADSP_21061_EZ_KIT_Lite_Man_Rev2.pdf [Jul. 04, 2018].
12. "ADSP-21061 datasheet". http://www.analog.com/media/en/technical-documentation/data-sheets/ADSP-21061_21061L.pdf [Jul. 04, 2018].
13. "AD1847 Codec datasheet". http://www.analog.com/media/en/technical-documentation/obsolete-data-sheets/1215246AD1847.pdf [Jul. 04, 2018].
14. G. Friedrich et al, "Señales vibro-acústicas y control activo de ruido", Proceedings of XV Congreso Argentino de Acústica AdAA 2017, Bahía Blanca, Argentina, 23-24 Nov. 2017. ISBN 978-987-24713-9-2. http://www.adaa.org.ar/adaa2017/AdAA2017-18.pdf, last access may 10, 2018.

Digitalización de Imágenes de ECG para la Detección del Síndrome de Bayés

Lorena G. Franco[1], Luis A. Escobar Robledo[2], Antoni Bayés de Luna[3], José María Massa[4]

[1]Universidad Nacional del Centro de Buenos Aires, Universidad Nacional Tecnológica, Arg, [2]Fac. Medicina, Universdidad CES, Colombia,[3]Fundació d'Investigació Cardiovascular, España,[4]INTIA, Fac. Cs. Exactas Universidad Nacional del Centro de Buenos Aires, Arg.
francol.edu.ar@gmail.com

Abstract. Bayes Syndrome is a recently recognized by medical community sickness. This desease has been studied in the last decades by its discoverer, MD. Antonio Bayés de Luna. Since several works shows that this desease is related to multiple symptoms, an early detection is considered relevant. Given that digital support of the EKG signal is mandatory for its analysis by a computer algorithm and considering that even with the technological advances, a big number of health institutions rely on paper or image digitalized support por EKGs. Hence, an image digitalization method that preserves the signal features that are relevant to diagnose the Bayes Syndrome is needed. In this paper, some alternatives of digitalization are analyzed for a representative dataset. Results are promising and shows that developed digitalization algorithm could be used for the further project stages that involves signal processing and classification.

Palabras Clave: Síndrome de Bayés, ECG, Bloqueo Interauricular Avanzado, Arritmias, procesamiento digital.

1. Introducción

El alcance y el recrudecimiento de las enfermedades cardiovasculares se han incrementado en los últimos años. La Organización Mundial de la Salud (OMS) indica que constituye una de las principales causas de muerte a escala mundial. Según el reciente informe de la OMS, las tasas de mortalidad global, debido a enfermedades cardiovasculares han disminuido. Sin embargo, el impacto de la enfermedad como causa de mortalidad sigue siendo alto [1]. Esto nos denota lo relevante de la prevención y el tratamiento adecuado.

Las señales electrocardiográficas son un elemento decisivo para el diagnóstico de enfermedades cardíacas. En la figura 1 se presenta un ECG normal y se señala la onda P de interés.

A partir del nacimiento y posterior desarrollo del diagnóstico automático de ECG, el mismo tiene un valor clínico muy significativo para el diagnóstico médico moderno. Hoy en día, muchas investigaciones se centran en la extracción de características y el reconocimiento de patrones de señales de ECG [2][3]. El análisis

de ECG basado en computadora requiere una detección precisa. Lograrlo representa un desafío, ya que una señal de ECG real generalmente contiene artefactos (es decir, ruido muscular, artefactos de movimiento y cambios de derivas en la línea base).

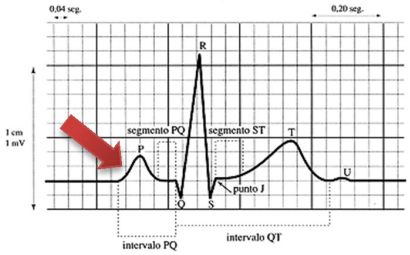


Fig. 1. ECG normal donde se indica la onda P

Dentro del marco de las enfermedades cardiovasculares, se puede mencionar que en los últimos años se ha demostrado la asociación del Síndrome de Bayés a múltiples afecciones médicas del sistema circulatorio. El concepto de BIA, bloqueo más frecuente y relevante a nivel auricular, significa que existe una conducción retrasada entre la aurícula derecha y la aurícula izquierda. Se dividió el BIA de la misma manera que a nivel ventricular, sinoauricular y auriculoventricular en primer grado o parcial, tercer grado o avanzado y segundo grado o intermitente [4] [5] [6].

Existe una fuerte evidencia de que el BIA avanzado (BIA-A), predice la fibrilación auricular (FA) y que cuanto mayor es la duración de la onda P, más fuerte es la asociación. La asociación de BIA-A con FA, ahora denominado Síndrome de Bayés, es más frecuente de lo que inicialmente se pensó. En personas muy mayores la prevalencia de BIA-A es del 25% [6]. En la figura 2 se puede observar la morfología de la onda P en un caso de BIA-A.

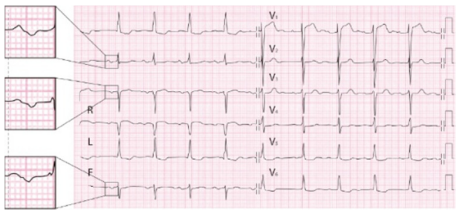


Fig. 2. ECG con BIA-A

Existen estudios que han informado la prevalencia de BIA superior al 40% en pacientes hospitalizados. A pesar de esto, el BIA permanece en gran medida subdiagnosticado. Este bloqueo presenta notables asociaciones con múltiples afecciones médicas que incluyen fibrilación auricular, isquemia miocárdica, agrandamiento de la aurícula izquierda y émbolos sistémicos [8]. En el artículo [9] se concluyó que el BIA debía ser considerado como un novedoso factor de riesgo para accidente cerebrovascular cardioembólico. En función de lo expuesto resulta de interés su reconocimiento en una etapa temprana. El diagnóstico de BIA parcial o

avanzado puede realizarse analizando el ECG. La Universidad de Glasgow que desarrolla programas de análisis de ECG [10], comercializados por algunas empresas, ha estado interesada en el Síndrome de Bayés. Sin embargo los resultados que han obtenido en la detección se encuentran en discusión.

Habitualmente los centros de salud y cardiología trabajan con estudios en papel o imagen. Sin aprovechar las ventajas de las nuevas tecnologías que permiten, como en el caso de los ECG, obtener los resultados en diferentes formatos. Ante la realidad mencionada y considerando que se dispone de imágenes de ECG que presentan las características de la onda P que define el Síndrome de Bayés, es necesario para avanzar en su detección obtener la señal digitalizada. Contemplando la variedad de opciones existentes en cuanto a herramientas de digitalización, en el contexto de esta publicación se optó por experimentar con un software disponible previo a la implementación de una herramienta propia. La herramienta desarrollada aplica para la segmentación el método de umbralización. El objetivo de la segmentación es simplificar y/o cambiar la representación de una imagen en otra más significativa y más fácil de analizar.

En este trabajo se presenta en la sección 2 aspectos teóricos relacionados al Síndrome de Bayés y al procesamiento de imágenes aplicables a las imágenes de ECG que se disponen. En la sección 3 se presentarán las alternativas abordadas a partir de los estudios disponibles y en la sección 4 los resultados obtenidos. Por último en la sección 5 se presentarán las conclusiones.

2. Procesamiento de imágenes con Síndrome de Bayés

Se presentarán en la sección 2.1 los aspectos relacionados al Síndrome de Bayés. A continuación se introducirán aspectos teóricos relacionados al procesamiento de imágenes. Los mismos resultan de interés para comprender los resultados que se observan en la sección 4.

2.1 Síndrome de Bayés

En 1979 eran escasos los estudios sobre alteraciones de la conducción a nivel auricular y tampoco existía una unificación de criterios con respecto a la terminología y sus definiciones [4]. Posteriormente se propuso dividir los trastornos de la conducción a nivel auricular en bloqueos intraauriculares e interauriculares (BIA). Se sugiere clasificar el BIA en:

Parcial: existe un retraso en la zona de conducción del Haz de Bachmann, pero la mayoría de la conducción de derecha a izquierda todavía se produce en el techo auricular.

Avanzado: existe un bloqueo total del Haz de Bachmann y la conducción hacia la aurícula izquierda se produciría a partir de la parte baja de la aurícula derecha con una dirección retrograda caudocraneal [4] [11].

A partir de estos descubrimientos se publican diversos artículos sobre BIA con diferentes definiciones [11]. Recientemente se propuso denominar a la asociación entre BIA-A y arritmias supraventriculares (fundamentalmente fibrilación auricular

[FA] y aleteo atípico) Síndrome de Bayés [12]. Esta denominación constituye un tributo a un reconocido especialista en Electrocardiología a nivel mundial [13] [14].

2.1.1 Bloqueo Interauricular

Se clasifica los BIA en: de primer grado (parcial), de segundo grado (bloqueo transitorio interauricular) o de tercer grado (avanzado).

El bloqueo de primer grado se define con una onda P≥120 mseg, generalmente bimodal. Se evidencia en las derivaciones I, II o III. La derivación V1 puede presentar una morfología negativa en su fase final.

El bloqueo de segundo grado se caracteriza por cambiar la morfología de la onda P de normal a patrones de BIA de primer grado que se transforma en uno de tercer grado y viceversa.

El bloqueo de tercer grado se define con una onda P≥120 mseg y una morfología de onda P bifásica o ± en las derivaciones II, III y AVF [15].

2.2 Procesamiento de imágenes

El procesamiento digital de imágenes incluye un conjunto de técnicas que permiten la representación digital de una imagen. El mismo constituye una herramienta útil para el diagnóstico, tratamiento y monitoreo de diversas enfermedades. A través de su aplicación se pretende facilitar el posterior análisis.

El análisis computacional de imágenes involucra aislar el objeto o región de interés del resto de la imagen (proceso de segmentación), determinar características como pueden ser la morfología o el tamaño y por último la clasificación o el reconocimiento de patrones.

A continuación se describirá el proceso de segmentación por medio de un umbral.

2.2.1 Umbralización

La umbralización de imágenes es uno de los métodos que abarca la segmentación de imágenes digitales. Se trata de una técnica de segmentación simple y eficiente que permite separar los pixeles de una imagen en escala de grises en dos categorías a partir de un valor umbral de intensidad. Éstas técnicas buscan obtener un valor de umbral que permita binarizar a la imagen separando adecuadamente el fondo (background) y el objeto a separar (foreground) [16]. Existen dos tipos de umbrales, el umbral global que es único sobre toda la imagen y el umbral local o adaptativo que consiste en dividir la imagen en subimágenes para encontrar un umbral apropiado para cada una de ellas [17]. Dichas subimágenes pueden ser de diferentes formas y tamaños, lo que se define en el momento de implementar el algoritmo para la determinación del umbral local.

Algunas de las técnicas de umbralización están basadas en la información estadística que brinda el histograma, sobre todo en aquellas imágenes donde los objetos tienen una superficie o textura homogénea y el fondo resulta principalmente

uniforme. El desafío de la umbralización es encontrar el valor del umbral adecuado, entre los valores de grises de la imagen, permitiendo una óptima separación entre el objeto y el fondo.

Si el histograma de intensidad de una imagen (niveles de gris) posee dos picos o lóbulos podemos separar dos zonas o regiones que destacarían el objeto y el fondo. En este caso el umbral óptimo es un valor que se encuentre entre ambos picos en el valle.

3. Alternativas de digitalización de ECG

El electrocardiograma (ECG) es uno de los métodos más practicados para detectar cualquier anomalía en la función cardíaca. Hoy en día los avances tecnológicos permiten que muchos equipos generen los resultados del ECG en más de un formato. A pesar de esta evolución muy pocos centros y especialistas en Cardiología guardan los ECG en formato digital. La mayoría de los resultados de la señal electrocardiográfica están disponibles como registros en papel o a lo sumo como imagen.

En este trabajo se utilizaron ECG obtenidos en porcinos y las imágenes de ECG de seres humanos con BIA correspondiente a estudios previos.

3.1. Exploración de viabilidad del uso de ECG de porcinos

Los ECG de porcinos se han convertido en un rival de los ECG de perros para experimentar en cardiología. Debido a las características anatómicas y fisiológicas de su corazón, presenta menos diferencias con el ser humano. El uso de cerdos en la investigación cardiovascular humana ha aumentado progresivamente debido a las similitudes en el tamaño del corazón, los diámetros de sus arterias coronarias entre otras características. Sin embargo para el proceso de análisis de la onda P presenta diferencias que a simple vista pueden notarse.

En la figura 3 se muestra una imagen de un ECG del dataset de porcinos en pdf.

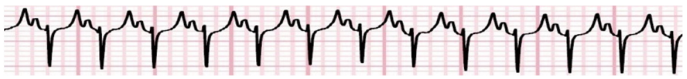


Fig. 3. Imagen de la derivación II de un cerdo con BIA.

Los valores normales de morfología y función cardíaca se han publicado para perros, ganado, hurones, entre otros. Sin embargo, falta información publicada acerca de la morfología y la función cardíaca en cerdos. Con el objetivo de mejorar cuando los datos de cerdos se extrapolan a humanos, se documentaron las medidas electrocardiográficas y ecocardiográficas de morfología y función cardíaca en 132 cerdos sanos. Se encontró como rasgo característico: una onda P bifásica, trifásica o multifásica con una fase negativa inicial en el 50% de los cerdos investigados. Se considera que los trastornos regionales de despolarización pueden sustentar la variación en la duración y la morfología de la onda P [18].

Para la detección del Síndrome de Bayés es necesario contar con las características de la duración como de la morfología de la onda P. Se realizaron búsquedas para obtener la información necesaria y previa a la aplicación de algún método sin obtener suficientes datos para avanzar en esta línea de investigación.

3.2. Digitalización de imágenes de ECG

Los resultados del ECG habitualmente se encuentran disponibles como registros en papel. Como consecuencia se genera la necesidad de almacenar y recuperar los registros de electrocardiogramas en papel del paciente para el diagnóstico de dolencias. Este proceso requiere un gran espacio de almacenamiento y un extenso esfuerzo manual. En esta sección se presentan las características del procesamiento para conversión de datos de imágenes de ECG a la forma de señal digitalizada. Se aplicaron dos métodos distintos de procesamiento, uno de ellos es utilizando una versión de prueba del software im2Graph y el otro mediante un algoritmo desarrollado con Matlab. En ambos casos se proporciona una estrategia efectiva y de muy bajo costo para la conversión de imágenes de ECG a señales digitales.

Los apartados siguientes de esta sección implican el procesamiento de imágenes de ECG mediante un conjunto de técnicas de procesamiento para la conversión de datos de imágenes de papel ECG a la forma de señal digitalizada, lo que resulta en almacenamiento y recuperación conveniente de información ECG.

3.2.1 Digitalización de imagen con Software Free

Considerando las opciones existentes en cuanto a herramientas de digitalización y el beneficio de utilizar software libre, en el contexto de esta publicación se utilizará Im2gph en su versión free para la digitalización de las imágenes de ECG de seres humanos. Este software de digitalización convierte gráficos en datos mediante rápidos y eficientes algoritmos de procesamiento de imágenes.

Para proceder a reconstruir la señal digital se utilizan los datos brindados por el software y mediante Microsoft Excel se procesaron.

3.2.2 Algoritmo en Matlab

Consiste en el desarrollo de una herramienta de software propio destinado a convertir una imagen correspondiente a un ECG a un archivo de datos, que puede procesarse y almacenarse de manera eficiente. Se trata de un enfoque alternativo a las opciones actuales.

El software permite ingresar el valor de un umbral global para realizar una segmentación de la imagen. A continuación el software permite segmentar la imagen por área ingresando las coordenadas específicas. Los resultados se presentan en la siguiente sección.

4. Resultados

En esta sección se presentan los resultados obtenidos al aplicar los algoritmos presentados en las secciones 3.2.1 y 3.2.2.

4.1 Resultados de digitalización con software

La figura 4 muestra el ECG original seleccionado, mientras que la figura 5 la señal obtenida a partir de los datos del software Im2graph. Como puede observarse la morfología de la onda P de interés se mantiene.

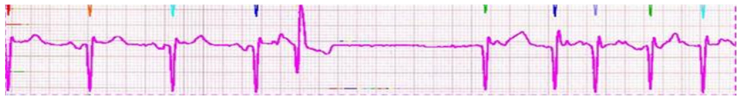


Fig. 4. Imagen Original seleccionada para obtener los puntos del ECG



Fig. 5. Gráfica obtenida a partir de los datos del software

4.2 Resultados de digitalización con Matlab

En las figuras 6 a 10 se pueden observar las diferencias al aplicar el algoritmo implementado en Matlab.

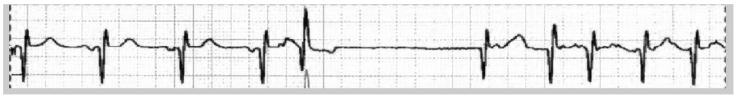


Fig. 6-a. Imagen del ECG original obtenida con Matlab

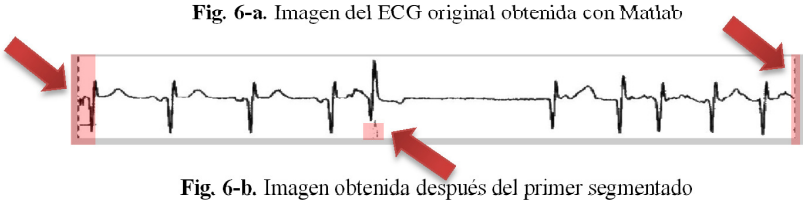


Fig. 6-b. Imagen obtenida después del primer segmentado

La imagen 6-a muestra la imagen original de la derivación II que se ha ingresado en el software. Al aplicarle un valor de umbral global, previo análisis del histograma de la imagen, se obtiene la figura 6-b. En la misma puede observarse que se ha logrado quitar la grilla.

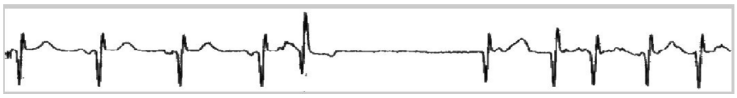


Fig. 7. Imagen final obtenida

La figura 7 muestra que se logra eliminar partes de otras derivaciones que habían quedado en la imagen y los detalles de los extremos, señaladas mediante flechas en la figura 6-b. Para finalizar se muestran en las figuras 8 a 10 la imagen original y las dos señales obtenidas.



Fig. 8. Imagen Original derivación II

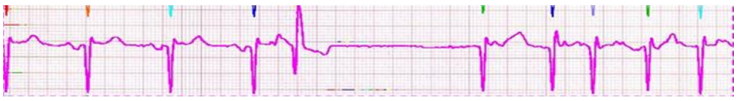


Fig. 9. Imagen Original en color rosa utilizada para prueba



Fig. 10-a. Imagen final obtenida con el algoritmo en Matlab.



Fig. 10-b. Imagen obtenida en Matlab de los datos procesados provenientes de Im2gph.

La figura 8 muestra la imagen original de la derivación II, mientras que en la figura 9 se puede observar la señal de la imagen seleccionada con el software Im2gph para posteriormente obtener los datos de interés. La imagen 10 es el resultado final obtenido por el algoritmo desarrollado. En la imagen 10-a se presenta la imagen obtenida con el algoritmo desarrollado en Matlab y la figura 10-b la señal reconstruida en base a los datos provenientes del software Im2gph.

A partir de los resultados obtenidos se puede visualizar que la morfología de la onda P que resulta de interés se logra mantener en el proceso de digitalización de imagen del ECG. Sin ser necesario realizar un proceso de filtrado sobre la misma. La preservación de la onda P fue evaluada cualitativamente por los expertos. Cabe aclarar que en el contexto del dataset disponible, no es posible llevar adelante una

evaluación cuantitativa de la onda P debido a que no se dispone de un ground truth de la misma.

5. CONCLUSION

La electrocardiografía desempeña un papel en el diagnóstico de alteraciones de conducción auricular y ventricular como de muchas cardiopatías hereditarias. El diagnóstico fiable es posible mediante el ECG.

El Síndrome de Bayés presenta notables asociaciones con múltiples afecciones médicas, en gran medida subdiagnosticado. Frente a esta realidad, resulta relevante la prevención y el tratamiento adecuado.

La permanente evolución de las técnicas y dispositivos para la adquisición de datos fisiológicos han permitido disponer hoy en día de los resultados en diferentes formatos. Los electrocardiógrafos eran de inscripción directa de tipo analógico y los registros de ECG por lo general se realizaban en papel termosensible. En la actualidad la mayoría de los ECG convierten la señal analógica de ECG en digital [19]. A pesar de contar con esta ventaja, los cardiólogos y centros de salud, en la mayoría de los casos, no se benefician de estos avances. Analizan los ECG en papel o en imágenes. Un formato digital de la señal permitiría ahorrar en almacenamiento, reducir el impacto medioambiental al eliminar el uso de papel y facilitaría el procesamiento para detección de anormalidades. Frente a este contexto y con los ECG disponibles, en este trabajo se han presentado dos alternativas para obtener la señal digitalizada de imágenes de ECG.

Los resultados obtenidos y derivados del uso de las herramientas de digitalización de ECG representan una alternativa cuando la señal obtenida sólo está disponible en papel o en imágenes. Los mismos son alentadores y muestran que se pueden utilizar para las etapas posteriores del proyecto que implica el procesamiento y la clasificación de la señal para detectar el Síndrome de Bayés.

Bibliografía

1. Organización Mundial de la Salud, http://www.who.int/cardiovascular_diseases/about_cvd/es

2. Houssein, Kilany, Hassanien, Snasel.: A two-stage Feature Extraction Approach for ECG signals. AECIA: International Afro-European Conference for Industrial Advancement (2016)

3. Nikan, Gwadry-Sridhar, Bauer.: Pattern Recognition Application in ECG Arrhythmia Classification. 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC), pages 48-56. (2017)

4. Bayés de Luna.: Bloqueo a Nivel Auricular. Rev Esp Cardiol.; 32:5. (1979)

5. Bayés de Luna, Cladellas, Oter, Torner, Guindo, Martí, et al.: Interatrial conduction block and retrograde activation of the left atrium and paroxysmal supraventricular tachyarrhythmia. Eur Heart J ; 9:1112-8. (1988)

7. Bayés de Luna, Baranchuk, Escobar Robledo, Massó van Roessel, Martínez-Sellés.: Diagnosis of interatrial block, Journal of Geriatric Cardiology. (2017)

8. Kitkungvan, Spodick.: Interatrial block: is it time for more attention? Journal of Electrocardiology 42 687–692. (2009)

9. Ariyarajah, Puri, Apiyasawat, et al.: Interatrial block: A novel risk factor for embolic stroke? Ann Noninvasive Electrocardiol. ; 12:15-20. (2007)

10. Macfarlane, Devine, Clark.: The University of Glasgow (Uni-G) ECG Analysis Program. Computers in Cardiology ; 32:451–454. (2005)

11. Conde, Baranchuk.: Bloqueo interauricular como sustrato anatómico-eléctrico de arritmias supraventriculares: síndrome de Bayés. Arch Cardiol Mex.; 84(1):32-40. (2014)

12. Bayés de Luna, Cladellas, Oter, Torner, Guindo, Martí, et al.: Bloqueo interauricular como sustrato anatómico-eléctrico de arritmias supraventriculares: síndrome de Bayés. Arch Cardiol Mex ; 84(1):32-40. (2014)

13. Baranchuk, Conde, Bayés de Luna.: Síndrome de Bayés: la Importancia de su Reconocimiento. Revista CONAREC; 31(129):76-78. (2015)

14. Zamora, Arakaki, Rivero, Guetta, Thierer, Boissonnet.: Síndrome de Bayés: Tributo Merecido a un Maestro de la Electrocardiografía Mundial, Revista CONAREC; 31(129):116-118. (2015)

15. Bayés de Luna, Platonov, Cosio, et al.: Interatrial blocks. A separate entity from left atrial enlargement: A consensus Report. J Electrocardiol. ; 45:445-51. (2012)

16. Triana, Jaramillo, Gutiérrez, Rodríguez.: Técnicas de umbralización para el procesamiento digital de imágenes de GEMFoils. Scientia et Technica Año XXI, Vol. 21, No. 4. (2016)

17. Gonzalez and Woods.: Digital Image Processing. N.J.: Prentice Hall, (2008)

18. Paslawska, Noszczyk-Nowak, Paslawski, Janiszewski, Kiczak, Zysko, Nicpon, Jankowska, Szuba, Ponikowsk.: Normal electrocardiographic and echocardiographic (M-mode and wodimensional) values in Polish Landrace pigs. Paslawska et al. Acta Veterinaria Scandinavica, 56:54. (2014).

19. Bayés de Luna.: Electrocardiografía Clínica. Editorial Permanyer. Ed. 7ma. (2012)

Prototipo Avanzado de un Simulador de Planificación de Tareas de Tiempo Real en Sistemas Heterogéneos

Martín Joel Rodríguez [1] y Claudio Aciti [1 2]

[1] Universidad Nacional de Tres de Febrero, Sede Caseros
Valentín Gómez 4828, Caseros (B1678ABJ), Buenos Aires, Argentina

[2] Universidad Nacional del Centro de la Provincia de Buenos Aires.
Pinto 399, Tandil (7000), Buenos Aires, Argentina

{mjrodriguez@untref.edu.ar, caciti@exa.unicen.edu.ar}

Abstract. Se desarrolló una aplicación que simula el comportamiento de un conjunto de tareas periódicas en sistemas de tiempo real con procesadores heterogéneos. El simulador permite elegir entre diferentes algoritmos de planificación, se implementaron PF y EDF, y es extensible a incorporar nuevos algoritmos. Para los casos de inclumplimiento de plazos, se implementaron los mecanismos Aborto y Finalización Tardía, y es extensible a incorporar nuevos mecanismos. En cada simulación, se debe definir previamente un conjunto de datos con las configuraciones de los procesadores y las tareas que serán simuladas, y se le debe especificar el número de iteraciones a evaluar. El simulador procesa esta información y entrega el resultado.

Keywords: Sistemas de tiempo real - Procesadores heterogéneos - Planificación de tareas

1 Introducción

Los sistemas heterogéneos avanzan paulatinamente ofreciendo altos niveles de performance y de optimización del uso de energía, comparados con los sistemas tradicionales [1]. Históricamente, los sistemas de tiempo-real (STR) se ejecutaban en sistemas monoprocesadores, sin embargo, en la actualidad los STR están migrando a arquitecturas heterogéneas, es decir, arquitecturas de computación que cuentan con múltiples elementos de procesamiento para proporcionar un balance óptimo entre rendimiento, latencia, flexibilidad, costo, entre otros factores [2, 3].

1.1 Antecedentes

Los sistemas heterogéneos, están formados por distintos procesadores pudiendo contener uno o más de cada tipo. Estos sistemas, pueden utilizar diferentes tipos de procesadores, por ejemplo, un procesador de propósito general (GPP) y un procesador

de propósito especial (GPU, FPGA, DSP, etc.) [4]. Dentro de un sistema heterogéneo es posible implementar un STR. Por definición, un STR, es aquel encargado de producir respuestas dentro de un determinado tiempo finito, es decir, la calidad de las respuestas del sistema se ve condicionada por el momento en que se producen las mismas [5, 6].

Uno de los principales retos que surge es la planificación de las tareas de tiempo-real en multiprocesadores heterogéneos. Los STR, por lo general, están constituidos por tareas periódicas que incluyen, entre sus parámetros, los instantes máximos en que las mismas deben finalizar su ejecución. Este parámetro extra se denomina deadline[1]. Si una tarea finaliza después de este tiempo, se dice que ha perdido su vencimiento [4, 7].

Los STR, por lo general, cuentan con un Sistema Operativo de Tiempo Real (RTOS) donde una de las principales funciones es la planificación de las tareas a ejecutar. El planificador debe elegir cual tarea ejecutar a continuación. Para hacer esto, primero, examina la cola de tareas listas para su ejecución y, dependiendo de la política de ejecución implementada, elige una para ejecutarla. Al planificar tareas en sistemas heterogéneos, la complejidad aumenta notablemente, porque no solo hay que decidir en qué procesador se va a ejecutar la tarea siguiente, sino que además el RTOS debe decidir en qué tipo de procesador se puede realizar la ejecución. Es por esto que, de antemano, el RTOS debe saber los tipos de procesadores de los cuales dispone y en cuales se puede ejecutar cada una de las tareas que integran el sistema [8].

En la actualidad, un conjunto de aplicaciones y frameworks existen en la literatura para la simulación de planificaciones de tareas en STR. Entre los relevados se pueden mencionar a STRESS [9], PERTS [10], YASA [11], Cheddar [12], RealTTS [13] y el simulador de la Université Libre de Bruxelles [14]. Aplicaciones como MAST [15] ofrecen herramientas de modelado, y otras como FORTISSIMO [16] presentan una plataforma para el diseño de simulaciones.

1.2 Motivación y Objetivo

La atracción generada en la última década por los sistemas heterogéneos produjo que el desarrollo de nuevas estrategias de planificación sea un activo campo de investigación debido a que los algoritmos de planificación juegan un importante papel en los sistemas heterogéneos [17]. Para explotar todo el potencial de un componente heterogéneo es necesario asignar correctamente la carga de trabajo. La planificación de tareas y el equilibrio de carga del sistema se vuelven más complejos en los sistemas heterogéneos. En los STR, la planificación de las tareas es fundamental para garantizar que las tareas se ejecuten correctamente dentro de los tiempos previamente asignados.

Considerando este contexto y la problemática de la planificación, que resulta fundamental en todo sistema, junto con el déficit generado por no encontrar un software capaz de realizar planificaciones de tareas en STR con procesadores

[1] Del inglés: vencimiento.

heterogéneos, se propone como objetivo el desarrollo e implementación de un prototipo de simulador de planificador de tareas en STR con procesadores heterogéneos.

Entre los objetivos parciales, el prototipo de simulador deberá permitir seleccionar el algoritmo de planificación y el mecanismo para el manejo de tareas no cumplidas por sobrecarga. El simulador realizará planificaciones de *N* tareas en *M* procesadores (pudiendo estos ser distintos), y a cada tarea se le podrá indicar en qué tipo de procesador podrá ejecutarse, ya que al trabajar con sistemas heterogéneos puede darse el caso de que una tarea sea ejecutable en más de un tipo de procesador o por lo contrario que una tarea pueda ejecutarse en una tipo de procesador y no en otro.

Se propone modularizar cada paso de la utilización de la herramienta, en aplicaciones virtualmente independientes, que pueden ser programadas en un lenguaje a elección, implementando una comunicación entre estos módulos, lo que permita al usuario reemplazar cualquier módulo por uno de su propio diseño, u obtener los datos en crudo de la simulación si así lo desea.

2 Definición formal del problema

Un STR consiste en un conjunto *P* de *N* tareas periódicas. Cada tarea $\tau_i \in \Pi$ está caracterizada por su período, T_i, el tiempo límite para ejecutarse, D_i, el tiempo de ejecución, C_i, el tiempo de inicio (offset o jitter), f_i y una prioridad, P_i. La unidad mínima de cómputo de una tarea T_i es una acción *J*. Una tarea es un conjunto de acciones similares que se repiten a lo largo del tiempo.

$$\Pi = \{\tau_i = (T_i, D_i, C_i, f_i, P_i)\} \quad (1)$$

Un sistema con multiprocesadores heterogéneos está compuesto por un conjunto Φ de tipos de procesadores diferentes φ_x.

$$\Phi = \{\varphi_{x,y},\ 1 \le x \le M,\ y \ge 0\} \quad (2)$$

El subíndice *"x"* indica el tipo de procesador y el subíndice *"y*" indica la cantidad.

2.1. STR de tareas periódicas en procesadores heterogéneos

Un STR de tareas periódicas con procesadores heterogéneos está compuesto por:

- Un conjunto de tareas $\tau_i \in \Pi$.
- Un conjunto de procesadores $\varphi_{x,y} \in \Phi$.
- Una lista de procesadores y tareas definida por pares ordenados $<\varphi_{x,y}, \tau_i>$

y sigue las condiciones que se especifican a continuación:

- Cada tarea τ_i tiene asignada una prioridad. La tarea que tiene mayor frecuencia es la que tiene mayor prioridad. Y la que tiene menor frecuencia es la que tiene menor prioridad.
- Una tarea τ_i puede ser ejecutada en uno o más tipos de procesadores $\varphi_{x,y}$.
- Una vez que una instancia q representada por τ_i, q de una tarea τ_i inició su ejecución, sólo puede continuar en procesadores del mismo tipo.
- Cada tarea deberá tener una definición para cada tipo de procesador en la que pueda ser ejecutada.

2.2. Paso a paso de la Planificación de tareas en un procesador

Los pasos propuestos para planificar una tarea en un sistema heterogéneo son:

A. El planificador chequea, en cada instante de tiempo *t*, si una tarea τ_i inicia un período nuevo (una instancia nueva) en la estructura *T*. En ese caso, $\tau_{i,q}$ debería agregarse al conjunto de tareas *LISTAS (L)*. Esto significa que la instancia anterior $\tau_{i,\,q-1}$ llegó a su tiempo de finalización.
B. Si la instancia $\tau_{i,\,q-1}$ completó su ejecución. Entonces $\tau_{i,\,q}$ es agregada al conjunto de tareas *LISTAS (L)*.
C. Si la instancia $\tau_{i,\,q-1}$ empezó su ejecución pero no la completó su ejecución. Entonces $\tau_{i,\,q-1}$ sigue a la espera en el conjunto de tareas *LISTAS (L)* y $\tau_{i,\,q-1}$ es descartada.
D. El planificador debe decidir qué tarea τ_i se va a ejecutar en el instante *t* siguiente.
E. Si no hay ninguna tarea en ejecución busca en las tareas *LISTAS* la de mayor prioridad y la ejecuta.
F. Si hay una tarea en ejecución, compara con la de mayor prioridad de las tarea *LISTAS*. Si la que está en ejecución tiene mayor prioridad o igual, entonces sigue. Caso contrario realiza un cambio de contexto.
G. Ejecuta la tarea.
H. Vuelve al punto A.

3 Solución propuesta

Se puede considerar la aplicación que realiza la simulación como la caja negra central, y las entradas y salidas de la aplicación como módulos separados independientes. Llamaremos a la aplicación que realiza la simulación backend, mientras que los módulos de entrada y salida serán parte del frontend, que serán las partes que tendrán interacción con el usuario. El módulo de entrada es el "Panel de definiciones" donde el usuario designará los procesadores y tareas que serán simulados. El módulo de salida constituye el "Gráfico de la simulación", y es donde el usuario podrá visualizar el resultado obtenido, analizar eventos y sacar conclusiones.

3.1 Lenguaje de comunicación

Al tratarse de módulos separados e independientes, se define una manera de transmitir las entradas, en un lenguaje que pueda interpretar, al backend del simulador. De la misma manera, la salida del backend debe poder ser interpretada por cualquier módulo que se quiera implementar, sin quedar restringido al funcionamiento del mismo.

Se opta por el lenguaje JSON (JavaScript Object Notation) para la intercomunicación entre módulos independientes. JSON es un formato de estándar abierto que usa texto legible por los seres humanos para transmitir datos de objetos que consisten en pares atributo-valor, muy utilizado para la comunicación asincrónica entre un servidor y el explorador.

De esta forma se puede implementar los módulos de entrada y salida en cualquier lenguaje, siempre y cuando la entrada pueda expresar la información en JSON y la salida pueda interpretar el JSON para realizar la visualización de datos.

3.2 Backend

El backend es el motor principal de la aplicación, ya que es el encargado de hacer la simulación. Recibe una entrada con la definición de procesadores y tareas que están involucrados y otros datos como el planificador de tareas que se utilizará y la cantidad de tiempo (en número de iteraciones) que se quiere simular.

Es implementado en el lenguaje Java principalmente por ser orientado a objetos, por ser multiplataforma y por tener un buen manejo de la memoria (Java cuenta con un garbage collector, o recolector de basura, para deshacerse de los objetos instanciados que ya no son accedidos). El simulador maneja muchas instancias de tareas que luego de ejecutadas son desalojadas de los procesadores y no son vueltas a utilizar.

La aplicación Java fue montada como una aplicación web en un servidor Tomcat. Esto permite que pueda ser accedida por un explorador web, enviando y recibiendo datos a través de peticiones HTTP. De esta forma, los módulos de entrada y salida

pueden comunicarse al backend del simulador estando virtualmente alojado en cualquier servidor.

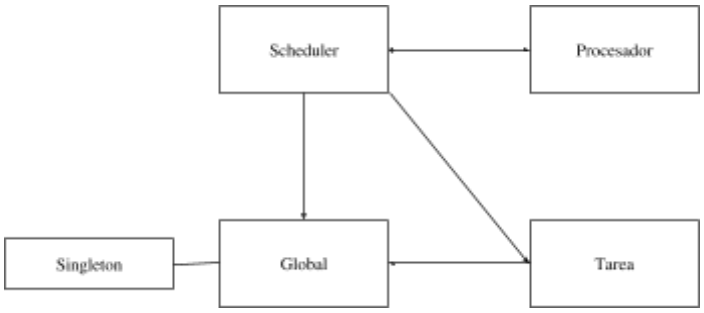


Fig 1: Diagrama de clases

Una primera aproximación es modelar los procesadores como si fueran procesadores homogéneos, es decir, que cada procesador se comporte de la misma forma y compute a la misma velocidad. En contraste, los procesadores heterogéneos sólo pueden recibir tareas específicas para un tipo de procesador, y su velocidad de cómputo puede variar según el caso. En este modelo cada procesador realiza una unidad de cómputo por unidad de tiempo.

Implementación de procesadores heterogéneos. Una vez implementada esta clase se modificó la arquitectura de la aplicación para hacer posible la existencia de procesadores heterogéneos, es decir, procesadores que solo pueden computar tareas de un tipo determinado. A su vez, los tipos de procesadores pueden diferir en su velocidad de procesamiento, haciendo que algunos computen más rápido que otros.

En principio se definieron las clases *Procesador* y *Tarea* como clases abstractas, de manera que no puedan instanciarse. Obligatoriamente los tipos de procesadores y tareas nuevas deben heredar de estas clases respectivamente.

Se define un atributo clock, cuyo valor por defecto es 1, que representa cuántas unidades de tiempo se necesitan para completar un cómputo de tarea. De esta manera, el valor 1 representa la velocidad más rápida, comportamiento esperado en la versión con procesadores homogéneos.

Para crear un tipo de procesador en particular se debe heredar de la clase Procesador. En su constructor se pasa además de su id, su velocidad de clock.

Es responsabilidad de la clase hijo sobreescribir el método *setTarea()* para restringir el tipo de tareas que puede recibir el nuevo tipo de procesador.

3.3 Frontend

El frontend constituye tanto el módulo de entrada de datos, como el de salida y visualización de la simulación. Al tratarse de módulos separados del backend de simulación, existen infinitas implementaciones que pueden ser compatibles con el simulador, siempre y cuando el lenguaje de intercomunicación sea JSON y la especificación de la transmisión de datos sea la acordada.

Para este trabajo se implementa un módulo de entrada y uno de salida como parte de una posible solución. La ventaja de tener el simulador modularizado de forma independiente es que pueden existir múltiples alternativas para las aplicaciones que se conecten al simulador.

3.4 Gráfico de la simulación

El gráfico es la última instancia de todo el recorrido de la aplicación y representa la visualización de la simulación. Al igual que el panel de definiciones, el gráfico de la simulación de este proyecto es sólo una de las infinitas soluciones que se pueden construir gracias a la modularidad del proyecto (Fig 2).

Si bien el gráfico es la salida del backend, se puede considerar a su vez como una segunda caja negra que recibe los datos de la simulación, los ordena y convierte en información gráfica para consumo del usuario. Esta transformación de los datos no es trivial y requiere algunas funciones que empaqueten los datos en formas que el usuario pueda comprender.

Representaciones. Existen varias formas de representar la simulación a partir de los datos recibidos. Debido a que se ponen en observación dos elementos distintos del sistema, procesadores y tareas, el gráfico puede representarse en función de uno o el otro. Se puede hacer una línea por cada procesador y cada casillero ocuparlo con la representación de la tarea en cuestión, o por el contrario, una línea por cada tarea y cada casillero ocuparlo con la representación del procesador en donde esta tarea se encuentra alojada. Por último se puede tener una vista cruzada, es decir, una línea por cada relación tarea-procesador posible.

Distintas vistas permiten llegar a una conclusión más exhaustiva de la simulación observada. Es por eso que para esta solución se implementaron tres visualizaciones distintas: Vista por Procesador, Vista por Tarea y Vista Cruzada.

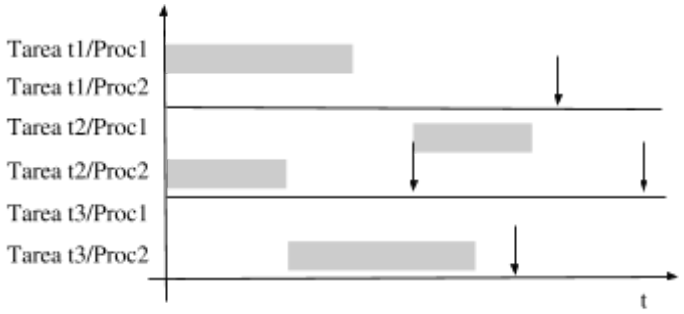


Fig. 2. Ejemplo de graficación en vista cruzada.

4 Conclusiones

El objetivo del presente trabajo ha sido desarrollar un prototipo de simulador de planificación de tareas de tiempo-real con multiprocesadores heterogéneos. Este objetivo ha sido alcanzado en su totalidad mediante el desarrollo de una solución modularizada de implementación independiente para el frontend y el backend, permitiendo una gran flexibilidad y futura explotación del simulador.

El backend de simulación se llevó a cabo como una aplicación Java alojada en un servidor web, y el frontend con HTML/CSS/JavaScript para poder armar las simulaciones y visualizarlas. Se incluye en el desarrollo la implementación de un planificador de prioridades fijas, con y sin desalojo configurable, y políticas de aborto para las fallidas. El simulador maneja tareas periódicas independientes con distintas definiciones de tiempo de inicio, cómputo, prioridad y plazo máximo de ejecución; simuladas en un grupo de procesadores heterogéneos con tipo y velocidad de cómputo configurable.

Se distinguieron y estudiaron los puntos críticos del planificador para los cuales se documentaron casos testigos para comprobar el correcto comportamiento del simulador en dichas circunstancias. Como resultado se observó que el simulador respondía de la forma esperada.

En términos generales, el simulador y la solución de frontend propuesta, brinda un entorno de estudio y desarrollo altamente expandible para futuros trabajos y soluciones alternativas, orientadas al estudio y planificación de STR.

5 Trabajos Futuros

Se destacan como trabajos futuros en la aplicación: Incorporar la posibilidad de simular tareas aperiódicas y esporádicas, definiendo una función de probabilidad para la aparición de estas, y habilitando la posibilidad de que puedan ejecutarse una única

vez. Dar acceso al simulador a través de un servidor web, aprovechando las ventajas del procesamiento dedicado y acceso remoto

Referencias

1. Brodtkorb, A. R., Dyken, C., Hagen, T. R., Hjelmervik, J. M., and Storaasli, O. O. (2010).
   State-of-the-art in heterogeneous computing. Scientific Programming, 18(1):1–33.
2. Luna, F., Nesmachnow, S., y Alba, E. (2010). Búsqueda local paralela para planificación de tareas en sistemas de computación heterogéneos. Universidad de Castilla, La Mancha
3. Orozco, J. D., Urriza, J. M., Cayssials, R., Fernández, E., Ferrari, M., Echaiz, J., Buckle, C., Barry, D., Páez, F. E., Olguín, G., et al. (2013). Sistemas de tiempo real con requerimientos heterogéneos: integración hardware-software. In XV Workshop de Investigadores en Ciencias de la Computación
4. Laplante, P. A. (1992). Real-Time Systems Design and Analysis: An Engineer's Handbook. IEEE Press, Piscataway, NJ, USA
5. Liu, C. L. and Layland, J. W. (1973). Scheduling Algorithm for multiprogramming in Hard-Real-Time Environment. Jornal of ACM.
6. Burns, A. y Wellings, A. J. (2010). Real-time systems and programming languages, volume 2097. Addison-Wesley
7. Urriza, J. , Ferrari, M., Orozco, J., Cayssials, R., Páez, F., Olguín, G., Schorb, L., Lucas, S., Tolosa, R., (2016). Sistemas de Tiempo Real Mixtos: Planificación en Sistemas Operativos de Tiempo Real Bajo Plataformas de Desarrollo Concretas. XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina)
8. Aguilar, J., Leiss, E., et al. (2004). Introducción a la computación paralela. Editorial Venezolana, Universidad de Los Andes, Mérida.
9. Audsley, N. C., Burns, A., Richardson, M. F., and Wellings, A. J. (1994). Stress: A simulator
   for hard real-time systems. Software: Practice and Experience, 24(6):543–564.
10. Liu, J. W.-S., Liu, C. L., Deng, Z., Tia, T.-S., Sun, J., Storch, M., Hull, D., Redondo, J., Bettati, R., and Silberman, A. (1996). Perts: A prototyping environment for real-tim e systems. International Journal of Software Engineering and Knowledge Engineering, 6(02):161– 177.
11. Golatowski, F., Hildebrandt, J., Blumenthal, J., and Timmermann, D. (2002). Framework for validation, test and analysis of real-time scheduling algorithms and scheduler implementations. In Rapid System Prototyping, 2002. Proceedings. 13th IEEE International Workshop on pages 146–152. IEEE.
12. Singhoff, F., Legrand, J., Nana, L., and Marcé, L. (2004). Cheddar: a flexible real time scheduling framework. In ACM SIGAda Ada Letters, volume 24, pages 1–8. ACM.
13. Diaz, A., Batista, R., and Castro, O. (2007). Realtss: a real-time scheduling simulator. In
    Electrical and Electronics Engineering, 2007. ICEEE 2007. 4th International

Conference
on, pages 165–168. IEEE.
14. Vroey, S. D., Goossens, J., and Hernalsteen, C. (1996). A generic simulator of real-time scheduling algorithms. In Simulation Symposium, 1996., Proceedings of the 29th Annual, pages 242–249. IEEE.
15. González Harbour, M., García, J. G., Gutiérrez, J. P., and Moyano, J. D. (2001). Mast: Modeling and analysis suite for real time applications. In Real-Time Systems, 13th Euromicro Conference on, 2001., pages 125–134. IEEE.
16. Kramp, T., Adrian, M., and Koster, R. (2000). An open framework for real-time scheduling
simulation. Springer.
17. Mednis, A., Strazdins, G., Zviedris, R., Kanonirs, G., and Selavo, L. (2011). Real time pothole detection using android smartphones with accelerometers. In Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on, pages 1–6. IEEE.

Workshop Innovación en Educación en Informática (WIEI)

Coordinadores

Cecilia Sanz (UNLP) - Beatriz Depetris (UNDTF) - Marcelo De Vincenzi (UAI)

Temas

Estrategias y metodologías innovadoras de Enseñanza
Descripción y evaluación de experiencias innovadoras de enseñanza
Herramientas estratégicas en el desarrollo de procesos educativos en la formación en Informática
Procesos colaborativos mediados por TIC
Competencias del profesional docente del área Informática /Ciencias de la Computación
Diseños curriculares innovadores
Ubicuidad en los proceso de enseñar y aprender
Incidencias de las actuales tecnologías móviles
Aprendizaje de la Informática centrado en el alumno

Adelantos en el desarrollo de un Protocolo de Interpretación Gráfica para no videntes y herramienta de implementación

Hernán Amatriain, Hernán Merlino, Sebastián Martins, Santiago Bianco, Nicolás Pérez y Darío Rodríguez

Laboratorio de Investigación y Desarrollo en Ingeniería de Software
Grupo de Investigación en Sistemas de Información.
Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús
http://www.unla.edu.ar/sistemas/gisi/
hamatriain@gmail.com, hmerlino@gmail.com, smartins089@gmail.com,
santiago.bianco.sb@gmail.com

Resumen. El presente trabajo busca cubrir la vacancia existente en un protocolo y un conjunto de herramientas que den soporte a los no videntes para comunicar ideas a través de diagramas de sistemas. Al no existir en la actualidad un conjunto completo de estas herramientas, no puede llevarse a cabo la inclusión completa de personas no vidente en la Educación Superior.

Palabras clave. Modelado de Sistemas, UML, DFD, DER, No Vidente, Discapacidad, Inclusión, Ciego.

1. Introducción

La Ley Nacional de Educación Superior establece que el Estado debe ofrecer acceso a la educación universitaria a todo aquel que quiera hacerlo y cuente con la formación y capacidad requerida, y garantizar la accesibilidad al medio físico, servicios de interpretación y los apoyos técnicos necesarios y suficientes, para las personas con discapacidad [Ley 25573, 2002]. Queda en evidencia que en algunas áreas de menor demanda los desarrollos son aún escasos. Las carreras de Sistemas de Información / Informática, utilizan asiduamente modelos gráficos para el diseño de funcionalidades y datos. La interpretación de estos modelos hace complicada la comprensión del tema para el estudiante no vidente, y muy difícil la tarea del docente para que el estudiante los asimile y utilice para su posterior evaluación. Las estructuras curriculares de estas carreras exigen cubrir estos temas.
En la Universidad Nacional de Lanús se presentó un caso puntual que puso a prueba la estructura y organización para cubrir estos casos, desnudando la falta de madurez y estrategias de la Institución para afrontarlos. Un estudiante no vidente, cursante de la carrera Licenciatura en Sistemas, del Departamento de Desarrollo Productivo y Tecnológico fue el detonante para comprender que debía trabajarse en este sentido y poder darle las herramientas necesarias, tanto al estudiante como a los docentes, para

facilitar sus estudios. Éste alumno, en su adolescencia perdió la vista y actualmente es difícil para él la comprensión de gráficos de sistemas. No existe un procedimiento para evaluar y entrenar a estos estudiantes en el manejo de formalismos gráficos, presentándose siempre como un desafío al docente que enfrenta tal situación.
El presente trabajo busca desarrollar un protocolo de comprensión gráfica para facilitar la tarea del docente en explicar estos modelos y la del estudiante no vidente comprenderlos y utilizarlos en la resolución de problemas, para su posterior evaluación. También se busca desarrollar un conjunto completo de herramientas que implemente dicho protocolo. En este trabajo presenta: el problema abordado (sección 2), la solución propuesta (sección 3), los adelantos en la etapa de validación del protocolo y herramienta desarrollados (sección 4) y se plantean algunas conclusiones y futuras investigaciones (sección 5).

2. Descripción del Problema

En la Universidad Nacional de Lanús, se viene haciendo un trabajo de interpretación y apoyo técnico a personas con discapacidad desde el Programa de Inclusión Universitaria para Personas con Discapacidad de la Dirección de Bienestar Universitario dependiente de la Secretaría de Cooperación y Servicio Público. Esta iniciativa se fundamenta en la necesidad de abordar una problemática social más amplia, que remite a las oportunidades educativas que poseen aquellos sectores de la población históricamente excluidos del Sistema de Educación Superior, partiendo de una nueva manera de concebir, diseñar e implementar políticas universitarias, asentada en el reconocimiento, respeto y afirmación de las diferencias inherentes a la población estudiantil, congruente con la política de democratización del conocimiento que lleva adelante la Universidad Nacional de Lanús. Hasta hace poco tiempo la discapacidad era pensada como una cuestión que concernía exclusivamente a las personas afectadas, quienes con asistencia y protección podían rehabilitarse a fin de adaptarse al medio y lograr la integración social. Centrado en los aspectos médicos, en lo patológico, este modelo médico-céntrico de la discapacidad perdía de vista los derechos de este grupo heterogéneo, como personas y como ciudadanos, ignorando que lo que está en juego es la calidad de vida y la construcción de una sociedad realmente inclusiva, capaz de enriquecerse de la diversidad.
Sin embargo se ha presentado un caso que pone en evidencia la falta de madurez que se tiene en ciertas áreas específicas con discapacidades puntuales. Particularmente se presentó una prueba difícil al equipo interdisciplinario cuando el estudiante no vidente de la carrera de Licenciatura en Sistemas del Departamento de Desarrollo Productivo y Tecnológico, tuvo que aprender y ser evaluado en algunas técnicas y metodologías de modelado de sistemas que son esencialmente gráficas como los Diagramas de Flujo Nassi-Shneiderman, Diagramas Entidad Relación (DER), Lenguaje Unificado de Modelado (UML), que son solo algunas de las herramientas gráficas que forman parte curricular de la mencionada carrera (acreditada por Resolución CONEAU 1089/12). Efectivamente, no existen herramientas ni procedimientos especiales para evaluar a un no vidente en estas áreas específicas.

No existe un procedimiento para evaluar y entrenar a estos estudiantes en el manejo de formalismos gráficos, presentándose siempre como un desafío al docente que enfrenta tal situación. Se han buscado distintas formas de darle una solución a este y otros inconvenientes que se presentan a los alumnos con estas características, trabajando con equipos interdisciplinarios, pero más allá de muy buenas ideas y apoyo que ha recibido el estudiante en las distintas áreas, el diseño utilizando modelado con gráficos ha sido hasta el momento un escollo insoslayable.
En un intento de solucionar este problema instalado en la actualidad, se buscaron herramientas especiales, encontrando que algunas posibles soluciones como ser el uso de [PSeInt, Planttext, PlantUML] son parciales e insuficientes.
El inconveniente con las herramientas mencionadas es que son incompletas en cuanto por un lado pueden generar gráficos a partir de un código, pero no pueden realizar el trabajo inverso, haciendo imposible que un ciego interprete un diagrama; y por otro lado, no abarcan todos los diagramas de modelado de sistemas, dejando afuera, por ejemplo, los Diagramas Entidad Relación (DER) que se utilizan para la modelización de datos (para la posterior creación de la Base de Datos).

3. Solución Propuesta

El presente trabajo busca desarrollar un protocolo de comprensión gráfica y un conjunto completo de herramientas para que un no vidente pueda interactuar con un grupo de estudio o docente (en un ambiente de aprendizaje o evaluación) o un grupo de desarrollo (en un ambiente laboral) a través de estos diagramas de modelado de sistemas. Particularmente se busca solucionar el problema a través de los siguientes desarrollos:
Desarrollo 1: Un lenguaje de interpretación gráfica que pueda codificar los diagramas de modelado de sistemas existentes más utilizados de forma tal que pueda ser interpretado por un no vidente, comprendiendo un gráfico así especificado o que pueda escribir el código que genera un diagrama específico. Este lenguaje debe ir acompañado de un proceso de evaluación para que los docentes cuenten con un protocolo o metodología de actuación a la hora de evaluar a un disminuido visual en el área de utilización de diagramas de modelado de sistemas.
Desarrollo 2: Un conjunto de herramientas de interpretación gráfica que pueda codificar los diagramas de modelado de sistemas existentes más utilizados de forma tal que pueda ser interpretado por un no vidente, comprendiendo un gráfico así especificado o que pueda escribir el código que genera un diagrama específico.

3.1. Protocolo de interpretación gráfica

El desarrollo de un protocolo para interpretación gráfica de diagramas de diseño de sistemas tiene dos objetivos. Por un lado, se busca conseguir un lenguaje a través del cual un no vidente pueda comprender un diagrama de diseño de sistemas (DFD, DER, UML, etc) que haya sido codificado con el mismo; y por otro lado ser la base para la creación de un conjunto de herramientas que puedan pasar tanto del lenguaje a un gráfico y viceversa.

La codificación de un diagrama de sistema a través del presente protocolo, debe tener en cuenta el objetivo de un gráfico, y la construcción de conocimiento de un individuo que carece de visión. Debe tenerse en cuenta que la representación interna del conocimiento de un no vidente es muy distinta a quien pudiendo ver, analiza las relaciones existentes en un diagrama (por ejemplo un DER) con una simple inspección visual. Los gráficos como son los diagramas de clases, DER, casos de uso, DFD, etc., además de ser el formalismo de un diseño, buscan que con un "golpe de vista" el diseñador tenga una primera impresión de la parte de sistema que se está modelando, o la pueda comunicar. Además de ello, un diseño con formalismo gráfico constituye una fuente de consulta permanente, de tal modo que el diseñador rápidamente puede observar alguna relación o característica que no terminó de memorizar o haya olvidado. Por ello es un desafío conseguir que un no vidente pueda interpretar un formalismo gráfico a partir de un código y que el mismo sea una fuente de consulta rápida. La forma de percibir de un no vidente un código escrito es de manera "lineal" a partir de un audio y al cual se accede secuencialmente. La solución que se ha planteado a este inconveniente, es la creación de un índice de rápido acceso, de tal forma que el individuo disminuido visualmente pueda navegarlo y detenerse en el elemento que desee revisar (sobre el cual quiere "refrescar la memoria").
Los diagramas sobre los cuales se ha desarrollado el protocolo son los siguientes: Diagrama de Flujo de Datos (DFD), que incluye Diagrama de Contexto (DC), Tabla de Eventos (TE), y Diagrama de Sistemas (DS); Diagrama de Entidad-Relación (DER); Diagrama de Clases: se utiliza UML; Diagrama de Casos de Uso: se utiliza UML; y Diagrama de Secuencia: se utiliza UML.

Diagrama de Contexto:
La forma de codificar el diagrama de contexto es indicando los flujos por sus nombres y de dónde hacia dónde se dirigen, anteponiendo el elemento generador (Entidad Externa o Sistema) y poniendo posteriormente el elemento receptor. Se cumple con la sintaxis:
Diagrama de Contexto:
<Entidad Externa>: <Flujo>: <Sistema>;
<Sistema>: <Flujo>: <Entidad Externa>;
Índex: <Entidad Externa>
Nota aclaratoria: el índice tendrá una lista de las entidades externas, de manera que al seleccionar cualquiera de ellas se tiene acceso a los flujos que parten de ella al sistema

Tabla de Eventos:
Se describe cada evento como un elemento u objeto con una serie de características.
Tabla de Eventos:
Evento <Nro>:
Tipo: <Externo / Temporal>
Entidad Externa: <Nombre EE / vacío>
Descripción: <Breve descripción del evento>
Estímulo: <Flujo que activa al sistema / vacío>
Respuesta: 1. <Respuesta 1>, 2. <Respuesta 2>
Función asociada: <Nombre función asociada>;
Índex: Evento <Nro> (<Descripción>)

Diagrama de Sistemas:
Diagrama de Sistemas:
Proceso <Nro>: //Nota: el número de proceso se corresponde unívocamente con el número de evento.
Nombre: <Nombre del proceso> //Nota: corresponde a la función asociada del evento.
Evento Asociado: (<temporal / externo>) <Descripción del evento>
Entidad Externa: <Nombre de todas las EE asociadas, se desprende de la descripción>
Demoras: <Nombre de las demoras asociadas, se desprende de la descripción>
Descripción: <Breve descripción coloquial de la funcionalidad que debe implementar el sistema, de manera secuencial>
Flujos:
1. <Entidad Externa>: <Flujo>: <Proceso>,
2. <Demora>: <Flujo>: <Proceso>,
3. <Proceso>: <Flujo>: <Entidad Externa>;
Índex: <Nombre Proceso>
Subindex (por cada elemento del índice):
<Descripción>
<Flujos>
<Acceso completo>

Diagrama Entidad-Relación:
El DER se representa a través de las definiciones de todas las entidades primero (con sus respectivos atributos), y posteriormente se definen las relaciones entre entidades. Se indexan solamente las entidades con un subíndice que indique si el usuario de esa entidad quiere conocer sus atributos o las relaciones con otras entidades.
Diagrama Entidad-Relación:
Entidades:
Entidad <Nombre>:
<Clave / Obligatorio / Opcional>: <Nombre Atributo>;
Relaciones:
Entidad <Nombre A>: <Relación>: Entidad <Nombre B>
Entidad <Nombre B>: <Relación>: Entidad <Nombre A>;
Nota: la relación debe ir acompañada por la modalidad y cardinalidad (número mínimo y máximo de relaciones entre entidades) debe indicarse la relación en ambos sentidos por cada par de entidades relacionadas.

Diagrama de Clases:
Para representar las clases se comienza con la declaración del tipo de clase seguida con el nombre de la misma. Se diferencian tres tipos de clases: interfaces, abstractas y concretas.
Clases concretas:
Clase <nombre clase>
Clases abstractas:
Clase <nombre clase> (Abstracta)
Interfaces:
Clase <nombre clase> (Interface)

Cuando una clase hereda de otra se indica de la siguiente forma:
Clase <nombre clase hija> HeredaDe <nombre clase madre>
Los atributos y métodos se indican a continuación del nombre de la clase anteponiendo el tipo de acceso, es decir si es privada (-), pública (+) o protegida (*):
Clase <nombre clase>
Atributos:
- atributo1
- atributo2;
Métodos:
+ metodo1 (<lista argumentos>) <tipo retorno>
+ metodo2 (<lista argumentos>) <tipo retorno>
+ metodo3 (<lista argumentos>) <tipo retorno>;
Finalmente se escriben las relaciones. Se codifican tres tipos de relaciones: herencia (ya explicada), parte de (composición, colección y agregación), y comunicación:
Relaciones:
<nombre clase hija> HeredaDe <nombre clase madre>
<nombre clase menor> EsParteDe <nombre clase mayor>
<nombre clase> una_a_muchas <nombre clase>
<nombre clase> muchas_a_muchas <nombre clase>
<nombre clase> una_a_una <nombre clase>
Índex: <Clases>

Diagrama de Casos de Uso:
Se indica el actor que interactúa con el sistema y la funcionalidad del sistema con que éste interactúa. Que es el mismo caso de uso:
<nombre actor> - <nombre caso de uso>
Los casos de uso que incluyen o extienden su comportamiento con otros casos de uso se codifican de la siguiente manera:
<caso de uso> - Incluye - <caso de uso>
<caso de uso> - Extiende - <caso de uso>
Escenarios de caso de uso:
Nombre de caso de uso: <nombre>
Identificación: <id>
Actores: <actores>
Pre condiciones: <pre condiciones existentes, contexto del caso de uso>
Tipo de señal: <interna/externa>
Secuencia principal:
1) <acciones-1>
…
N) <acciones-N>;
Secuencia alternativa:
1) <acciones-1>
…
M) <acciones-M>;
Índex: <Casos de uso>

Diagramas de Secuencia:

Este diagrama describe cada una de las secuencias del sistema. Se identifica cada una de manera numérica e indica el caso de uso asociado. Luego, el desarrollo de la misma se describe como una sucesión de flujos finalizados con punto y coma.
Diagrama de Secuencia:
Secuencia <Nro>:
Caso de Uso: <Caso de Uso asociado>
Desarrollo:
<Actor>: <Tipo mensaje> <Mensaje> (<Argumentos>): <Nombre Objeto Receptor>
<Nombre Objeto Emisor>: <Tipo mensaje> <Mensaje>(<Argumentos>): <Nombre Objeto Receptor>;
Índex: <Casos de Uso asociados>

3.2. Herramienta para desarrollo de Clases UML

El primer desarrollo en busca de tener un conjunto completo de herramientas se centró en la generación de gráficos de clases con el formalismo implementado por el Lenguaje Unificado de Modelado (UML). Esta herramienta recibió el nombre de Percepción UML, y a ella se le adicionarán, posteriormente, la interpretación del resto de los diagramas propuestos para modelar con UML.
Percepción UML, es una aplicación web, que permite a cualquier persona crear Diagrama de Clases (DC), y poder seguir trabajando sobre esta base de datos en cualquier otro lugar, cualquier otra computadora, crear distintas versiones y modificaciones de este DC, compartirlos por mail, permitir que cualquier persona con un usuario y contraseña puedan acceder a la misma para que sigan trabajando o para que simplemente hagan uso del diseño. La particularidad de Percepción UML es que el DC es transformado a audio, de manera que una persona puede crearlo y otra puede verlo o percibirlo auditivamente (escuchar un audio con el protocolo de comprensión gráfica que previamente fue enseñado). A continuación se observan capturas de pantalla de ejemplo.



Figura 1: Diagramas disponibles para seleccionar y Clases disponibles en el diagrama seleccionado

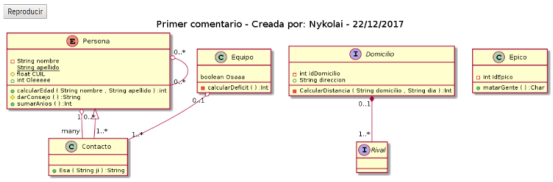


Figura 2: Diagrama de clases creado

4. Experiencia de validación del Protocolo y Herramienta desarrollados

El protocolo desarrollado inicialmente era sustancialmente distinto al presentado en este trabajo, y su cambio responde a una evolución del mismo a través de la validación realizada con el estudiante ciego de la UNLa.
Esta validación fue una experiencia interesante donde se comenzó a comprender la forma que un no vidente enfrenta a su entorno, como lo percibe y como interactúa con él. Particularmente fue adaptándose el protocolo y la herramienta de acuerdo a como lo requería el estudiante en cuestión.
Con esta experiencia es que se comprendió la necesidad de un índice para ayudar al acceso rápido. También que prefería en la medida de lo posible descripciones coloquiales antes de la utilización de signos de puntuación.
A medida que se validaba el protocolo, se aplicaba a la evaluación de una materia de la carrera: Ingeniería de Software I (materia de segundo año). El estudiante pudo desarrollar un examen al igual que sus compañeros, obteniendo puntuaciones en práctica de un 80% en el primer parcial y un 60% en el segundo, aprobando, así, la cursada de esta materia cumpliendo con las exigencias curriculares de desarrollar diagramas de flujo de datos, tabla de eventos, diagramas entidad-relación, diagramas de clases, diagramas de casos de uso y diagramas de secuencia (más flojo es éste último punto). La herramienta desarrollada también fue bien recibida y adaptada a sus necesidades.

5. Conclusiones y futuras líneas de investigación

Si bien éste fue el primer estudiante en que se aplica este protocolo y herramienta, se puede concluir que hasta aquí la experiencia fue altamente positiva, pudiéndose considerar al protocolo y herramienta desarrollados como un éxito.
Actualmente está ingresando en la misma carrera de la UNLa un estudiante disminuido visual, con el que se espera aplicar el mismo protocolo y herramientas para darle la oportunidad de valerse por sí mismo en el estudio de Sistemas.
Respecto al estudiante testigo con el que se validó el protocolo y la herramienta, debe aclararse que el tiempo insumido en la enseñanza de los mismos fue mayor que el que requirió el resto de los estudiantes para ver el mismo tema (sea DFD, DER o UML).
Actualmente se están desarrollando las herramientas para tablas de eventos, diagrama entidad-relación, diagrama de sistemas y diagrama de casos de uso. Se espera poder contar con estas herramientas para comienzos del año entrante. También se ampliará el protocolo a otros diagramas de modelado (como PERT u otros diagramas UML no desarrollados aún). Por último, se espera desarrollar en una segunda fase, herramientas de procesamiento visual e interpretación de patrones (herramientas de inteligencia artificial) para ayudar al estudiante no vidente en tiempo real y poder desarrollar gráficos en un pizarrón y éste pueda comprenderlos.

6. Financiamiento

El presente trabajo es financiado por el Proyecto 80020170400005LA de la Secretaria de Ciencia y Tecnología de la Universidad Nacional de Lanús.

7. Bibliografía

ANSI/IEEE, (2007). Draft IEEE Standard for software and system test documentation. ANSI/IEEE Std P829-2007.

Argimón J. (2004). Métodos de Investigación Clínica y Epidemiológica. Elsevier España, S.A. ISBN 9788481747096.

Arnett, K. (2003). Teacher adaptations in core French: A case study of a one grade 9 class. The Canadian Modern Language Review, 60 (2), 173-198

Basili, V. (1993). The Experimental Paradigm in Software Engineering. En Experimental Software Engineering Issues: Critical Assessment and Future Directions (Ed. Rombach, H., Basili, V., Selby, R.). Lecture Notes in Computer Science, Vol. 706. ISBN 978-3-540-57092-9.

Beardon, C; Lumsden, D; Holmes, G (1991). Natural Language and Computational Linguistics. An introduction. Ellis Horwood.Berge, N. & Berge, Z. (1998). Integration of Disabled Students into Regular Classrooms in the United States and in Victoria, Australia. The Exceptional Child, 35 (2), 107-117.

Booth, T. y Ainscow, M. (2000). Índice de inclusión, desarrollando el aprendizaje y la participación en las escuelas. Bristol: CSIE UNESCO.

Casanova, M. y Cabra, M. (Coords.) (2009). Educación y personas con discapacidad: presente y futuro. Madrid: Fundación ONCE.

Constitucón Nacional (1994). Constitución de la Nación Argentina. Art. 75: Declaraciones, Convenciones, y Pactos complementarios de derechos y garantías.

Correa, P. (2010, diciembre). Los mapas táctiles y diseño para todos los sentidos. Trilogía. Ciencia, Tecnología, Sociedad, (22), 77-87.

Delrieux, C. (2003). Introducción a la Computación Gráfica. Departamento de Ingeniería Eléctrica, Universidad Nacional del Sur, Agosto, 2003.

Echeita, G. y Duk, C. (2008). Inclusión educativa. Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambios en Educación, 6 (2). Recuperado de http://redalyc.uaemex.mx/pdf/551/55160201.pdf

Edler, R. (2008, enero-junio). Políticas de la educación especial. Revista Intercontinental de Psicología y Educación, 10 (001), 15-28.

Edward, A. (2000) Interactive Computer Graphics: a Top-Down Approach With Opengl
2nd edition, AddisonWesley, 2000.

Fernández Fernández, G.; Sáez Vacas, F. (1995). Fundamentos de informática: [lógica, autómatas, algoritmos y lenguajes].Editorial Anaya Multimedia.

Foley, J.D.; van Dam, A.; Feiner, S.K.; Hughes, J.F. (1995). Computer Graphics: Principles and Practice, 2nd edition, AddisonWesley 1995

García Martínez, R., Britos, P. (2004). Ingeniería de Sistemas Expertos. Editorial Nueva Librería. ISBN 987-1104-15-4.

Giné, C. (2001). Inclusión y sistema educativo. Presentado en el III Congreso "La atención a la diversidad en el sistema educativo", Universidad de Salamanca-Instituto Universitario de Integración en la Comunidad (INICO).

Grishman, R. (1991). Introducción a la Linguística Computacional. Visor.

González López, P. y García-Consuegrea Bleda, J. (1998). Informática Gráfica. Ediciones de la Universidad de Castilla-La Mancha-Cuenca, 1998.
González, M. (2008). Diversidad e inclusión educativa: algunas reflexiones sobre el liderazgo en el centro escolar. Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambios en Educación, 6 (2) http://redalyc.uaemex.mx/src/inicio/ArtPdfRed.jsp?iCve=55160208
Hernández, C. (2011). Desarrollo de las concepciones educativas de las personas con discapacidad visual. La Habana: Pueblo y Educación.
IEEE, (1997). IEEE Standard for Developing Software Life Cycle Processes. IEEE Std 1074-1997 (Revisión of IEEE Std 1074-1995; Replaces IEEE Std 1074.1-1995)
Jacobson, I., Ng, P. W., McMah, P. E., & Jaramillo, C. M. Z. (2013). La esencia de la ingeniería de software: El núcleo de Semat. Revista Latinoamericana de Ingeniería de Software, 1(3), 71-78.
Ley de Educación Superior (2002). Ley Nacional Nro. 25573 (modifica Ley 24521). Publicada en el Boletín Oficial del 30-abr-2002, Número: 29888, Página: 1
May, H. & Bridger, K. (2010). Developping and embedding inclusive policy and practice in higher education. York, Reino Unido: Higher Education Academy.
McLean, P., Heagney, M. & Gardner, K. (2003). Going Global: The implications for students with a disability. Higher Education Research & Development, 22 (2), 217-228.
Moreno Sandoval, A. (1998). Linguística Computacional. Síntesis.
Oktaba, H., Garcia, F., Piattini, M., Ruiz, F., Pino, F., Alquicira, C. (2007). Software Process Improvement: The Competisoft Project. IEEE Computer, 40(10): 21-28. ISSN 0018-9162.
ONU (2006). Convención sobre el derecho de las personas con discapacidad. Sede de las Naciones Unidas en Nueva York, Diciembre 2006.
OpenGL: sitio oficial disponible en https://www.opengl.org.
Orlansky, M. (1982). Education of Visually Impaired Children in the USA. Current Issues in Service Delivery, the Exceptional Child, 29 (1), 13-20.
Planttext: para diagramas UML. Sitio oficial disponible en https://www.planttext.com.
Plantuml: para diagramas UML. Sitio oficial disponible en http://plantuml.com.
PSeInt: para diagramas Nassi-Schneiderman. Sitio oficial disponible en http://pseint.sourceforge.net
Quijano, G. (2008). La inclusión: un reto para el sistema educativo costarricense. Revista Educación, 32 (1), 139-155.
Riveros, H. y Rosas, L. (1985). El Método Científico Aplicado a las Ciencias Experimentales. Editorial Trillas. México. ISBN 96-8243-893-4.
Rumbaugh, J., Jacobson, I., Booch, G. (1999). The Unified Modeling Language, Reference Manual. Addison Wesley, ISBN-10: 02-0130-998-X.
Sabato J, Mackenzie M. (1982). La Producción de Tecnología: Autónoma o Transnacional. Instituto Latinoamericano de Estudios Transnacionales - Technology & Engineering. ISBN 9789684293489.
Shieber, S.M. (1989). Introduccióna los Formalismos Gramaticales de Unificación. Teide.
Stainback, S. y Stainback, W. (1999). Aulas inclusivas. Madrid: Narcea, SA de Ediciones.
UNESCO (2009). Informe de seguimiento de la EPT en el mundo 2009: Superar la desigualdad por qué es importante la gobernanza. París.

Checkers Game, un agente de Software para jugar a las Damas

Yamil Lacoste, Paula Tristán, and
Oscar Goñi

INTIA, Fac. Cs. Exactas, UNICEN, Tandil, Argentina
yalacoste@alumnos.exa.unicen.edu.ar,ptristan3@gmail.com,oegoni@exa.unicen.edu.ar
http://www.exa.unicen.edu.ar/

Resumen *En la era digital la inclusión de las nuevas tecnologías de la información y la comunicación en el ámbito educativo tienen un impacto significativo en el desarrollo del aprendizaje de los estudiantes y en el fortalecimiento de sus competencias favoreciendo su inserción en la sociedad. La educación entonces enfrenta importantes desafíos, la inclusión de áreas de conocimiento como programación y robótica resultan elementales en la formación de profesionales del futuro. Estas áreas de conocimiento involucran que los alumnos adquieran ciertas habilidades y capacidades como: resolución de problemas, nivel de abstracción, atención, concentración, etc. Este trabajo propone la implementación de un agente de software para jugar a las damas como herramienta base para incentivar las áreas de programación, robótica y visión artificial. El entorno propuesto permite conocer y comprender conceptos básicos de cada temática y la forma en que éstas se relacionan, así como experimentar nuevas alternativas*

Keywords: Programación, robótica, visión artificial

1. Introducción

Actualmente la educación se enfrenta una serie de dificultades en cuanto a la preparación de los alumnos tal que les garantice el acceso a nuevas competencias, habilidades, aplicaciones y conocimientos. Las áreas como programación, robótica y visión artificial en etapas tempranas del aprendizaje resultan uno de los pilares fundamentales en la formación profesionales del futuro.

La programación, robótica y visión artificial forman parte de los conocimientos principales de las asignaturas de tecnología. Sin embargo, su enseñanza es una tarea ardua: «La mayoría de los educadores y docentes de programación coinciden en señalar que son pocos los estudiantes que afirman que aprender a programar una computadora es una tarea sencilla» [1]. «El hecho de que los robots se perciban como algo asociado a la alta tecnología, e incluso a la ciencia ficción, nos permite incentivar los estudios científico-tecnológicos mediante esta materia» [2]. Este trabajo aborda los diferentes pasos seguidos en el diseño

de una plataforma HW/SW que incorpora inteligencia para realizar una tarea (en este caso, jugar a las damas), sistema de interacción con el usuario, sistema de visión artificial y sistema de interacción con el entorno mediante un brazo robótico. Dada la definición precisa de sus partes y el bajo costo de implementación, este podría ser orientado a educación en nivel inicial o secundario. El trabajo interdisciplinario motiva al estudiante al diseño e implementación de su propio robot para cumplir con un área específica. Por ejemplo, un robot hexápodo podría reproducir el comportamiento de una araña en materias relacionadas con la biología. Esta propuesta fue realizada en el contexto de un trabajo final de grado de la carrera Ingeniería de Sistenas, en la cual se aborda la realización de un Brazo Robótico para Jugar a las Damas proporcionando un entorno básico para que los alumnos utilicen y puedan interactuar con estas nuevas tecnologías. La propuesta se enmarca como un mecanismo de transposición didáctica para ser llevado a las escuelas, en donde los alumnos puedan diseñar sus propios brazos robóticos, puedan aprender herramientas de hardware y software libre (por ejemplo, Arduino [3]), puedan entender los diferentes componentes electrónico y se incentive la programación.

2. Áreas del conocimiento involucradas

Tal como se expresó en la sección anterior, este trabajo propone una alternativa para jugar a las damas; que permite conocer, comprender y experimentar tres importantes áreas de conocimiento: Programación para realizar el juego, Visión Artificial para poder detectar los estados del tablero real y Mecánica con Robótica para poder realizar un brazo artesanal (Figura 1). El sistema didáctico elaborado, como se observa en la Figura 2, se divide en un sistema de control compuesto por una laptop y una placa Arduino. Allí se reciben señales de entrada del entorno (sistema sensorial), se realiza el procesamiento y al finalizar se envía el control para comandar el sistema de accionamiento de la estructura mecánica y así orientar las posiciones del brazo robótico. La inteligencia para realizar una actividad de manera autónoma, en este caso, Jugar a las Damas, es brindada por un agente Software. En el esquema del funcionamiento del agente propuesto que se indica en la Figura 3, se observan cada una de las etapas de la solución propuesta indicando además a que área del conocimiento corresponde cada una de ellas. El área de programación incluye la lógica para determinar la nueva jugada del jugador virtual, la visión artificial se encarga de captar el estado del tablero mediante el sistema sensorial y la robótica de accionar esa jugada. Al inicio es posible elegir entre varios modos de juegos: PC vs PC, Humano vs PC, Humano vs Robot, Humano vs Humano. En este trabajo en particular, sólo se descibirá el modo Humano vs Robot. El siguiente paso es realizar la configución del tablero y la evaluación de las condiciones del entorno para reconocer cada casillero y ficha del tablero adecuadamente. Una vez completada la calibración del entorno, se da inicio al juego. El comportamiento del juego se basa en un ciclo que comienza con la detección del estado del tablero y en función del jugador de turno se simulará la jugada o se va a propondrá una jugada mediante uno de los

método de interacción propuestos. Para la propuesta se incluyen los planos del brazo robótico artesanal y un instructivo de conexión. Ademas se incluye una interfaz para que los alumnos puedan elaborar sus propios brazos y que sean capaz de manipular ciertos parámetros de la interfaz con el objetivo de comprender como estas áreas se relacionan. Algunos de los parámetros a modificar son: los umbrales para detectar las fichas del tablero, los colores de la ficha del tablero, los algoritmos heurísticos para obtener la mejor jugada, los tamaños y pesos de los diferentes brazos, los sistemas de accionamiento a utilizar,etc.

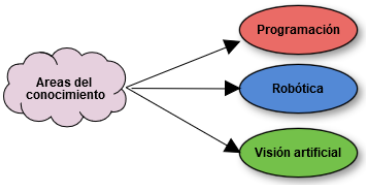


Figura 1. Áreas del conocimiento involucradas

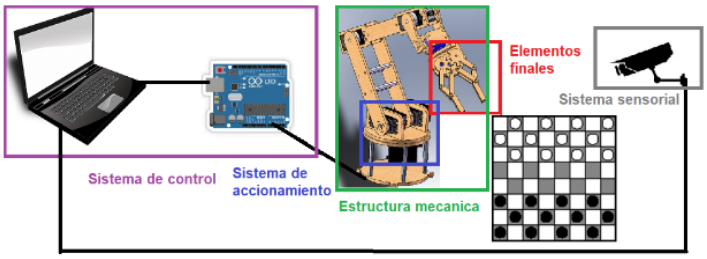


Figura 2. Morfología de un robot

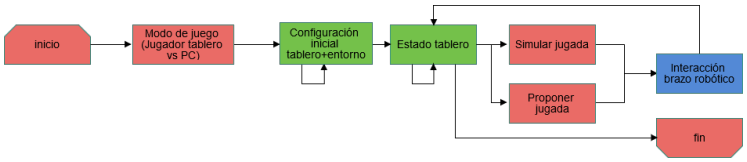


Figura 3. Esquema de funcionamiento del agente

2.1. Visión artificial

En esta sección, se hace un recorrido por las diferentes etapas del procesamiento digital de imágenes(PDI) utilizadas en la visión artificial con el objetivo de mejorar la calidad o facilitar la detección de los diferentes estados del tablero real [4].

Adquisición Para realizar la visión por computador, se utilizó en primer lugar una cámara IP de 2 Megapixel sostenida con un trípode a una altura aproximada de 30 cm en posición zenital como se observa en la Figura 4, permitiendo desplazar y aumentar la altura de la cámara con facilidad mejorando así la calidad de la captura. Además, se utilizó iluminación mediante luz natural.

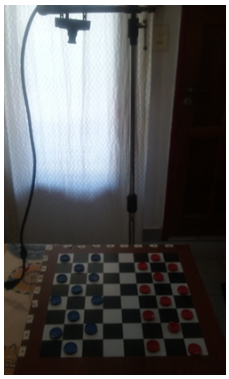


Figura 4. Captura

Realce y Segmentación En esta etapa, se realiza la detección de todos los casilleros del tablero. Las sucesivas pruebas con diferentes algoritmos indicaron que el que mejor se adaptó a la solución fue el algoritmo *goodFeaturesTotrack* implementado en OpenCV [5]. Este algoritmo permite encontrar todos los vértices del tablero de una forma automática. Este algoritmo necesita realizar una calibración del tablero. Para esto, es preciso indicarle al usuario que se realizará tal acción. Se puede observar en la Figura 5 que varios de los puntos detectados se encuentran fuera del tablero, por lo tanto es necesario recortar la foto lo máximo posible al borde hasta encontrar solamente los 80 puntos correspondiente a los vértices del tablero (Figura 6.)

Extracción de características e Interpretación Una vez aplicado el algoritmo *goodFeaturesTotrack*, se prosigue a distinguir los casilleros a los que pertenece cada ficha basándose en los 80 vértices detectados. De esta forma, es posible completar una estructura de doble entrada que permita conocer la ubicación de cada



Figura 5. Ventana de configuración del tablero



Figura 6. Resultado deseado de la configuración del tablero

una de las fichas dentro del tablero.
La determinación del color de cada casillero se obtiene mediante la identificación de los puntos centrales de cada casillero, determinando mediante un umbral si pertenece al color Blanco, Negro, Rojo, Azul, Verde, Amarillo.

2.2. Programación

El área de programación es la encargada de controlar la parte lógica de nuestro sistema como iniciar juego, indicar fin de juego, simular jugadas y fundamentalmente tiene el objetivo de encontrar las jugadas óptimas para el jugador virtual. Para esto, es necesario dotar de inteligencia al jugador virtual, se implementó el algoritmo NegaMax con poda alfa-beta para obtener la mejor jugada dentro del conjunto de alternativas [6]. La idea del algoritmo es evaluar el tablero en donde puede ocurrir una de tres opciones:

- Hay ganador: Si hay ganador se retorna un valor máximo.
- Estado Hoja: Si se llegó a un estado hoja ya sea porque la profundidad es cero, o no hay movimientos posibles se realiza una evaluación heurística para determinar una puntuación al estado del tablero.

- Estado Intermedio: Se obtiene las jugadas posibles y por cada jugada realiza un análisis en profundidad. Al finalizar, se eligen sólo los mejores valores de los parámetros de alfa y beta.

2.3. Robótica

Esta área permite interaccionar con el entorno mediante un brazo robótico con el objetivo de accionar una jugada [7]. Se proponen dos modelos de brazo robótico: uno realizado artesanalmente y otro más industrializado impreso en 3D donde para cada uno se propondrá un tablero con fichas correspondientes.

Interacción mediante brazo robótico artesanal Durante esta etapa, se crearon y diseñaron varios prototipos de brazos robóticos caseros hasta encontrar el que mejor se adaptaba (Figura 10). Estos fueron cortados en madera fibrofacil de 5mm y además se realizó un tablero artesanal con madera melamina de color marrón y de un tamaño de 40x40cm. A esta, se le realizó gráfica autoadhesiva para representar un tablero de 36x36cm, con alto y ancho de casillero de 4 cm. Para el tablero se reutilizaron las fichas del juego de mesa Cuatro en Línea, las cuales son de color rojo y azul con diámetro de 2,5 cm como se las puede observar en la Figura 6 junto al tablero. Además, para representar las fichas de Damas se colocaron unos adhesivos de color amarillo y verde al revés de la ficha como se puede observar en la Figura 9. Para la propuesta se incluyen los planos del brazo robótico artesanal(Figura 8) y un instructivo de conexión como se puede observar en la Figura 7.

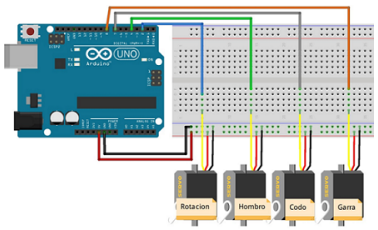


Figura 7. Instructivo de conexioń del servo a arduino

Interacción mediante brazo robótico impreso en 3D Debido a las limitaciones que presentaba el brazo robótico casero de propagación de errores y deficiencia al momento de tomar la ficha, se decidió probar un nuevo prototipo realizado con impresora 3D, el cual se lo puede observar en la Figura 11. Este donde consta con 3 grados de libertad pero con un diseño profesional, con medidas completamente exactas y sin tanta exigencia mecánica sobre los servos como el realizado de forma artesanal. Los servos utilizados son los mismos que

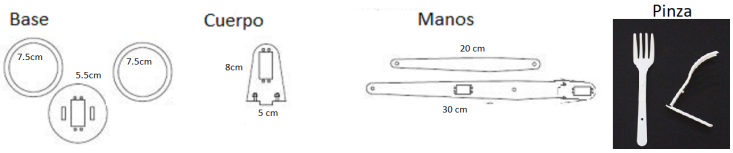


Figura 8. Planos del brazo robótico

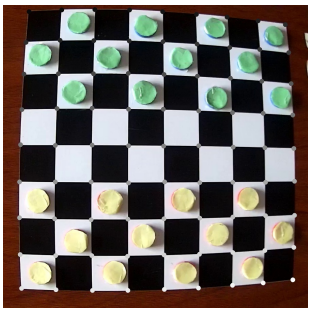


Figura 9. Tablero con 80 vértices + Fichas Damas

se utilizaron en el brazo robótico artesanal. La limitación de este modelo es que requiere de un tablero mucho más chico de 8x8cm debido a su baja amplitud de desplazamiento. Las fichas se representaron con tuercas metálicas. Esto permitió la utilización de un electroimán para realizar la recolección de la ficha. La correcta coordinación de movimiento, junto con la activación de activación del electroimán, permite la recolección de la ficha dentro del tablero y su liberación fuera de este (Figura 12).

3. Resultados de la herramienta propuesta

De manera general se puede afirmar que el entorno propuesto permite jugar a las damas con un jugador inteligente de forma física mediante procesamiento



Figura 10. Prototipo del brazo robótico casero terminado



Figura 11. Prototipo de brazo robótico en 3D con tablero y ficha

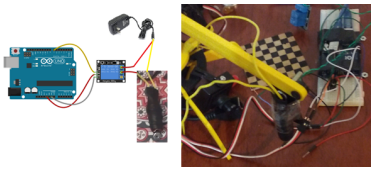


Figura 12. Prototipo de brazo robótico impreso en 3D + electroimán

de imagen, para eso es fundamental hacer una etapa de calibración para poder distinguir correctamente todas las fichas del tablero antes de jugar debido a que por condiciones invariantes de luz no se detectan correctamente todos los casilleros del tablero. Debido al problema de la luz, se hizo un análisis de la cantidad de errores detectados durante la etapa de configuración del tablero como se puede observar en la Figura 13. Para esto, lo que se hizo fue reunir los diferentes resultados obtenidos durante la etapa de configuración del tablero aplicando diferentes condiciones de luz para obtener una métrica de errores detectados.

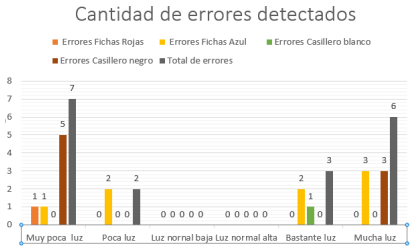


Figura 13. Resultados del procesamiento de imagen

Por otra parte, para el brazo robótico se realizaron pruebas de posición en el plano xy, los resultados obtenidos es que el prototipo va a fallar en +/- 0,5 cm

al posicionarse en una coordenada. A pesar de esto, el trabajo proporciona una alternativa didáctica que puede ser muy fructífera para que los alumnos aprendan de forma conjunta las tres áreas de conocimientos fundamentales para los estudiantes del siglo XXI de forma practica y resolviendo un problema concreto.

4. Conclusiónes y Trabajos Futuros

Mediante la realización de este trabajo, se logró cumplir con el objetivo de realizar una herramienta que permita jugar con un tablero físico sin la necesidad de un contrincante humano. El trabajo es fácilmente llevado a las escuelas para que los alumnos puedan elaborar sus propios diseños de brazos robóticos y puedan aprender resolviendo problemas reales.

En cuanto a la aplicación didáctica en educación, las etapas del diseño de la plataforma, se encuentran bien delimitadas, de manera que es posible especializar cada una por separado de manera independiente. El trabajo multidisciplinario relaciona áreas como programación, electrónica, física e inteligencia artificial.

El armado del brazo robótico representa de un gran desafío tanto en el ámbito académico asi cómo actividad extra curricular. Los requerimientos de hardware de bajo costo y software libre utilizados para este proyecto es realmente bajo y por lo tanto aplicable, aún en escuelas con bajos recursos. El desafiío que representa esta actividad en alumnos de corta edad representa un gran estimulo para el perfil del futuro profesional.

Actualmente y para garantizar la continuidad de este trabajo, se esta estructurando la propuesta didáctica ajustándola de acuerdo al nivel o ámbito en el que se pretende poner en practica.

Referencias

1. Ricardo Coppo, Javier Iparraguirre, Germán Feres, Gustavo Ursua, and Ana Cavallo. Aplicación del aprendizaje basado en problemas y la tecnología informática a la enseñanza de programación en los primeros años de ingeniería. In *XVIII Congreso Argentino de Ciencias de la Computación*, 2013.
2. Gabriel Ocaña Rebollo, Isabel María Romero Albaladejo, Francisco Gil Cuadra, and Antonio Codina Sánchez. Implantación de la nueva asignatura robótica en enseñanza secundaria y bachillerato. *Revista Investigación en la Escuela*, 87:65–79, 2015.
3. Arduino CC. Pagina oficial de arduino, Enero 2018.
4. Rafael C. Gonzlez. *Digital Image Processing*. Prentice Hall, 2002.
5. Varios Autores. Entorno de desarrollo opencv, Septiembre 2017.
6. P. Brassard, G. ; Bratley. *Fundamentos de Algoritmia*. Prentice Hall, 1997.
7. Antonio Barrientos. *Fundamentos de robtica*. S.A. MCGRAW-HILL / INTERAMERICANA DE ESPAA, 1996.

Propuestas de aplicaciones móviles desde la cátedra Sistemas de Información

Lía G. Rico[1], Laura R. Villarrubia[1]

[1] Calle Ítalo Palanca N°10 - Cátedra de Sistemas de Información - Facultad de Ingeniería - UNJu - Jujuy

liagrico@hotmail.com, lauraritavillarrubia@yahoo.com.ar

Resumen. Esta ponencia presenta el desarrollo de un trabajo de campo que se realiza en la cátedra Sistemas de Información, de las carreras Licenciatura en Sistemas, Ingeniería en Informática e Ingeniería Industrial, de la Facultad de Ingeniería de la Universidad Nacional de Jujuy. Dicho trabajo se desarrolla en forma interdisciplinaria combinando conocimientos y capacidades de estudiantes de las distintas carreras, volcados en distintos roles; además se organizan equipos de trabajo, esperando el desarrollo de la competencia de trabajo en equipo, como la participación, colaboración, comunicación, compromiso y respeto mutuo. Finalmente, como resultado del trabajo, se exponen las propuestas de las aplicaciones móviles creativas que los distintos equipos lograron modelar y las conclusiones de la experiencia.

Palabras clave: trabajo de campo - trabajo en equipo - interdisciplinariedad - desarrollo de software - prototipos

1 Introducción

La asignatura en la que se llevó a cabo esta experiencia, se denomina Sistemas de Información I, se dicta en el tercer año de la carrera de Licenciatura en Sistemas, y cuarto año de la Carreras Ingeniería Informática e Industrial, de la Facultad de Ingeniería de la Universidad Nacional de Jujuy.

El objetivo de la asignatura es lograr que el estudiante sea capaz elegir y aplicar modelos de desarrollo de soluciones informáticas, definir el tipo de sistema de información adecuado a determinada situación y diseñar una solución informática acorde a las necesidades de información especificadas.

Para el logro de este objetivo sobre equipos de trabajo interdisciplinarios fue necesario incorporar a las actividades del trabajo de campo competencias sociales y actitudinales relacionadas al desempeño efectivo en equipos de trabajo y comunicarse con efectividad, y así poder encontrar las mejores soluciones tecnológicas a retos y problemas del entorno.

En el presente trabajo se pueden visualizar los proyectos de software arribados desde el aprendizaje cooperativo.

2 Marco teórico

2.1 Trabajos de Campo

Hasta fines del siglo XX la enseñanza consistía en la transmisión de la información. De esta manera se esperaba que el estudiante desarrolle destrezas cognitivas, interactivas y manuales, y desarrolle actividades más bien rutinarias, la evaluación de los saberes consistía en una reproducción literal de saberes y estereotipos de acción, sin pretensión de cambio. El enfoque estaba en la "enseñanza", importaba el "saber". Luego el enfoque se trasladó al "aprendizaje", cobró importancia el "saber hacer". Desde fines del siglo XX se enfatizaron las competencias cognitivas, interactivas y manuales, a ser aplicadas a tareas no rutinarias; la enseñanza viró a una transferencia creativa de saberes diversos esperando un aprendizaje y dominio de competencias a ser aplicadas a situaciones nuevas y dinámicas.



Fig. 1. Antecedentes en la Educación [1]

Mucho se ha dicho en la literatura pedagógica que los propósitos de la enseñanza deben extenderse más allá de los contenidos curriculares circunscritos al aula de clase, por tanto existe la necesidad de adquirir y desarrollar múltiples capacidades en distintos contextos para afrontar las demandas de la sociedad actual. En este sentido, el trabajo de campo es un buen ejemplo pues constituye una convivencia circunstancial contextualizada fuera de la institución educativa cuya implementación requiere de una enseñanza estratégica [2]
Los trabajos de campo requieren en primer lugar un plan detallado y completo. Un segundo elemento fundamental es un compromiso especial de todas las partes en su ejecución. Finalmente, una evaluación para medir el éxito del mismo.
En la planificación de los trabajos de campo con fines didácticos, vale la aseveración que hace Jorge Steiman cuando menciona el valor pedagógico de los proyectos de cátedra: Defino al proyecto de cátedra como una propuesta académica en la educación superior en la que se explicitan ciertas previsiones, decisiones y condiciones para la práctica didáctica en el aula y que intenta hacer explícitos ciertos acuerdos que conforman aquello que

puede objetivarse del contrato didáctico que se establece con los alumnos/as y con la institución [3]
El compromiso en la ejecución del trabajo de campo se debe asumir seriamente por todos los involucrados: los docentes son una guía permanente en el trabajo, aconsejan, corrigen, miden tiempos, resultados; los estudiantes tienen que realizar un gran esfuerzo, no sólo aplican los conocimientos teóricos a la práctica, sino que intentan comprender la realidad, al tiempo que coordinan y organizan tareas, tiempos y recursos que se deben emplear para lograr el objetivo del trabajo.
Es claro que la evaluación de un trabajo de campo no sólo consistirá en "medir" saberes; la evaluación que se realiza no es al final del trabajo, sino durante todo el proceso de realización del trabajo, por lo tanto esta tarea consiste en definir los aspectos a tener en cuenta para un registro completo del desenvolvimiento de los grupos de trabajo, de sus integrantes, y de la evolución minuciosa de las soluciones y resultados que los grupos plantean.

2.2 Competencia genérica Trabajo en Equipo

La competencia de trabajo en equipo se impone a la individualización laboral. El cambio de estructura y proceso de las organizaciones de la sociedad actual ha generado un gran impacto en la nueva manera de trabajar. Las tareas han aumentado su dificultad, haciendo que su resolución individual sea imposible. Es por este motivo, que las organizaciones del trabajo reclaman, hoy más que nunca, la competencia transversal de trabajo en equipo [4]
Aurelio Villa y Manuel Poblete definen la competencia Trabajo en Equipo como "Integrarse y colaborar de forma activa en la consecución de objetivos comunes con otras personas, áreas y organizaciones"
El dominio de esta competencia está estrechamente relacionado con: Buena socialización e interés interpersonal elevado. Fuertes valores sociales que lleven a creer en la integridad, honestidad y competencia de los otros. Capacidad de comunicación interpersonal. Madurez para afrontar las diferencias de criterio. Convicción de la eficacia del trabajo compartido. Voluntad e interés por compartir libremente ideas e información. Valor de colaboración, solidaridad [5]
Manuel Poblete distingue tres niveles de dominio:

1. Participar y colaborar activamente en las tareas del equipo y fomentar la confianza, la cordialidad y la orientación de la tarea conjunta.
2. Contribuir en la consolidación y desarrollo en equipo, favoreciendo la comunicación, el reparto equilibrado de tareas, el clima interno y la cohesión.
3. Dirigir grupos de trabajo, asegurando la integración de sus miembros y su orientación a un rendimiento elevado.

3 Propuesta didáctica para el trabajo de campo

El objetivo de la propuesta es lograr que los estudiantes realicen, en equipos de trabajo, prototipos de aplicaciones móviles, siendo estas soluciones factibles a situaciones problemáticas reales del medio. Se espera que los estudiantes logren volcar conceptos teóricos a la práctica propuesta, experimenten la realización de técnicas y prácticas de relevamiento de información como entrevistas, cuestionarios, observaciones, apliquen herramientas de modelado de sistemas, desarrollen los prototipos, realicen las pruebas y expongan en forma correcta sus resultados.
El aprendizaje esperado es una combinación de la adquisición de destrezas del análisis y diseño de los sistemas como también la incorporación de las habilidades que implica la competencia de trabajo en equipo. La propuesta incluye dos fases:

1- ***Armado de equipos de trabajo e identificación del caso a estudiar***: los equipos de trabajo son organizados por los docentes de la cátedra y están conformados por estudiantes de las distintas carreras. Cada equipo escoge un nombre que lo identifica, con lo cual el "equipo" se constituye en una entidad, y sus integrantes son partes importantes en él.

La cátedra entrega a cada equipo una guía para la formulación del proyecto para generar un contexto propicio de aprendizaje. La primera tarea en equipo es la identificación de un problema auténtico en el medio que requiera una solución sistémica y que sea relevante dentro del campo disciplinar. El problema y/o oportunidad a considerar se debe acompañar de una justificación basada en un relevamiento de información, que demuestre la viabilidad del proyecto del equipo. Entran en juego varias capacidades: deben participar y colaborar en la presentación de propuestas, discusión, evaluación, decisión, análisis y resolución de problemas. Esta actividad favorece la comunicación y la orientación a la tarea y decisión conjunta.

2- ***La guía de actividades*** **para el desarrollo del proyecto de software. Cada etapa se ejemplifica en forma gráfica, con una muestra extraída de los trabajos finales de los estudiantes.**

- **Identificación del proyecto:** hace referencia a la identidad visual o corporativa del equipo de trabajo, presentan logotipos, miembros del equipo con sus roles y perfiles.



Fig. 2. Identidad corporativa. Collage de imágenes de trabajos finales.



Fig. 3. Equipos de Trabajo. Collage de imágenes de estudiantes.



Fig. 4. Merchandising corporativo.

- **Presentación del objetivo y alcance de la aplicación:** el equipo expone los motivos por los que la aplicación móvil debe realizarse, es decir, por qué se considera que la aplicación móvil es posible, adecuado, y/o conveniente. También es importante que definan la cobertura funcional del prototipo. Se solicita en esta actividad visualizar los resultados obtenidos en el relevamiento de la información.

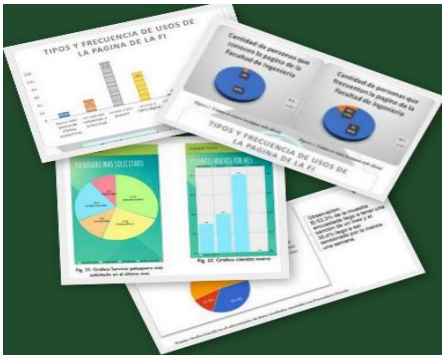


Fig. 5. Justificación del Proyecto. Resultados Encuestas Web.

- **Planificación del trabajo:** se diseñan las actividades del desarrollo de la aplicación móvil en el tiempo empleando herramientas como GANTT y PERT.



Fig. 6. Planificación de Actividades. Collage de imágenes de trabajos finales.

- **Análisis costo-beneficio:** realizan un cálculo estimado del costo del proyecto y se muestran los beneficios de implementar la aplicación.



Fig. 7. Análisis Costo-Beneficio. Collage de imágenes de trabajos finales

- **Desarrollo de la aplicación:** el equipo de trabajo elige la metodología de desarrollo de la aplicación móvil, decide el modelo de ciclo de vida, organiza las actividades de sus miembros de acuerdo a roles y funciones. Se solicita mostrar en esta actividad los requisitos funcionales y no funcionales, historias de usuario, iteraciones y pruebas de la aplicación.



Fig. 8. Metodología de software. Collage de imágenes de trabajos finales

- **Presentación del prototipo de aplicación móvil:** La exposición de cada trabajo de campo, se realizó en forma oral, cada equipo expuso en forma ordenada (Introducción-Desarrollo-Conclusión) el objetivo, alcance y la operatividad de la aplicación móvil ; y en forma escrita mediante la presentación de un informe.



Fig. 9. Interfaces Prototipos.

4 Resultados: Aplicaciones Móviles

4. 1 Prototipo de Aplicación Móvil "JUJUY EVENTOS": El objetivo de la aplicación móvil es proporcionar información suficiente, confiable y diversa de los eventos socio-culturales y deportivos de la provincia de Jujuy a todos los usuarios de la misma. Resuelve la problemática real de los jujeños que reclaman no tener al alcance la información suficiente para poder conocer los diferentes posibilidades de entretenimiento y pasatiempos, eventos de estrenos de películas en cines, obras de teatro, conciertos musicales, festivales públicos y privados, etc. Muchos usuarios se ven forzados a buscar la información navegando en Internet por distintos sitios, utilizando diarios, recorriendo las calles de la ciudad buscando carteles, yendo personalmente hasta los distintos lugares para enterarse y/o adquirir una entrada; o se enteran casualmente escuchando radio o comentando con amigos o conocidos. El alcance del trabajo es:

- El Sistema mostrará información referida a eventos como cartelera de cines, obras de teatro, festivales públicos y privados y eventos deportivos.
- Permitirá realizar búsqueda de eventos por nombre.
- Permitirá realizar búsqueda de eventos por fecha.
- Para las obras de teatro, festivales y eventos deportivos mostrará información sobre fecha, hora, lugar en donde se realizará el evento, costo y en el caso de cine mostrará además, tráiler, y sinopsis de la película.
- El usuario podrá puntuar a las películas y realizar comentarios para ampliar la información.
- El usuario podrá compartir la información del evento elegido a sus contactos a través de WhatsApp.

Existen beneficios intangibles, que no están relacionados con las ventas ni la rentabilidad. Son aquellos que a partir de la implementación del prototipo ayudarán a crecer a las personas cuyos talentos no se conocen o se ignoran. Poder permitir mostrar la diversidad que presenta la provincia de Jujuy en materia de cultura y actividades de recreación.

4.2 Prototipo de Aplicación Móvil "BIBLIOFI": El objetivo de BiblioFI es realizar consultas de disponibilidad y préstamos de libros de la biblioteca de la FI (Facultad de Ingeniería) destinada a facilitar esta actividad a los estudiantes y demás personal de la FI de la UNJu y en el caso de disponer de un libro prestado, notificaciones con antelación del próximo vencimiento de los mismos.
Alcance:

- Mostrar a los socios de biblioteca información de su estado: activo, pasivo o sancionado.
- Permitir al socio consultar material bibliográfico que tiene o puede pedir prestado.
- Notificar al socio, con 24 hs. de antelación, la devolución del libro que obtuvo mediante préstamo.
- Consultar la disponibilidad del material académico.

- Consultar fecha de devolución de un determinado libro que fue prestado, en el caso de no encontrar disponibilidad de otra unidad del mismo en la biblioteca.
- Permitir la búsqueda de libros por autor, palabras clave, título, tema.

Beneficios: Facilidad para consultar el material bibliográfico disponible sin necesidad de apersonarse en la biblioteca, lo cual ahorra tiempo y gastos de traslado; las notificaciones reducen la posibilidades de ser sancionado. Previene equivocaciones en la devolución de los libros, cuando un usuario dispone de más de un libro prestado.

4.3 Prototipo de Aplicación Móvil "Laphi Anuncios de Jujuy": El objetivo es desarrollar una aplicación móvil con el fin de brindar información sobre las distintas publicidades, ofertas y anuncios sobre los comercios de San Salvador de Jujuy a los ciudadanos y turistas.

Alcance:

- Laphi brindará información del comercio mejor puntuado y pagado.
- Laphi mostrará diferentes comercios en orden de conveniencia económica, según la franja horaria del día (por ejemplo, al mediodía los restaurantes más convenientes; a la mañana cafeterías y panaderías en primer lugar, etc.)
- Laphi determina franjas horarias y categorías de comercios (comida, vestimenta, salud, construcción, educación, etc.)
- Laphi contará con un buzón de comentarios para promover posibles mejoras en el comercio (atención al cliente, calidad de sus productos)
- Laphi implementará un sistema de calificación para los comercios (estrellas 1-5).
- Laphi mostrará la ubicación del comercio seleccionado (mapa)
- Laphi permitirá la publicación de anuncios de los comercios de S.S. de Jujuy.
- Laphi determinará los comercios "Líderes Anunciantes" en función de la calificación y comentarios que obtenga dicho comercio dentro de su categoría.
- Esta función toma en cuenta el horario y el día en que se entra en la aplicación. Según estos parámetros se ofrece las distintas opciones de comercio al usuario común.

Beneficios:

Para el propietario del comercio: Incremento de ingresos por ventas a partir de una mayor difusión y conocimiento de sus productos por medio de la app; reducción de los costos en anuncios publicitarios tradicionales (tv, radio, diarios, volantes) Como beneficios intangibles: aumento en el prestigio del comercio que utiliza la app; mejora de calidad en su producto o servicio ofrecido gracias a una mejor comunicación con los clientes.

Con respecto al usuario final, los beneficios son principalmente intangibles: ahorro de tiempo de búsqueda de comercios de determinado ramo, ó en búsqueda de mejores precios; información de comercios nuevos y productos que ofrecen a disposición del ciudadano, sin que este se traslade.

4.4 Prototipo de Aplicación Móvil "Hacela Corta": Objetivo y Alcance: "Hacela Corta" es una aplicación móvil, destinada a la Gestión de Turnos para peluquerías de la ciudad de S.S. de Jujuy con proyección a todo el país, siendo también una herramienta

para la toma de decisiones a partir de información valiosa que se podrá obtener a partir de la misma. Esta aplicación permitirá por un lado a los clientes: solicitar, consultar y seleccionar, el tipo de servicio que desea recibir y además elegir un determinado peluquero. Por otro lado, permitirá al peluquero contar con calendario de turnos y una lista de sus clientes. Al administrador o dueño de la peluquería, le brindará información para la toma de decisiones acerca de su local.
Beneficios: En nuestros días, las peluquerías, están evolucionando y adaptándose a clientes cada vez más exigentes por lo que requieren herramientas adecuadas para alcanzar dichas expectativas. El uso de la aplicación por parte de peluquerías y clientes puede representar una ventaja competitiva para las peluquerías que la adopten y una mejor imagen de estas. Tanto la peluquería como el cliente ahorran tiempo, organizan y coordinan mejor sus actividades.

5 Conclusiones

La propuesta permitió al estudiante avanzado en informática:

- Realizar un trabajo en el campo experimental, brindando una solución sistémica a un problema real.
- Hacer un abordaje interdisciplinario, integrando las perspectivas de las diversas formaciones disciplinares de los miembros del grupo.
- Analizar las diferencias y proponer alternativas de resolución, identificando áreas de acuerdo y desacuerdo, y de negociar para alcanzar consensos.

Las soluciones tecnológicas a las que se arribaron solo se pueden lograr con trabajo en equipo, imaginación, creatividad, sentido común, conocimiento académico propio y el de pares, lo que se puede observar en las imágenes.
Se realizó un proceso de autoevaluación a la práctica de enseñanza tomando como fuente de información al estudiante, siendo este un proceso de juicio de valor exitoso, ya que todos los puntos evaluados fueron validados como positivos por la mayoría.
Como propuesta a futuro se está trabajando junto a la Secretaría de Extensión, Coordinación y Planificación para dar a conocer las soluciones tecnológicas logradas por los equipos de trabajo.

6 Integrantes de los grupos

Equipo 1 "**Solutions Group":** Ávalos, Héctor Federico - Peralta, Mauricio Agustín- Zerpa, Ernesto Gabriel - Castillo, Valeria Gisela - Díaz, Érica Cecilia - Farfán, Sebastián Jorge - Lazarte, Gisela Belén - Méndez, Yesica Melisa
Equipo 2 "TecnoJuy": Guanca, Jorge Daniel Antonio - Subía, Pedro Elías Alejandro -

Molloja, Josué Emanuel - Ramos, Abel Armando - Hoyos, Alicia Micaela - Paredes, María Victoria - Sajama, Ayelen Abigail Valeria - Solíz, Betsabé Fátima Del Huerto - Soliz, Carhué Fátima Del Huerto

Equipo 3 "Laphi": Calizaya, Inés Jorgelina - Chuca, Ronaldo Javier - Contreras, Facundo - Llampa, Alvaro Facundo - Mena, Lucas Maximiliano - Meruvia, Facundo Emanuel - Ortega, Wilfredo Gabriel - Puma, Maximiliano Nicolás - Vilte, Rodolfo Alberto

Equipo 4 "Yeddi Servicios Informáticos": Ali, Luis Mariano - Fernández, Karen Marisa - García, Rocío Agustina - Ibarra Ortega, Marcos Antonio - Murillo, Nicolás Matías - Quispe, Liliana Griselda - Ramos, Pablo Alejandro - Suyo, Diego Fabián - Zelaya, Sebastián Víctor Andrés

Bibliografía

1. IAFRANCESCO V., GIOVANNI M. Currículo y plan de estudios: Estructura y planeamiento. Serie Escuela Transformadora, Libro 5. Editorial Magisterio, Bogotá, Colombia, 2004.

2. LARA DE GONZÁLEZ, L., El trabajo de campo desde la perspectiva del docente. Sapiens. Revista Universitaria de Investigación. [Fecha de consulta 01 de Junio de 2018]. Disponible en <http://www.redalyc.org/articulo.oa?id=41030367705>ISSN 1317-5815

3. STEINMAN, Jorge. "*Más Didáctica en la educación superior*". 1era edición. Buenos Aires. Unsam Edita. ISBN: 978-84- 96571-80- 8 (2008)

4. TORELLES C., COIDURAS, J., Sofía Isus, F. Xavier Carrera, Georgina París, José M. Cela. Competencia de trabajo en equipo: Definición y categorización. Revista Profesorado. Revista de currículum y formación del profesorado. VOL. 15, Nº 3 ISSN 1138-414X (edición papel). ISSN 1989-639X (edición electrónica). (2011)

5. VILLA, A., POBLETE, M. *Aprendizaje basado en competencias*. España. Edición: Mensajero. (2007)

Construyendo FrankLab: Una Plataforma Web de Robótica Educativa

Juan de la Fuente, Marcos Picucci, Daiana Bonet Peinado,
Rafael Zurita, Gerardo Parra, Jorge Rodríguez, and Laura Cecchi

Grupo de Investigación en Lenguajes e Inteligencia Artificial
Departamento de Teoría de la Computación - Facultad de Informática
Universidad Nacional del Comahue
Buenos Aires 1400, Neuquén, Argentina
{juan.delafuente,marcos.picucci,daiana.bonet,
rafa,gparra,j.rodrig,lcecchi}@fi.uncoma.edu.ar

Resumen El movimiento *CS for All* ha logrado captar la atención durante los últimos años. En este contexto se evidencia la necesidad de desarrollar recursos didácticos específicos al campo de la Educación en Ciencias de la Computación que posibiliten una mejor aproximación de la disciplina al ámbito de la educación obligatoria.
En este trabajo se presenta una plataforma Web de robótica educativa, *FrankLab*, diseñada para favorecer el desarrollo de los procesos de enseñanza y de aprendizaje de conceptos y prácticas en el área Algoritmos y Lenguajes de Programación durante toda la educación obligatoria.
FrankLab es un entorno de programación en línea que permite a estudiantes de escuelas primarias y secundarias programar robots educativos de la familia Frankestito. Dicho ambiente está compuesto por un conjunto de herramientas de programación que soporta Programación Tangible, Programación Basada en Bloques y Programación basada en texto, como Python. Los estudiantes ejecutan sus programas desde un navegador web sobre robots físicos o sobre un ambiente simulado.

Keywords: Robótica Educativa, Educación en Ciencias de la Computación, Enseñanza de la Programación, Entorno Web para Robótica Educativa, Programación Tangible, Programación Basada en Bloques

1. Introducción

El movimiento *CS for all*[9] ha recibido considerable atención durante los últimos años. En este marco se construye acuerdo en relación a la necesidad de ampliar la participación de las Ciencias de la Computación en las propuestas de enseñanza para la educación obligatoria, como forma de aumentar las posibilidades de comprensión del mundo y como política para contribuir a democratizar el acceso, para todos los estudiantes, al conocimiento producido en el ámbito de la computación [3,9].

La Robótica Educativa se presenta como un dispositivo pedagógico utilizado con frecuencia, en diferentes niveles del sistema educativo, para enseñar Ciencias de la Computación a estudiantes de todas las edades.

En este contexto existe un amplio consenso sobre que las experiencias educativas que involucran la programación de robots tienen un gran potencial para mejorar los procesos de enseñanza de aspectos de la computación, que frecuentemente se presentan como difíciles de percibir y comprender.

Una de las principales potencialidades de la Robótica Educativa es exponer y hacer tangibles conceptos abstractos de la computación. De esta forma, se favorece la construcción de conocimiento principalmente, en el área de conocimiento Algoritmos y Lenguajes de Programación. Además, se promueve el desarrollo de habilidades enmarcadas en el pensamiento computacional, como la resolución de problemas y el trabajo colaborativo.

Los trabajos realizados en el ámbito del desarrollo de plataformas de robótica educativa son numerosos y demuestran efectividad para facilitar la comprensión de conceptos fundamentales de la computación, que generalmente presentan dificultad en su enseñanza. Sin embargo, persisten campos problemáticos no explorados suficientemente en la dirección de ubicar la premisa *CS for all* y, en particular, las experiencias con robots en el plano de lo posible para más niños y adolescentes: en general las plataformas no soportan múltiples modos de programación, lo que reduce el grupo de estudiantes al que se orienta la propuesta. El acceso a los robots físicos suele ser escaso, por lo que las actividades que involucran su programación se desarrollan en espacios y tiempos limitados. Finalmente, pocas plataformas soportan capacidades que posibiliten la colaboración masiva.

Teniendo en cuenta este contexto, en este trabajo se presenta FrankLab, una plataforma Web de robótica educativa diseñada con el fin de mejorar las posibilidades de aprender Ciencias de la Computación, ampliando la población estudiantil destinataria. Así, estudiantes desde el nivel inicial hasta los últimos años de la educación secundaria, e inclusive aquellos que hayan superado el nivel medio, o aquellos que tengan alguna discapacidad dispondrán de herramientas de programación adecuadas a sus particularidades.

Esta plataforma, basada en tecnologías Web, soporta múltiples modos de programación en línea: tangible, basada en bloques, basada en texto y posibilita el abordaje de problemas complejos. Asimismo, busca mejorar la disponibilidad de los recursos, posibilitando el acceso remoto a laboratorios físicos y simulados. En cuanto a la gestión de programas, integra características como compartir, comentar y remixar proyectos para promover la colaboración masiva.

FrankLab tendrá licencia open source o compatible de acuerdo a las tecnologías Web utilizadas en su desarrollo y será accesible en forma pública y gratuita.

Este trabajo se organiza de la siguiente manera: en la siguiente sección se describe el marco de referencia para el desarrollo de FrankLab. En la sección 3 se presenta la plataforma Web de robótica educativa, analizando su estructura, los robots físicos a los que está orientada y los modos de programación que soporta. Por último, en la sección 4 se detallan las conclusiones.

2. Marco conceptual

Enseñar y que los estudiantes aprendan a programar presenta diferentes dificultades y posibilidades de acuerdo a la edad y singularidad del aprendiz. Además del desafío de diseñar soluciones algorítmicas para resolver problemas específicos, el estudiante debe desarrollar competencias para expresarlas en un lenguaje en particular.

El pensamiento computacional es una habilidad fundamental para todos, no solo para los informáticos. Entre muchos otros aspectos, pensamiento computacional es usar abstracciones y descomposiciones al intentar solucionar una tarea compleja o al diseñar un sistema de tamaño considerable. Es elegir una representación apropiada de un problema o modelar los aspectos relevantes del problema para hacerlo tratable. Normalmente, también requiere la capacidad de manejar múltiples niveles de abstracción.

La programación tangible ofrece oportunidades de aprendizaje a niños y adolescentes no alfabetizados. En este modo de programación, una colección de objetos físicos se convierten en elementos de un lenguaje reduciendo significativamente la edad para participar de este tipo de actividades formativas [1,7,8].

Por otro lado, el enfoque de programación basado en bloques ayuda a definir fácilmente el concepto de pensamiento computacional basándose en tres dimensiones: concepto, práctica y perspectiva computacional en un lenguaje de programación, que cumple con la difícil tarea de tener “piso bajo”, “techo alto” y “paredes anchas” . Esto es, facilidad para comenzar a programar, oportunidades de crear proyectos de complejidad creciente y soportar diversidad de tipos de proyectos e intereses, características primordiales para aprender a programar [2].

En trabajos anteriores a este artículo, presentamos la arquitectura de Frankestito, un robot Myro compatible de bajo costo, diseñado y construido bajo licencias libres, para ser utilizado como recurso didáctico disciplinar. Una característica distintiva de este robot educativo es la incorporación de un sistema de visión y comunicación Wi-Fi [4]. Teniendo en mente las ventajas de las capacidades mencionadas, se trabajó sobre el robot educativo Multiplo N6 Max, modificando su arquitectura de hardware y de software embebido, a fin de hacerlo compatible con Frankestito[10].

El robot educativo Frankestito se utiliza regularmente para enseñar Algoritmos y Lenguajes de Programación a estudiantes secundarios, en actividades de divulgación científica e investigación [6].

Los trabajos descriptos en esta sección presentan abordajes parciales al campo problemático en cuestión. Se considera que el tratamiento en estas áreas debe plantearse como integrado y estudiarse en forma articulada para avanzar en la elaboración de una plataforma que logre recuperar aportes realizados, en función de ampliar la posibilidades de concretar experiencias educativas basadas en la robótica educativa.

En este sentido, FrankLab es una solución integradora que intenta recuperar las potencialidades descritas en esta sección.

3. FrankLab

FrankLab es una herramienta Web diseñada con la finalidad de proveer un ambiente para la programación de robots educativos compatibles con Frankestito. Este entorno permite la codificación en línea de programas, los cuales son luego ejecutados por un robot físico o simulado, cuyo accionar puede ser visualizado en tiempo real por los usuarios.

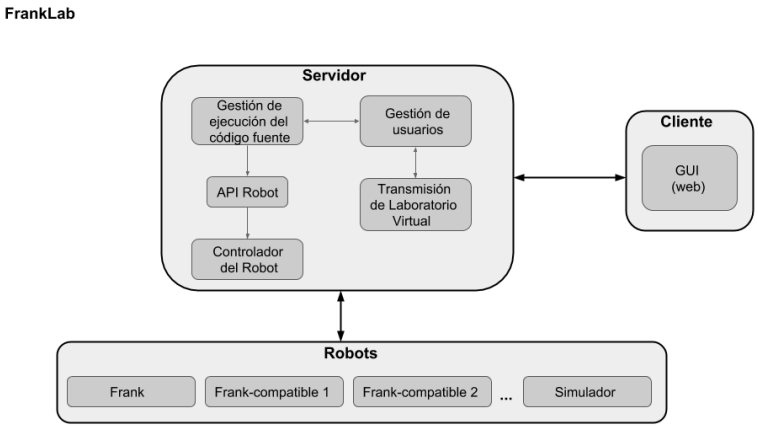


Figura 1. Estructura FrankLab

3.1. Estructura de la Plataforma

La plataforma, que se muestra en la Figura 1, está basada en el modelo de comunicación cliente-servidor. La estructura lógica del servidor está agrupada en cinco módulos, distribuidos en base a su funcionalidad:

- *Gestión de usuarios:* Este módulo aborda, por un lado, problemas en relación al modelo de datos: roles y privilegios de acceso de los distintos tipos de usuarios (estudiante, profesor y administrador) y administración de perfiles y archivos locales a cada uno de éstos. Por otro lado, comprende aspectos en relación a la edición de código fuente y a la colaboración que puede existir entre un grupo de alumnos sobre el mismo programa. Asimismo, cubre aspectos sobre la administración de los turnos que dispone cada usuario, con el objetivo de lograr un uso equitativo de los recursos físicos.
- *Transmisión de Laboratorio Virtual:* Este módulo es el encargado del proceso de entrega de los datos de la visualización en vivo, capturada por la cámara del laboratorio remoto hacia el navegador del cliente.

- *Gestión de ejecución del código fuente:* Este módulo es el encargado de realizar la compilación o interpretación del código fuente, para los lenguajes que así lo requieran.
- *API del robot:* Un lenguaje común para todos los robots físicos o simulados es definido y utilizado para controlarlos, de forma de proveer una capa de abstracción entre el lenguaje de programación elegido y la implementación real. Este módulo es el encargado de la traducción del comando de alto nivel al conjunto de instrucciones de bajo nivel que lo implementan.
- *Controlador del robot:* Este módulo gestiona las rutinas de bajo nivel que dan soporte al manejo y comunicación de robots físicos o simulados. Ejemplos de ellas son, el análisis de las imágenes capturadas por la cámara del robot y la transmisión y recepción de datos hacia y desde el mismo, respectivamente.

En lo que respecta a la interacción del cliente con el servidor, inicialmente el usuario debe acceder mediante su navegador a la página principal de FrankLab e iniciar una sesión identificándose.

Una vez validada su identidad, tendrá acceso a la selección del modo de programación a utilizar, esto es: tangible, basada en bloques o basada en texto. Dependiendo de su elección, se despliegan los lenguajes soportados para el modo elegido; por ejemplo, Python y C++ en el caso de la programación textual, e inicialmente Blockly cuando se trata del enfoque de programación por bloques.

Concretada la selección, se provee al usuario del respectivo entorno de codificación para la edición de sus programas. Al ser éstos ejecutados, el usuario puede visualizar en directo lo que la cámara principal del robot y del laboratorio remoto están capturando en ese momento, es decir, la vista en primera y tercera persona de los movimientos del robot, basados en la ejecución de su programa.

En el caso de no haber disponibilidad de robots para su utilización, se puede en forma alternativa visualizar la ejecución del programa en un ambiente simulado. Por último, dado que el usuario puede proveer y recibir feedback hacia y desde el robot, se dispone de una sección para ingresar/recibir datos de entrada/salida e informar acerca del estado actual de la ejecución.

3.2. Robots Físicos

Frankestito es una familia de robots físicos diseñados y construidos como hardware y software libre en la Facultad de Informática. Actualmente, el modelo más utilizado en las actividades de programación con robots es el N6-Frankestito[4,6,10].

Frankestito y N6-Frankestito son robots educativos con visión. Físicamente están compuestos de un chasis con dos motores continuos y una tercera rueda de movimientos libres. La potencia de los motores está gestionada por un microcontrolador más un integrado para tal fin. También, sobre el chasis, se encuentra empotrado una computadora de una única placa (del inglés SBC: Single Board Computer) con un microprocesador MIPS, interfaz usb (en donde está conectada una cámara de video), interfaz Wi-Fi (para comunicaciones TCP/IP)

y un dispositivo UART para la comunicación serie entre los dos procesadores (microcontrolador y microprcesador MIPS).

El software está compuesto de un firmware que se ejecuta en el microcontrolador, y una versión de Linux embebida en la SBC. El firmware interpreta el nivel mas bajo del protocolo de MyRO-Scribbler , y controla los motores en base a los paquetes enviados y recibidos mediante la interfaz UART. En el Linux embebido existe, por otro lado, un conjunto de programas que se encarga de la captura de video de la cámara, y de la recepción de comandos y envío de información a los programas clientes. De esta manera, un programa cliente puede, a través de una comunicación Wi-Fi, controlar los motores del robot, y sensar del ambiente mediante mensajes sencillos definidos en el protocolo.

3.3. Modos de Programación Soportados

Con la intención de hacer de la programación de robots una experiencia posible para más estudiantes, FrankLab soporta diferentes modos de programación que la hacen accesible a niños y adolescentes desde el nivel inicial hasta después de la Escuela Secundaria. La Figura 2 muestra los modos de programación soportados por FrankLab en relación a los niveles del sistema educativos.

La categoría Modos de Programación como forma de organización teórico conceptual y su relación con diferentes momentos de la escolarización obligatoria corresponden a construcciones teóricas elaboradas por los autores de este documento.

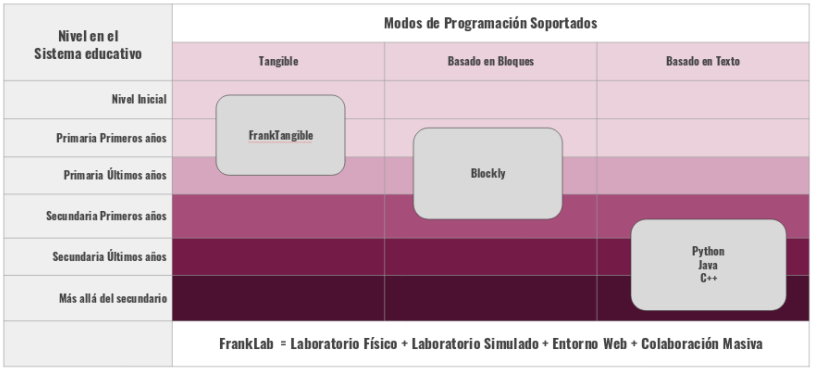


Figura 2. Modos de Programación soportados por FrankLab

Programación Tangible La Programación Tangible que incorpora el entorno *FrankLab*, permite programar robots educativos compatibles con *Frankestito* mediante la manipulación directa de objetos físicos, como estrategia para reducir el

esfuerzo cognitivo y la habilidades previas requeridas para aprender conceptos fundamentales sobre Algoritmos y Programación.

Este enfoque permite que Frankestito llegue, en los procesos de enseñanza de la programación, a una población más amplia de estudiantes, incluyendo aquellos no alfabetizados. Es necesario destacar que, al trabajar con objetos físicos nos brinda la posibilidad de *integración* en las clases de programación, es decir, unificar la herramienta de trabajo para todos los estudiantes sin distinción.

En este ambiente se trabaja con una colección de piezas tangibles, como bloques de madera o trozos de cartulina, que se secuencian para construir programas. Cada pieza corresponde a un comando o estructura de control, de forma similar a lenguajes basados en bloque.

A diferencia de otros entornos donde se trabaja con programación física, estas piezas no utilizan ningún tipo de hardware, permitiendo que el factor económico no sea un impedimento para la enseñanza de Ciencias de la Computación.

De este modo, se trabaja inicialmente con los objetos hasta lograr la secuencia buscada, luego se requiere tomar una fotografía al trabajo realizado para ser enviado al robot a través de FrankLab y poder observar los resultados del programa creado.

Para que esto sea posible, es necesario definir un lenguaje de programación tangible que permita representar las principales funcionalidades de Frankestito y las estructuras de control básicas. Además, se considera el desarrollo de un *parser* que traduzca especificaciones en este lenguaje a uno de los soportados por los robots educativos compatibles con Frankestito.

En este sentido, en la plataforma FrankLab se desarrolla un apartado que incluya el lenguaje con objetos físicos, el parser para producir un código ejecutable a partir de esos objetos, el medio en que esos programas se ejecutan, en nuestro caso los robots físicos.

Teniendo en cuenta la estructura planteada, en la Figura 1, de FrankLab, los siguientes módulos ejecutarán acciones específicas respecto a este modo de programación:

- ***Cliente:*** Luego de que los estudiantes definen la secuencia de figuras, es necesario transferir el código generado al robot que lo ejecute. Para ello se toma una fotografía del programa construido con objetos físicos, que es subida al la plataforma FrankLab en la sección de *programación tangible*.
- ***Gestión de Ejecución del Código Fuente***: En este módulo la imagen es procesada para identificar las distintas funcionalidades y estructuras utilizadas en el programa. Luego mediante el *parser* se traducirá del lenguaje con objetos físicos a un lenguaje común para todos los robots compatibles con Frankestito.

Programación Basada en Bloques La programación basada en bloques, referencia a la posibilidad de generar código comprensible por *Frankestito* mediante una gramática basada en “bloques de programación”, gráficos apilables e interactivos. Los mismos permiten generar programas sin recurrir a un lenguaje

de programación basado en texto, lo que reduce la curva de aprendizaje en la utilización de los robots educativos.

La idea es dotar a *FrankLab*, con bloques de programación que representen todas las funciones con las que cuenta *Frankestito*, a fin de ser utilizados dentro del entorno para controlar el robot. De este modo, se logra un entorno de programación sencillo y transparente con comunicación efectiva y simplificada.

Cada bloque tiene una estructura diferenciada por tipo de instrucción, permitiendo que al apilarlos produzca porciones de código sintácticamente correctas, facilitando el aprendizaje de las estructuras de programación de forma gráfica y explícita.

Además, estos bloques concentran toda la funcionalidad provista por la arquitectura de los robots, por lo que la expresividad del lenguaje representado en los bloques es análoga a cualquier otro lenguaje de la plataforma, permitiendo que los proyectos puedan ser interpretados o traducidos si el usuario cambia de modo o lenguaje de programación.

En la Figura 1 se puede observar como esta funcionalidad se encuentra distribuida entre dos módulos bien diferenciados:

- ***Cliente***: Del lado del *Cliente*, se encuentra la ejecución de las librerías que proveen los bloques gráficos al navegador Web, permitiendo la utilización de los mismos en el ambiente de programación.
- ***Gestión de Ejecución del Código Fuente***: una vez que el usuario decide *ejecutar* el código desde su navegador, el mismo es traducido a un código equivalente en un lenguaje de alto nivel basado en texto. Asimismo, este módulo es quién iteractúa con el código traducido desde los bloques hacia el lenguaje intermedio del robot.

Es necesario aclarar la separación entre los bloques gráficos de programación y el lenguaje representado por los mismos. Los *bloques gráficos* son una implementación propia en base a las librerías del proyecto de código libre de programación visual *Blockly*, que provee una gramática de componente visual para facilitar la codificación de diversos lenguajes de programación[5]. En la Figura 3 se muestra un ejemplo de bloques de programación para Frankestito desarrollada en el ambiente *Blockly*.



Figura 3. Ejemplo de bloques de programación para *Frankestito*

Por otro lado, el lenguaje que representan y que, en última instancia, es generado por el módulo de *Gestión de Ejecución del Código Fuente* de la plataforma es propio de la arquitectura *Frankestito* y corresponde propiamente al robot[4].

Programación basada en texto La plataforma FrankLab ofrece a los estudiantes la posibilidad de seleccionar el lenguaje de programación basado en texto de su preferencia, entre los soportados por Frankestito, para codificar sus programas. Es posible interactuar con el robot a través de una diversidad de lenguajes de programación como Python, C y C++, entre otros.

La implementación incorpora un entorno de desarrollo integrado en línea, que consiste en un editor de textos para programar, al que se accede a través de un navegador Web. Esta opción permite editar código fuente sin la necesidad de instalar un IDE en la máquina del usuario, reduce los tiempos de instalación y minimiza los requerimientos de hardware. Por otro lado, permite programar desde cualquier lugar y dispositivo que cuente con un navegador Web instalado.

- ***Cliente***: el editor de código en línea permite a los estudiantes seleccionar el lenguaje de programación y codificar sus programas. Integra las características habituales de un entorno de desarrollo integrado en línea, para facilitar el proceso de codificación, como por ejemplo por ejemplo, coloreado de palabras reservadas, reconocimiento de sintaxis y autocompletado de código. En conjunto estas funcionalidades contribuyen a reducir la curva de aprendizaje.

- ***Gestión de Ejecución del Código Fuente***: En este módulo se gestiona la compilación o interpretación del código. Asimismo, se espera que en casos de detectar errores del tipo léxico o sintáctico, se provea asistencia al usuario mediante la descripción del error encontrado.

La Plataforma Web de Robótica Educativa presentada en este trabajo permite ir más allá de la resolución de problemas clásicos con el objeto de enseñar a programar. El *techo* dado por la enseñanza de la programación abordando problemas simples, sino que es posible atacar problemas de mayor complejidad considerados en la categoría de la investigación o la innovación. Ejemplos de problemas en esta clase son, entre otros:

- Procesamiento de imagen: actualmente los robots compatibles con Frankestito pueden detectar colores, su cantidad y posición. Una mejora podría ser la detección de figuras geométricas o de rostros en una imagen.
- SLAM (Simultaneous Localization And Mapping): los robots navegan en ambientes desconocidos. La construcción y actualización de un mapa del sector, mientras que realiza una estimación de su ubicación y trayectoria, permitiría ampliar el conjunto de aplicaciones para los robots compatibles con Frankestito como vehículos autónomos.
- Sistemas Multiagentes: las aplicaciones que involucran un conjunto de agentes coordinados para lograr un objetivo común, requieren de comunicación sincronizada entre ellos. Ejemplo de esto, es la búsqueda de sobrevivientes en una catástrofe por grupos de robots autónomos.

4. Conclusiones

En este trabajo se ha presentado FrankLab, una plataforma Web de robótica educativa delineada como una solución integradora a características detec-

tadas en la enseñanza de Ciencias de la Computación, como por ejemplo, la singularidad del aprendiz, los lenguajes de programación adecuados para lograr competencias particulares y la escasa disponibilidad de los recursos físicos, entre otras.

Su arquitectura basada en el modelo cliente-servidor y estructurada en forma modular, posibilita el soporte de diferentes modos de programación (tangible, por bloques y basada en texto) y de diversos lenguajes de programación. Así, de acuerdo a su diseño, la plataforma que fue proyectada con el fin de mejorar las posibilidades de aprender Ciencias de la Computación, permite ampliar la población estudiantil destinataria, ofreciendo las herramientas adecuadas para tal fin.

Actualmente, está en desarrollo el primer prototipo de la plataforma con licencia open source. Se espera validar a FrankLab en actividades de extensión que se llevan a cabo con estudiantes de diferentes niveles educativos.

Referencias

1. P. Blikstein, A. Sipitakiat, J. Goldstein, J. Wilbert, M. Johnson, S. Vranakis, Z. Pedersen, and W. Carey. Project bloks: designing a development platform for tangible programming for children. *Position paper, retrieved online on*, pages 06–30, 2016.
2. K. Brennan and M. Resnick. New frameworks for studying and assessing the development of computational thinking. In *In AERA 2012*, 2012.
3. K.-. C. S. F. S. Committee et al. K-12 computer science framework. 2016.
4. E. Grosclaude, R. Zurita, J. Riquelme, R. del Castillo, and M. Lechner. Designing A Myro-Compatible Robot For Education As Copyleft Hardware. In *CACIC 2014*, pages 372 – 382. UNLAM, 2014.
5. E. Pasternak, R. Fenichel, and A. N. Marshall. Tips for creating a block language with blockly. In *2017 IEEE Blocks and Beyond Workshop (B B)*, pages 21–24, Oct 2017.
6. J. Rodríguez, G. Grosso, R. Zurita, and L. Cecchi. Intervención de la facultad de informática en la enseñanza de ciencias de la computación en la escuela media basada en robótica educativa. In *XI Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2016)*, 2016.
7. T. Sapounidis and S. Demetriadis. Educational robots driven by tangible programming languages: A review on the field. In *International Conference EduRobotics 2016*, pages 205–214. Springer, 2016.
8. T. Sapounidis, I. Stamelos, and S. Demetriadis. Tangible user interfaces for programming and education: A new field for innovation and entrepreneurship. In *Innovation and Entrepreneurship in Education*, pages 271–295. Emerald Group Publishing Limited, 2016.
9. M. Smith. Computer science for all. *The White House*, 2016.
10. R. Zurita, J. d. l. Fuente, M. Bucarey, D. Bonet, R. d. Castillo, G. Grosso, L. Cecchi, J. Rodríguez, et al. Mejorando las posibilidades de aprender a programar, ampliación del robot educativo multiplo n6 max a frankestito. In *XII Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET, La Matanza 2017).*, 2017.

Ciencias de la Computación y escuelas ¿una didáctica específica?

Claudia Queiruga[1], Claudia Banchoff Tzancoff[1], Paula Venosa[1], Soledad Gómez[1], Glenda Morandi[2]

[1] LINTI. Facultad de Informática. Universidad Nacional de La Plata. La Plata, Argentina
{claudiaq, cbanchoff, pvenosa, sgomez}@info.unlp.edu.ar
[1]Facultad de Periodismo y Comunicación Social. Universidad Nacional de La Plata. La Plata, Argentina
glenda.morandi@presi.unlp.edu.ar

Abstract. Este trabajo presenta la propuesta curricular de la "Especialización docente en didáctica de las Ciencias de la Computación" de la provincia de Buenos Aires, Argentina. La misma está enfocada en la formación docente de nivel secundario en el campo disciplinar de las Ciencias de la Computación o Informática. Su relevancia se caracteriza por el déficit de espacios de formación destinados a esta temática en los niveles obligatorios de la escolaridad. El proyecto surge en el marco de una convocatoria a Universidades Nacionales con carreras de Informática o Ciencias de la Computación, impulsada por la Fundación Sadosky perteneciente al Ministerio de Ciencia, Tecnología e Innovación de la Nación. Una de las características destacadas es el reconocimiento de los Institutos Superiores de Formación Docente, como referentes por excelencia de formación docente continua. La articulación entre las instituciones participantes promovió, como resultado, la elaboración de un diseño curricular y el dictado de una carrera que recupera contenidos de Ciencias de la Computación desde un enfoque de formación docente.

Keywords: Informática, Ciencias de la Computación, escuela secundaria, enseñanza de programación, pensamiento computacional.

1 Introducción

La inclusión de la enseñanza de la Informática en los sistemas formales educativos de niveles obligatorios es un hecho actual y novedoso en la Argentina y el mundo. Existen múltiples motivos y condiciones que ameritan hacer lugar al mismo, por un lado, las enormes oportunidades que brindan las tecnologías y el "software" como motor de desarrollo económico y social y, por otro, en un mundo cada vez más conectado, en el que la computación es ubicua, se torna esencial que los ciudadanos comprendan sus elementos constitutivos, y sean capaces de adaptarlos y modificarlos a sus necesidades específicas.

La resolución 263 de la Asamblea del Consejo Federal de Educación del año 2015 de la República Argentina, declaró la importancia estratégica para el sistema educativo argentino la enseñanza y el aprendizaje de la programación durante la escolaridad obligatoria, para fortalecer el desarrollo económico-social de la Nación. En forma simultánea, en los últimos años, un conjunto de iniciativas y políticas federales, han contribuido a consolidar una perspectiva que fortalece las razones por las cuales resulta necesario situar la enseñanza de la Informática como campo disciplinar. Ejemplo de ello es el surgimiento de políticas educativas, tales como: el proyecto "Program.AR" [1], la creación del "Programa Conectar Igualdad", el "Plan Nacional Integral de Educación Digital" (PLANIED) [2], entre otros, que permiten situar social e históricamente, la necesidad de pensar en clave educativa el desarrollo, avance y sinergia en este campo. Más recientemente, en 2017, el proyecto "Secundaria 2030" [3] aprobado por el Consejo Federal de Educación propone incorporar gradualmente un enfoque de enseñanza basado en capacidades y competencias digitales transversales.

La declaración de su estratégica incorporación a la educación formal obligatoria argentina, torna necesario trabajar de manera sostenida en la formación de docentes en condiciones de liderar los procesos de aprendizaje de la Informática. Actualmente, la formación docente en este campo disciplinar en la provincia de Buenos Aires, constituye un área de relativa vacancia. Si bien se identifican algunos Profesorados Universitarios en Informática y/o afines en algunas universidades del territorio bonaerense, tales como la UNICEN (Universidad Nacional del Centro de la Provincia de Buenos Aires) y la UBA (Universidad de Buenos Aires), resulta estratégico situar un Postítulo en enseñanza de las Ciencias de la Computación, que articule con el sistema de Educación Superior de la provincia y profundice procesos de articulación entre las Universidades Nacionales y los Institutos Superiores de Formación Docente (ISFD) de la jurisdicción provincial. A su vez, la emergencia del tema en las agendas educativas públicas se caracteriza por las escasas referencias en las que basarse para organizar los contenidos y la metodología sobre cómo enseñarlos. Es por ello que es fundamental formar a los docentes en ambos aspectos y de manera conjunta por equipos interdisciplinarios expertos en la disciplina Informática y en la didáctica.

En respuesta a atender la necesidad planteada, Program.ar en el año 2016 realizó una convocatoria a universidades nacionales con carreras de Informática o Ciencias de la Computación que tuvieran interés en diseñar y dictar una especialización en formación docente en asociación con los ISFD de sus jurisdicciones. De las universidades que se postularon, resultaron seleccionadas 8 en base a la calidad de las propuestas y la trayectoria profesional de sus equipos. En la provincia de Buenos Aires resultaron elegidas las UNLa (Universidad Nacional de Lanús) en asociación con el ISFDyT Nº 24 "Dr. Bernardo Houssay", la UNLP (Universidad Nacional de La Plata) en asociación con Escuela Normal Superior Nº1, "Mary O. Graham"- ISFDNº 95 y la UNICEN en asociación con el ISFDyT Nº 166 "José de San Martín". La provincia de Buenos Aires contará con 3 propuestas orientadas a destinatarios distintos: una formación sobre la disciplina y su didáctica orientada a docentes de nivel primario (UNICEN - ISFDyT Nº 166), una orientada a docentes de nivel secundario con formación de base en la disciplina o afines como matemática, física,

química y tecnología (UNLP - ISFD Nº 95) y una última orientada a docentes de nivel secundario con formación en otras asignaturas, prioritariamente en ciencias sociales (UNLa-ISFDyT Nº 24). Cada una de estas ofertas propone un recorte de contenidos y una didáctica particular para su abordaje según los destinatarios de la formación así como los beneficiarios últimos -estudiantes- de cada oferta.

Las secciones restantes de este trabajo están organizadas de la siguiente manera. En la sección II se analiza el campo disciplinar de la Informática en la escuela, la sección III describe el diseño curricular de la "Especialización docente en didáctica de las Ciencias de la Computación" elaborado por las autoras de este trabajo. La sección IV presenta la situación actual en relación a la implementación de la primera cohorte de la Especialización y las conclusiones se presentan en la sección V.

2 La relevancia del campo disciplinar de la Informática en la escuela

Actualmente nuestro cotidiano está rodeado de objetos tecnológicos, desde el celular que nos acompaña todo el tiempo, semáforos y casas inteligentes, pasando por drones que nos sobrevuelan, ropa tecnológica, entre otros elementos que cada vez son más evidentes en nuestra vida diaria. Estos elementos modifican nuestros hábitos, nuestra forma de relacionarnos, de entretenernos, de estudiar, aumentan nuestras capacidades cognitivas y es por ello que es necesario entender qué pasa allí adentro. Esta realidad plantea un desafío en el campo educativo: la formación de ciudadanos/as que puedan conocer, entender y operar sobre el mundo que los rodea, incluido su entorno digital, ubicándolos como sujetos críticos y creadores de innovaciones con tecnologías digitales, por sobre la pasividad y el mero consumo tecnológico. Es por ello que se torna necesario que los saberes Informáticos se incorporen en los diseños escolares de los sistemas educativos nacionales.

La enseñanza de Informática es actualmente una preocupación global, países como Nueva Zelanda, Estonia, Japón, Finlandia y Reino Unido, entre otros, han actualizado sus currículos escolares incluyendo la enseñanza de la programación en las escuelas [4][5]. Varios estados de los Estados Unidos también han implementado políticas activas en respuesta al respaldo de la industria tecnológica mediante el movimiento "LearnToCode", liderado fundamentalmente por las iniciativas globales code.org y codeacademy. En Argentina, a lo largo de las últimas décadas, el uso de TIC se ha ido incorporando a las prácticas educativas en los distintos niveles de la escolaridad obligatoria. Algunos ejemplos, en la provincia de Buenos Aires, son el nuevo diseño curricular de la educación primaria [6], el cual contiene un módulo sobre la inclusión de TIC que intenta incorporar en forma transversal el uso de tecnologías digitales en las distintas áreas curriculares y, el espacio curricular NTIC [7] de la escuela secundaria donde algunos conceptos relacionados a la disciplina Informática se intentan incluir. Sin embargo, la Informática como disciplina aún no ha sido legitimada en el ámbito de la educación escolar, aunque hay algunas experiencias donde esto comienza a transformarse, aún no se ha llegado a los consensos necesarios

sobre la inclusión del campo en las currículos escolares a nivel federal. Las escuelas de educación secundaria técnica son las únicas que cuentan con trayectos formativos en Informática. De esta manera el estudio sobre la disciplina no se da en forma transversal, ni forma parte integral de los contenidos a los que acceden la mayoría de los estudiantes en su paso por la escuela. Incorporar en los diseños escolares la disciplina Informática constituye un logro y un espacio de seguridad para el desarrollo de procesos cognitivos vinculados al razonamiento lógico que permite predecir, analizar y explicar, a la formulación de algoritmos, a la descomposición de problemas en partes más simples, a la abstracción para manejar la complejidad, a la generalización mediante el descubrimiento de patrones y similitudes y, a la evaluación. Estas ideas, que identifican al "pensamiento computacional"[8][9] son ampliamente aplicadas en la resolución de problemas usando computadoras y en la comprensión de los sistemas más allá del espacio escolar. A su vez, la enseñanza y el aprendizaje de "programación" es un aspecto de la Informática que en la escuela se percibe como el más desafiante. Es por ello que se torna necesario que los docentes se formen en los conceptos, en las prácticas y en los enfoques del pensamiento computacional, para que éstos resulten accesibles y atractivos para los estudiantes y fundamentalmente les permitan apropiarse de ellos para entender cómo funciona el mundo digital y cómo usar este conocimiento para programar.

El diseño curricular que aquí se presenta recupera la relevancia de la formación docente en la enseñanza y aprendizaje de la Informática con especial atención en la "programación". La intencionalidad transversal que orienta al abordaje de la "programación" está orientada a desarrollar un trabajo con los docentes que los coloque en la situación de involucrarse en una práctica de la programación que reflexione en torno de cada uno de los procesos cognitivos que supone este desarrollo. Asimismo, se propone un enfoque situado y no abstracto de la programación, interpelando los problemas de la vida cotidiana que esta práctica ayuda a resolver y mejorar. La atención está puesta en aprovechar las posibilidades que brinda la programación en relación con sus usos creativos, promoviendo que los sujetos superen la condición únicamente de consumidores de software.

3 Especialización docente en didáctica de las Ciencias de la Computación

3.1 La propuesta

La Especialización está dirigida a docentes y profesionales en el ejercicio de la enseñanza de Informática en la escuela secundaria. Abarca tanto a los docentes que dictan asignaturas del área de Tecnologías e Informática como a los que trabajan en las áreas de Matemática, Física y Química. Estos docentes tienen una formación inicial en temas como lógica proposicional, modelos y abstracción, resolución de problemas, uso de lenguaje científico; requerido como formación de base para la Especialización. Los primeros temas a abordar corresponden a los aspectos políticos

y pedagógicos que traen aparejados la enseñanza de la Informática. En este marco se trabajan temas tales como las potencialidades y alcances de la enseñanza de esta disciplina en el contexto socio-histórico actual atravesado por el desarrollo tecnológico y la innovación en este campo, destacando el análisis del concepto de soberanía tecnológica y sus relaciones con las ideas detrás del software libre y el software privativo. Estas temáticas, junto con el análisis de los aspectos sociales, políticos y éticos vinculados con la tecnología y su uso en la vida cotidiana y las buenas prácticas en el uso seguro y responsable de las redes de datos y de las tecnologías digitales en general, conforman las bases relacionadas a la formación en ciudadanía digital.

Como se mencionó en la sección previa, el contenido de la Especialización pone especial atención en la enseñanza y aprendizaje de la programación de computadoras, es por ello que una gran cantidad de horas están dedicadas a este tema y el sentido del mismo es transversal a todos los módulos. Se parte de los aspectos más básicos como ser la comprensión y análisis de algoritmos, finalizando con el desarrollo de un proyecto de software utilizando distintas herramientas informáticas, algunas de ellas basadas en la programación en bloques y otras que utilizan lenguajes de programación textuales. En este sentido se plantean cuatro módulos de complejidad creciente en los cuales se trabajan tanto los aspectos básicos de la programación como ser secuencias, estructuras de control y abstracción, como conceptos de lógica proposicional y desarrollos de software libre. Durante el primer año de la Especialización se desarrollan también aspectos relacionados al funcionamiento interno de las computadoras y al rol de los sistemas operativos, y en el segundo año se trabajan los conceptos fundamentales sobre redes de datos. Estas temáticas se abordan no sólo desde el punto de vista teórico de su funcionamiento sino con actividades prácticas que promueven el uso de los recursos disponibles en las escuelas: desde las salas de computadoras, robots educativos, netbooks y/o celulares inteligentes y tablets.

El abordaje de estos temas aplicados en la realización de dos prácticas situadas, permitirán formar egresados que puedan diseñar, coordinar y evaluar situaciones didácticas innovadoras centradas en el desarrollo del pensamiento computacional e incorporar en las prácticas de enseñanza un enfoque sobre la programación, que recupere estrategias que favorezcan procesos de creación de aplicaciones y contenidos digitales.

Este enfoque promueve el intercambio con docentes de diferentes áreas disciplinares de la curricula escolar, en la búsqueda de la vinculación de estrategias del pensamiento computacional con los procesos de enseñanza y aprendizaje de las mismas.

3.2 La articulación con el ISFD

La propuesta de la Especialización reconoce a los ISFD como referentes por excelencia de formación docente continua y a las universidades nacionales con carreras de Informática o Ciencias de la Computación como las generadoras de conocimiento específico del campo. Es por ello que la Especialización docente que

aquí se presenta responde a la necesidad de crear dispositivos sustentables de formación docente poniendo en relación ambas instituciones. Asimismo el desarrollo de la Especialización fue concebido en términos de adecuación a las necesidades y posibilidades de la jurisdicción en la que está suscripta. En el caso de nuestra propuesta fue desarrollada en asociación con la Escuela Normal Superior Nº1,"Mary O. Graham"- ISFD Nº95 y está orientada a docentes de nivel secundario con formación de base en la disciplina o asignaturas afines como matemática, física y química.

3.3 La metodología de enseñanza

La puesta en marcha del desarrollo de la caja curricular de la Especialización obligó al equipo de trabajo a repensar las formas sobre la enseñanza que cada uno desarrolla a diario, en nuestro rol como docentes. En este sentido la interdisciplinariedad acompañó el proceso, dado que el equipo está compuesto por profesionales de la Informática, las Ciencias de la Educación y de la Comunicación. El proyecto propone enseñar desde una didáctica específica los contenidos de las Ciencias de la Computación, destinados a docentes en ejercicio del nivel secundario. Estos docentes pueden o no tener formación en Informática y pueden o no tener formación pedagógica en el campo, lo cual se presentó como un gran desafío. Nuestra propuesta pretende enseñar a enseñar algo potencialmente desconocido y quienes lo conocen pueden no contar con las herramientas para enseñarlo. La decisión se basó en el consenso del equipo de trabajo que pretende romper con algunas lógicas de trabajo áulico que no fortalecen el desarrollo del pensamiento computacional, sino que más bien promuevan prácticas de trabajo propicias para el desarrollo del conocimiento tecnológico desde un sentido puramente instrumental. El aprendizaje basado en problemas y metodología de trabajo por indagación [10] resultaron opciones coherentes para potenciar el trabajo con tecnologías y promover prácticas de producción con ellas.

La metodología de trabajo que propone el aprendizaje basado en problemas (ABP), muy vinculado a las nuevas perspectivas sobre la enseñanza de las Ciencias de la Computación [11], resuelve la problemática de la enseñanza con ciertas dificultades, en tanto nos preocupa el desconocimiento sobre las temáticas propuestas. En este sentido, para el equipo, esta forma de trabajo áulico obligó nuevamente a repensar las formas en que se produce el conocimiento y los procesos cognitivos que pretende ayudar a producir. Los docentes, en la mayoría de los casos, se encontrarán con el desafío de conocer la tecnología y sus posibilidades, sin embargo la preocupación del equipo de trabajo radica en las posibilidades de resignificarla y ponerla en práctica. Desde esta perspectiva pareciera que la preocupación del equipo de trabajo es el mero uso y aplicación, sin embargo se parte de creer que en los estadios de un proceso crítico de apropiación de conocimiento es necesario, primero conocer para luego transformar en praxis [12]. Volviendo a la metodología ABP, su incorporación en las diferentes propuestas del plan de estudios estuvo ligada al recorrido que viven los estudiantes desde el planteamiento original del problema hasta su solución. El trabajo

colaborativo en pequeños grupos, permite compartir en esa experiencia de aprendizaje la posibilidad de practicar y desarrollar habilidades, de observar y reflexionar sobre actitudes y valores que en el método convencional expositivo difícilmente podrían ponerse en acción. En este sentido, coincidimos con Guevara Mora [13] cuando explica que "la experiencia de trabajo en el pequeño grupo orientado a la solución del problema es una de las características distintivas del ABP. En estas actividades grupales los alumnos toman responsabilidades y acciones que son básicas en su proceso formativo".

La indagación, práctica central y constitutiva del ABP, se vincula fuertemente con el desarrollo tecnológico en tanto se ve favorecida cada vez más con el acceso a la información y la era de los flujos. Los estudiantes se reconocen en estas prácticas pues son constitutivas de sus usos tecnológicos cotidianos, que fortalecen sus capacidades de indagar, investigar y crear. Es importante, comprender que esta capacidad de creación y descubrimiento definida como indagación es la base para el desarrollo de procesos de pensamiento crítico y significativo. Comprender esto de forma amplia permite concebir a la tecnología como una posibilidad y una herramienta fundamental, no solo por el actual contexto que nos rodea y condiciona, sino porque el acceso al conocimiento y las redes que genera. Poder pensar la forma en qué, para qué y por qué enseñar Ciencias de la Computación constituye un desafío diario para este equipo de trabajo, el cual considera que las tecnologías nos atraviesan y condicionan nuestras formas de vivir, por lo tanto su abordaje en los espacios educativos formales tiene que estar asegurado, no solo en la currícula sino desde un enfoque que promueva, crear con tecnologías.

3.3 La organización curricular

La carrera se estructura en diez módulos, con una carga horaria de 320 horas de módulos disciplinares y 80 horas de práctica docente situada, distribuidos en dos semestres en cada año de cursado. Los módulos que componen el plan de estudios se articulan en torno de tres áreas de conocimiento, que posibilitan agrupar objetos de enseñanza específicos de las Ciencias de la Computación, y los desarrolla en forma gradual en niveles de complejidad creciente, partiendo de una secuencia que se articula con los procesos de aprendizaje que se espera desarrollen los estudiantes de las escuelas secundarias.

Las áreas de conocimiento delimitadas son:

- Ciencias de la Computación, ciudadanía digital y educación, que incluye los módulos "Marco político pedagógico de la enseñanza de las Ciencias de la Computación" y "Uso seguro y responsable de la tecnología y de los servicios de Internet".
- Arquitectura, sistemas operativos y redes en la enseñanza secundaria, que incluye los módulos "Las computadoras y los sistemas operativos", "Redes de datos e Internet" y "Práctica Situada I".
- La Programación y su enseñanza, que incluye los módulos "El algoritmo como objeto de aprendizaje y forma de organización del pensamiento", "La

enseñanza de la programación a través de lenguajes visuales", "La enseñanza de programación en lenguaje real de producción de software con especial atención al software libre", "El proceso de enseñanza del desarrollo de un proyecto de software" y "Práctica Profesional Situada II".

En esta organización curricular se destaca un eje de formación práctica supervisada -Práctica Situada I y Práctica Situada II-, que se desarrolla en dos espacios curriculares diferentes, en cada uno de las cuales los cursantes elaboran un proyecto o diseño de enseñanza en el que deben incluir las estrategias y contenidos abordados en la carrera. La Tabla I resume la organización de la carrera en 2 años.

Tabla I. Organización de la carrera en años

Primer Año			
Marco político pedagógico de la enseñanza de las Ciencias de la Computación (30 hs.)	El algoritmo como objeto de aprendizaje y forma de organización del pensamiento (30 hs.)	Las computadoras y los sistemas operativos (40 hs.)	La enseñanza de la programación a través de lenguajes visuales basados en bloques (50 hs.)
Práctica Profesional Situada I (30 hs.)			
Segundo Año			
Redes de datos e Internet (40 hs.)	La enseñanza de programación en lenguaje real de producción de software con especial atención al "Software" libre. (50 hs.)	El proceso de enseñanza del desarrollo de un proyecto de "Software". (40 hs.)	Uso seguro y responsable de la tecnología y de los servicios de "Internet". (40 hs.)
Práctica Profesional Situada II (50 hs.)			

4 La implementación de la especialización

La Especialización aquí presentada ha sido aprobada por el Consejo General de Cultura y Educación de la provincia de Buenos Aires en marzo de 2018 y cuenta con despacho resolutivo del Director General de Cultura y Educación de la provincia. Esta resolución establece el inicio de su implementación en el año 2018, a ciclo cerrado, y la provisión de puntaje a los docentes que realicen el Postítulo. En este sentido, se encuentra actualmente en proceso de implementación, para ello se está considerando una buena relación docente-alumno en cada uno de los módulos, dado que las actividades se centrarán mayormente en metodologías de trabajo basadas en taller. Asimismo se pondrá atención en el seguimiento del desarrollo de los módulos, para

evitar el desgranamiento y abandono de los cursantes, propiciando la terminalidad de la formación. Es intención que esta política de formación federal pueda accionar lo más pronto posible en la formación de nuestros adolescentes en su paso por la escuela secundaria. El cuerpo docente de la Especialidad está conformado por especialistas en los temas del campo disciplinar Informático que se abordan y docentes del área de comunicación y educación en los módulos que requieren una clara comprensión de las políticas que acompañan el desarrollo de la disciplina en el campo educativo.

5 Conclusiones

El desarrollo de esta Especialización es el fruto de un trabajo interdisciplinar en el cual un equipo integrado por docentes de la UNLP y del ISFD Nº 95 de La Plata coordinó esfuerzo y experiencia en un trabajo que cubre un área de vacancia en la formación docente.

Los planes y programas estatales, tales como Escuelas del Futuro del PLANIED, que entre sus iniciativas propone distribuir en las escuelas de Argentina recursos tecnológicos, requieren de una formación específica para promover aprendizajes significativos, ubicando a los estudiantes como creadores de artefactos digitales. Asimismo los docentes y directivos están interpelados por estos nuevos materiales y requieren de formaciones como la presentada en este trabajo que les permite organizar y planificar sus actividades con estas tecnologías que se suman al espacio áulico.

Resulta relevante recuperar que la “Especialización en didáctica de las Ciencias de la Computación” presentada en este trabajo es promovida desde un programa federal, el cual a su vez impulsó la implementación de otras siete especializaciones orientadas a la formación docente en Informática en el ámbito escolar, que se dictarán en otras provincias del país durante el año 2018. Esto habla de una perspectiva federal en relación a la enseñanza y aprendizaje de la Informática en la escolaridad obligatoria, resultando en un gran avance en relación la inclusión de Informática en el aula y también propone un punto de partida para la reformulación de los contenidos curriculares de escolaridad obligatoria en el sistema educativo argentino.

Referencias

1. Program.AR: programa de la Fundación Sadosky, Ministerio de Ciencia, Tecnología e Innovación Productiva de la Nación Argentina (s.f). Recuperado de: http://www.fundacionsadosky.org.ar/programas/programar/. Último acceso: 11/07/2018
2. PLANIED: Plan Integral de Educación Digital del Ministerio de Educación y Deportes de la Nación Argentina (s.f). Recuperado de: http://www.tic.siteal.iipe.unesco.org/normativa/2033/resolucion-1536-e2017-plan-nacional-integral-de-educacion-digital-planied
3. Secundaria 2030: Transformar la Secundaria para transformar vidas (s. f). Recuperado de: https://www.argentina.gob.ar/educacion/innovacionycalidadeducativa/2030. Último acceso: 11/07/2018

4. National curriculum in England: computing programmes of study (2013). Recuperado de: https://www.gov.uk/government/publications/national-curriculum-in-england-computing-programmes-of-study. Último acceso: 11/07/2018
5. Dirección General de Cultura y Educación de la Provincia de Buenos Aires (2018). Diseño curricular para la educación primaria: primer ciclo y segundo ciclo; coordinación general de Sergio Siciliano. - 1a ed. ISBN 978-987-676-095-9. Recuperado de: http://servicios.abc.gov.ar/lainstitucion/organismos/consejogeneral/disenioscurriculares/primaria/2018/dis-curricular-PBA-completo.pdf. Último acceso: 11/07/2018
6. The New Zealand Curriculum on-line (s.f). Disponible:http://nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum/Technology. Último acceso: 11/07/2018.
7. Dirección General de Cultura y Educación de la provincia de Buenos Aires (2010). Diseño Curricular para la Educación Secundaria Ciclo Superior. ES4: Nuevas Tecnologías de la Información y la Conectividad. Coordinado por Claudia Bracchi. -1a ed. ISBN 978-987-1266-98-2.
Recuperado de:
http://servicios.abc.gov.ar/lainstitucion/organismos/consejogeneral/disenioscurriculares/secundaria/materias_comunes_a_todas_las_orientaciones_de_4anio/tic_4_final_web.pdf.
Último acceso: 11/07/2018.
8. Wing, Jannette M. (2006). Computational Thinking. Communications of the ACM, vol. 49, 33-35.
9. Wing, Jannette M. (2008). Computational thinking and thinking about computing. Philosophical Transactions of The Royal Society A, vol. 366, 3717–3725.
10. Morales, Patricia, Landa, Victoria. Aprendizaje Basado en Problemas. Theoria [en línea] 2004, 13. Disponible en: http://www.redalyc.org/articulo.oa?id=29901314 ISSN 0717-196X. Último acceso: 11/07/2018
11. Echeveste, M.E. y Martínez, M.C. (2016). Desafíos en la enseñanza de Ciencias de la Computación. Virtualidad, Educación y Ciencia, 12 (7), pp. 34-48.
12. Freire P. Pedagogía de la autonomía: saberes necesarios para la práctica educativa. México: Siglo XXI; 2004.
13. Guevara Mora, Gabriela. Aprendizaje Basado en Problemas como Técnica Didáctica para la Enseñanza del Tema de la Recursividad. InterSedes: Revista de las Sedes Regionales [en línea] 2010, XI (Sin mes). Disponible en: http://ucsj.redalyc.org/articulo.oa?id=66619992009. ISSN 2215-2458. Último acceso: 11/07/2018

Workshop Seguridad Informática (WSI)

Coordinadores

Javier Diaz (UNLP) - Hugo Ramón (UNNOBA) - Claudio Aciti (UNCPBA)

Temas

Administración de Configuración de Software y Despliegue
Administración de Proyectos de Software
Arquitecturas y Diseño de Software
Calidad de Software
Dependabilidad y Confiabilidad del Software
Desarrollo Ágil de Software
Desarrollo de Software basado en Componentes
Entornos para el Desarrollo de Software
Evolución del Software
Estándares para el Desarrollo de Software
Herramientas de Software
Herramientas y Tecnologías de la Web 2.0
Ingeniería de Requerimientos de Software
Ingeniería de Software Distribuido
Ingeniería de Software Embebido y Tiempo Real
Ingeniería de Software Empírica
Ingeniería de Software para Sistemas Ubícuos y Móviles
Ingeniería de Software para la Web
Ingeniería Reversa
Mantenimiento de Software
Metodologías de Desarrollo de Software
Métricas de Software
Modelamiento de Software
Modelamiento de Procesos de Negocio
Orientación a Aspectos
Patrones y Frameworks
Plataformas como Servicios
Políticas para el Software y para las TICs
Privacidad
Procesos de Software
Reuso de Software
Sistemas de Información
Sistemas de Software basados en Cloud Computing
Sistemas de Software Sensibles al Contexto
Software como Servicio
Soporte Computacional para Tareas Colaborativas
Testeo de Software
Teoría y Métodos Formales
Verificación y Validación de Software
Workflows

Application of Keystroke Dynamics Modelling Techniques to Strengthen the User Identification in the Context of E-commerce

Germán M. Concilio, Enrique P. Calot,
Jorge S. Ierache, and Hernán D. Merlino

Laboratorio de Sistemas de Información Avanzados,
Facultad de Ingeniería, Universidad de Buenos Aires,
Buenos Aires, Argentina
{gconcilio,ecalot,jierache,hmerlino}@lsia.fi.uba.ar

Abstract. Keystroke dynamics is a biometric technique to identify users based on analysing habitual rhythm patterns in their typing behaviour. In e-commerce, this technique brings benefits to both security and the analysis of patterns of consumer behaviour. This paper focuses on analysing the keystroke dynamics against an e-commerce site for personal identification. This paper is an empirical reinforcement of previous works, with data extracted from realistic conditions that are of most interest for the practical application of modelling keystroke dynamics in free texts. It was a collaborative work with one of the leading e-commerce companies in Latin America. Experimental results showed that it was possible to identify typists with an accuracy of 89% from a sampling of 300 randomly selected users just by reading comment field keystrokes.

Keywords: keystroke dynamics, biometrics, e-commerce, user identification, user classification

1 Introduction

The characteristics that help make a handwritten signature a unique human identifier also provide a unique digital signature in the form of a stream of latency periods between keystrokes. The handwritten signature has a parallel on the keyboard. The same neurophysiological factors that make a written signature unique are also exhibited in a user's typing pattern[12]. Biometric characteristics are unique to each person and have advantages as they cannot be stolen, lost, or forgotten[19,1,5].

During a user session, every key-press generates two events, one when the key is pressed and other on its release. A typing rhythm, consisting of the intervals between keys, albeit random at first sight and naturally variable, can provide meaningful information about the user like its handedness[18] or its emotional state[6]. What is more, timing data can be enough to identify the typist. The objective of keystroke dynamics analysis is to study biometric timing signatures and, which information can be inferred from them and their sources of noise.

Passwords and fixed short texts fit the general framework of keystroke dynamics well; being short sequences typed in a row and repeated often, their characteristic timings tend to be reasonably stable in a broad sense. The free text does not enjoy the same privileges. Typing errors, corrections, misspellings, interruptions, pauses to think and attention lapses which are impossible to predict, interfere with the source timing data. Furthermore, short-term variations due to daily tiredness, stress or emotional shifts and long-term effects of health and re-education of typing skills are strictly unavoidable. Due to the mentioned difficulties, the problem of free text analysis of keystroke dynamics has only recently begun to be attacked[11] and in slightly controlled conditions. In [9] the finite context modelling of keystroke dynamics in free text is examined, reporting promising results.

In this paper, we present new experimentations as an empirical reinforcement of the work done by Calot *et al.*[3,10]. Also, we use data extracted from realistic conditions that are of most interest for the practical application of modelling keystroke dynamics in free texts. Similarly to [10], we examine the performance of techniques for keystroke dynamics analysis in free texts under evaluation conditions which are much harsher than the used ones initially[11,2,9] but we extend it to the e-commerce domain.

According to [9], establishing the universality of results has proven very difficult in keystroke dynamics based models due to the inherent complexities of classifying this noisy behaviour. Replication and implementation efforts using different training and evaluation datasets have often lead to astonishingly varying error rates in comparison with the original study, both for static passwords[5,15] and free texts[17,13].

This paper is a collaborative work between the LSIA and one of the leading e-commerce companies in Latin America. We propose to analyse if it is possible to strengthen the `user_id` from the collection of time series data describing the dynamics of typing, or typing patterns for short. The technique of modelling by finite contexts[9] has been adapted for the specific context of data experimentation of an e-commerce platform. The company facilitated the employment of its dataset for our experimentation. The general framework we used admits many implementation parameters and selection strategies; their effects on classification performance were studied.

2 Description of the Problem

In the specific context of e-commerce, it is possible to generate an account matching model —or authentication problem—. It consists of comparing previous behaviours of a given user with the current behaviour to discover if it is the same person, i.e., it seeks to validate that the person who is interacting with the platform is who he is claiming to be. In this context, we present a classification problem about two mutually exclusive classes, namely false or positive, depending on whether the current session is considered valid for the user with

whom it was previously authenticated based on the previous behaviour with the platform's points of contact.

Nevertheless, in the present paper, we propose to analyse the more complicated problem of user identification, less treated in previous works but of more significant interest for the forensic analysis. Identification is about finding the actual person with the same pattern over an extensive collection of typing patterns, while authentication is only checking against the assumed-right person. There are endless possibilities here, and some examples are: detecting fraudulent accounts (different `user_id`, same real person), detecting the identity of an intruder on a valid account (detecting the `user_id` of the intruder), or clustering compromised accounts; all by analysing keystroke dynamics on comments written on the platform.

A significant contribution of the present work is that we based it on an experimental sampling randomly selected without any bias, where users used their everyday keyboards resulting in entirely different settings for each of them. None of them were aware of this study, and, of course, the real identity of the person behind the user of the e-commerce platform was kept confidential, employing only the `user_id`. The device, web browser, operating system used, the configuration and type of keyboard may vary among different users and even for the same user between sessions. In addition, there is a temporal variability between sessions.

Keystrokes are fundamentally noisy; not only do we face the unpredictable and uncontrollable variations that occur due to physiological changes of the user —such as tiredness, emotional states, and consumption of medicines or other substances that may affect motor performance in a subtle way[6]—, but also we take information from an ecosystem of keyboards fixed, portable, tactile, among others; with different configurations and we also intend to integrate the modelling of keystroke dynamics in practical systems.

2.1 Free and Static Texts

The continuous authentication of users while typing employs the same structures and techniques as the verification of static keys (parameters, classifiers, distance metrics, characteristic vectors) although it adds some distinctive limitations. Among them, the need to fragment the input text to avoid noise originating from unnecessary pauses in data entry as effective measurements of parameters; and the impossibility of using metrics that take into account the correlation between keys since we expect the verified text to be substantially different from the training text[8]. In static texts such as passwords, the correlation of the types of events is very high —especially if the passwords do match— with a higher number of comparable events per amount of typed text. Unfortunately, as this work is on free texts, we are forced to use less accurate models to prevent the rate of false rejections from growing excessively.

Not all the suitable classifiers for static keys perform well with free texts, and in many cases, it is not even obvious how to adapt them to this situation.

Among the general purpose classifiers, those based on metric spaces have provided excellent results[7,16]. However, two new techniques of classification of specific purpose called metric R and metric A, initially proposed by Bergadano, Gunetti and Picardi in [2] and [11], stand out above the rest and have been used by additional authors[17] after optimising them for execution in real time, since the computational cost of these techniques is not irrelevant[10].

2.2 Theoretical Considerations and Common Techniques

The most commonly used characteristic parameters are the hold time —latency between key-press and release—, wait time —latency between key release and next key-press— and flight time —latency between successive key-press events— ; to a lesser extent, average typing speed, probability of error or usage frequency of backspace and delete, key usage patterns and key release order when writing uppercase letters. Usually, the timing between consecutive keys —called digraphs— are used, but occasionally latencies of longer groups are chosen[2].

According to [10], probably every technique for classification and machine learning has been used for the analysis of keystroke dynamics, ranging from simple metric spaces —with norms such as Euclidean, Manhattan, Mahalanobis and an infinitude of variations and combinations, also approaches such as outlier counts— and k-NN to state-of-the-art classifiers such as support vector machines or random forests; artificial neural networks, fuzzy logic and genetic algorithms have been tried too, with not very promising results. Killourhy and Maxion's review[15] has become a classic; an updated one can be found at [14].

3 Experiment and Results

3.1 Dataset

In an e-commerce application, there are limitless points of contact connecting an individual and the purchase process where the keyboard is used enabling acquisition of keystroke dynamics while the user is ignoring that his typing rhythms are being analysed to improve his security.

Keystrokes extracted from the contact points related to user authentication (password entry, email and username) and checkout (credit card data) are useful regarding security as a second factor of authentication and are compatible schemes with verification of static keys. On the other hand, keystrokes extracted from forms —such as an inquiry of a buyer to a seller about a particular product and its response, the rating of a purchase from the buyer to the seller and vice versa, claims, among others— are of variable length and they have a considerably long text extension, which is why it is more appropriate to model them with keystroke dynamics techniques for free texts; and more appropriate for our analysis.

We acquired a dataset from a production environment and not in a laboratory setting. As a plus, we did not choose the subjects from a university population but a global population of e-commerce site users. Different keyboards with

diverse regional configurations were used, even by the same individuals, and should be the ones the users have at home. Typing skills vary from mediocre to excellent. Noise factors affecting typing cadences such as interruptions, attention lapses and short-and-long-term emotional states, mood or health variations of the individuals were not excluded.

For each session, the timing of key down and key up events was recorded with a precision of 1 millisecond and the presumed user identity, previously authenticated with the valid credentials, is included. Due to the real world nature of the capture setup, we cannot guarantee that all sessions logged in with a particular identity are not used afterwards by a different user and, indeed, we suspect this is the case with at least some of them, but not enough to ruin the experiment. At worst, the reported results can be considered artificially degraded by this artefact, though not to a large extent[9].

The original dataset for experimentation provided by the company had 97,184 sessions of 5,673 users, many of the sessions of the original dataset did not have enough keystrokes to be considered for evaluation due to the length requirement. We accepted as valid sessions only the ones having at least 20 characters. The quantity minimum of valid sessions per user was set at 20. Since the task is unrestricted, some keystrokes might correspond to arrow keys and other special keys used for edition and navigation; in order not to negatively bias the results, the keystroke count referred to above considers only alphanumeric characters. Besides, all the sessions that have pressed the paste keys (ctrl+V or command+V) were discarded from the analysis, as they noticeably alter the user's natural keystroke behaviour.

After an adaptation and normalisation of the original dataset, to evaluate the proposed method, the final dataset consists of time series describing the keystroke dynamics from 372 users over eight months. Each user has in average 32 keystroke patterns (within 20–80), resulting in a collection of 11,766 keystroke patterns in total. Although all those cadences of less than 20 characters are filtered, the extension of the sessions shows a considerable variability reaching up to 300 key-pressures; the average extension for the entire data set is 56 keys per session. The total number of characters per user is approximately 1,650 characters (with a minimum of 1,000 and a maximum of 4,000). The intervals between successive sessions of the same user are also strongly variable, with values ranging from a few hours to several days.

3.2 Experimental Setup

We assume that the real person behind a `user_id` is always the same. We used various samples of randomly selected users to evaluate the behaviour of distinct identification methods with many customisable aspects. For each user, their keystrokes were subdivided into two mutually exclusive subsets, labelled A and B, previously applying a shuffle operation because the dataset was chronologically sorted. In this way, we generated two patterns with the same identity or `user_id`. For example, for the initial sample of 50 users, 100 patterns were obtained, two per user, one belonging to group A and another one to B. Through

the modelling of the keystroke dynamics of each pattern we try to discover which other pattern is more likely to share the same `user_id`. We count one success when we identify the `user_id` of set A with the least pattern distance to the same `user_id` of set B.

3.3 Reporting Metrics

The simple metric used to grade the quality of the identifying method is the percentage of correctly matched users, that is the amount of `user_id` from the set A matching the same `user_id` in the set B over the total amount of users in the sample.

3.4 Implementation and Results

Specific tools were developed to test several configurations of the proposed method in experimental patterns in e-commerce.

We adopted the modelling by finite contexts, employing the keyboard event as a context of order N (differs from the implementation of González *et al.*[9] that utilises key-based contexts and models flight, wait and, hold time). Here, all the contexts are analysed without differentiating flight, hold and wait time of a specific key but using event-related contexts where the combination of action —press or release— and key are considered a unique event.

To find the optimal configuration of the identification algorithm, we randomly selected a sample of 200 users from a pool in the dataset. We varied the context sizes from 1 to 10. To calculate the distance between two users, we selected only the common intersection of events and built two vectors with mean and variance per user. Finally, we fed the distance algorithm with the mean vector for both users —normalised Euclidean additionally required the variance vectors—. We employed the following distance metrics: Metric A[11], Euclidean norm ($\mathcal{L}_2$), normalised Euclidean norm ($\bar{\mathcal{L}}_2$), Manhattan norm ($\mathcal{L}_1$), and Metric R[11]. Table 1(a) shows the experimental results, registering an advantage in effectiveness rates of Euclidean distance and the R metric over the other ones.

Due to computational power constraints, we used up to a maximum context order of 10; notwithstanding this limitation, Table 1(a) attests that the optimal classification performance peaks around context lengths in between four and seven events.

Table 1(b) presents the experimental results obtained by varying the sample size in steps of 50 from a range of 50–300, all with a fixed context of 5 events arbitrarily chosen. The same table, illustrated in Fig. 1, uncovers that the effectiveness rates achieved by all the adopted distances are stabilised starting from the 200-user samples.

3.5 Key Observations

It was found that considering timing data of certain special keys —including modifiers (shift, ctrl, alt), navigation (arrows, page up, page down) and correctors

Table 1. Experimental results in user identification for (a) different context orders and (b) different user sampling sizes

Order	A	$\mathcal{L}_2$	$\bar{\mathcal{L}}_2$	$\mathcal{L}_1$	R
1	77	84	74	80	81
2	83	83	80	83	85
3	83	86	78	82	86
4	83	85	77	82	87
5	83	85	76	81	87
6	82	85	76	82	86
7	82	85	77	81	86
8	81	85	77	81	85
9	81	84	78	81	85
10	79	84	77	81	84

(a) Different context orders with de sample of 200 users.

Users	A	$\mathcal{L}_2$	$\bar{\mathcal{L}}_2$	$\mathcal{L}_1$	R
50	88	96	92	94	90
100	82	90	85	87	90
150	74	85	78	80	86
200	83	85	76	81	87
250	80	86	78	79	88
300	78	86	77	80	89

(b) Different user sampling sizes with 5 as context order.

(backspace and delete)— worsened the classifier's performance even though their consideration in the contexts of the next keystrokes improved it.

4 Concluding Remarks

Despite all the noise, we obtained an effectiveness of between 96% (50 users sampling size) and 89% (300 users sampling size) in the identification of users. The experimental conditions detailed in Sec. 3.2 show that the times taken from the keystrokes typed in online purchases rating forms inside the e-commerce platform can be employed as a practical application method to strengthen the identification of users.

Table 1(a) reveals that the best classifiers are those that adopted the metric R in almost all cases. Still, the versatility and faster execution times of the classifiers based on the Euclidean distance makes it a very competitive alternative. Fig. 1 shows that the effectiveness rates achieved by all the adopted distances are stabilised starting from the 200-user samples.

Besides, it was established that the method proposed in [9] and replicated in [10] is still useful in the context of electronic commerce. The classification performance cannot compete with other biometric systems but has the advantage of being entirely non-intrusive, transparent, and not requiring additional hardware or user actions. The selected method seems promising, and there is still room for further improvements in this particular context.

4.1 Future Lines of Research

As future lines of work, we are planning to operate with other forms of model selection (the highest order context); expand the time lapse of the experiment

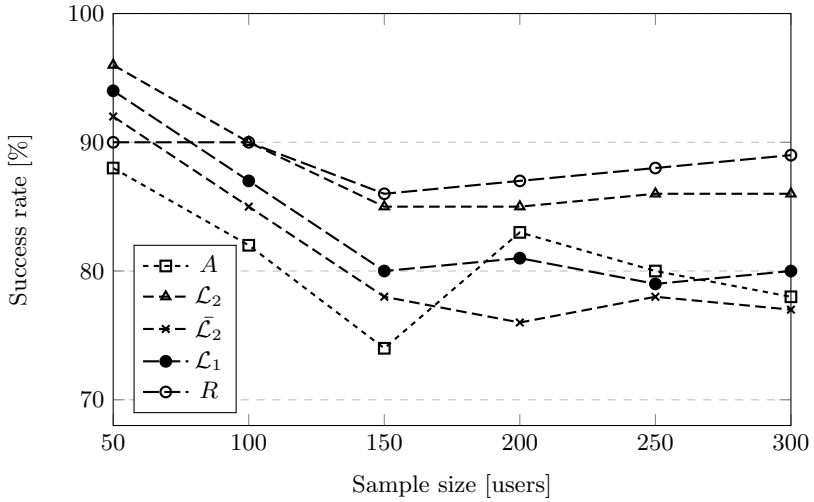


Fig. 1. Success percentage for samples from 50 to 300 for five metrics.

to make it last at least a year; utilise the algorithms to assert if the current user already existed on the platform (threshold at the minimum distance to be considered the same user); and optimise the code to be able to perform empirical tests with a higher number of users.

Our future lines of research will focus on eduction of emotions, initially contrasted against facial images and voice recognition, the results had already been explored in the UBACyT 2014–2017 GEF and will also be capitalised concerning the use of biometric EEG markers[4]; additionally, its applications in e-commerce. Also, the influence of emotional states and other noise sources to mitigate their adverse effects on classification.

References

1. Araujo, L.C.F., Sucupira, L.H.R.J., Lizarraga, M.G., Ling, L.L., Yabu-uti, J.B.T.: User authentication through typing biometrics features. Signal Processing, IEEE Transactions on 53(2), 851–855 (2005), `http://lsia.fi.uba.ar/papers/araujo05.pdf`
2. Bergadano, F., Gunetti, D., Picardi, C.: User authentication through keystroke dynamics. ACM Transactions on Information and System Security (TISSEC) 5(4), 367–397 (Nov 2002), `http://doi.acm.org/10.1145/581271.581272`
3. Calot, E.P.: Keystroke dynamics keypress latency dataset. Database (Jan 2015), `http://lsia.fi.uba.ar/pub/papers/kd-dataset/`
4. Calot, E.P., Ierache, J.S.: Multimodal biometric recording architecture for the exploitation of applications in the context of affective computing. In: Proceedings del XXIII Congreso Argentino de Ciencias de la Computación (La Plata, 2017). pp. 1030–1039. No. 10529 (2017)

5. Calot, E.P., Rodríguez, J.M., Ierache, J.S.: Improving versatility in keystroke dynamic systems. In: Proceedings del XIX Congreso Argentino de Ciencias de la Computación. No. 5606 (2013), http://ir.cs.uns.edu.ar/downloads/cacic_2013/11wsi.pdf
6. Epp, C., Lippold, M., Mandryk, R.L.: Identifying emotional states using keystroke dynamics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 715–724. CHI '11, ACM, New York, NY, USA (2011), http://doi.acm.org/10.1145/1978942.1979046
7. Flior, E., Kowalski, K.: Continuous biometric user authentication in online examinations. 2014 11th International Conference on Information Technology: New Generations 0, 488–492 (Apr 2010), http://lsia.fi.uba.ar/papers/flior10.pdf
8. González, N.: Utilización de contextos finitos para el modelado de la dinámica de tecleo en esquemas de autenticación mixta (Apr 2014), http://lsia.fi.uba.ar/papers/gonzalez14.pdf, internal paper from Laboratorio de Sistemas de Información Avanzada, Facultad de Ingeniería, Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina
9. González, N., Calot, E.P.: Finite context modeling of keystroke dynamics in free text. In: Biometrics Special Interest Group (BIOSIG), 2015 International Conference of the. pp. 1–5 (Sep 2015), http://ieeexplore.ieee.org/document/7314606/
10. González, N., Calot, E.P., Ierache, J.S.: A replication of two free text keystroke dynamics experiments under harsher conditions. In: 2016 International Conference of the Biometrics Special Interest Group (BIOSIG). pp. 1–6 (Sep 2016), http://ieeexplore.ieee.org/document/7736905/
11. Gunetti, D., Picardi, C.: Keystroke analysis of free text. ACM Transactions on Information and System Security (TISSEC) 8(3), 312–347 (Aug 2005), http://doi.acm.org/10.1145/1085126.1085129
12. Joyce, R., Gupta, G.: Identity authentication based on keystroke latencies. Commun. ACM 33(2), 168–176 (Feb 1990), http://doi.acm.org/10.1145/75577.75582
13. Kang, P., Cho, S.: Keystroke dynamics-based user authentication using long and free text strings from various input devices. Information Sciences 308, 72–93 (2015)
14. Karnan, M., Akila, M., Krishnaraj, N.: Biometric personal authentication using keystroke dynamics: A review. Applied Soft Computing 11(2), 1565–1573 (2011)
15. Killourhy, K.S., Maxion, R.A.: Comparing anomaly-detection algorithms for keystroke dynamics. In: International Conference on Dependable Systems & Networks (DSN-09). pp. 125–134. EEE Computer Society Press, Los Alamitos, California (2009), http://lsia.fi.uba.ar/papers/killourhy09.pdf
16. Leggett, J., Williams, G.: Verifying identity via keystroke characterstics. International Journal of Man-Machine Studies 28(1), 67–76 (1988)
17. Messerman, A., Mustafic, T., Camtepe, S.A., Albayrak, S.: Continuous and non-intrusive identity verification in real-time environments based on free-text keystroke dynamics. In: Biometrics (IJCB), 2011 International Joint Conference on. pp. 1–8. IEEE (2011)
18. Monrose, F., Rubin, A.D.: Authentication via keystroke dynamics. In: Proceedings of the 4th ACM conference on Computer and communications security. pp. 48–56. ACM (1997)
19. Polemi, D.: Biometric techniques: review and evaluation of biometric techniques for identification and authentication, including an appraisal of the areas where they are most applicable. Reported prepared for the European Commision DG XIIIC 4 (1997), http://lsia.fi.uba.ar/papers/polemi97.pdf

An Analysis of Convolutional Neural Networks for detecting DGA

Carlos Catania[1], Sebastian Garcia[2], and Pablo Torres[1]

[1] Facultad de Ingenieria. UNCuyo. Mendoza. Argentina
[2] CTU - Czech Technical University. Prague. Czech Republic

Abstract. A Domain Generation Algorithm (DGA) is an algorithm to generate domain names in a deterministic but seemly random way. Malware use DGAs to generate the next domain to access the Command Control (C&C) communication channel. Given the simplicity and velocity associated to the domain generation process, machine learning detection methods emerged as suitable detection solution. However, since the periodical retraining becomes mandatory, a fast and accurate detection method is needed. Convolutional neural network (CNN) are well known for performing real-time detection in fields like image and video recognition. Therefore, they seem suitable for DGA detection. The present work is a preliminary analysis of the detection performance of CNN for DGA detection. A CNN with a minimal architecture complexity was evaluated on a dataset with 51 DGA malware families as well as normal domains. Despite its simple architecture, the resulting CNN model correctly detected more than 97% of total DGA domains with a false positive rate close to 0.7%.

Keywords: Neural Networks, Network Security, DGA Detection

1 Introduction

A domain generation algorithm (DGA) is used to dynamically generate a large number of pseudo random domain names and then selecting a small subset of these domains for the Command Control (C&C) communication channel. The idea behind the dynamic nature of DGA was to avoid the inclusion of hard-coded domain names inside malware binaries, complicating the extraction of this information by reverse engineering [7]. The first DGA detection attempts depended on published lists of domains already detected as DGA. However, given the simplicity and velocity associated to the domain generation process, detection approaches that relied on static domain blacklists were rapidly rendered ineffective [5]. Nowadays, DGA detection methods can be classified in two main groups: (A) Context-based or (B) Lexicographical-based . The so-called context-based approaches mostly rely on the use of context information such as the responses from the DNS servers. A typical example would the fact that DGA domains are not usually registered and therefore the DNS response is usually *NXDomain*. On the other hand, lexicographical approaches classify the domains by studying the statistical properties of the characters conforming the domain name.

Regarding Lexicographical approaches, Natural Language Processing (**NLP** emerged as one of the most useful techniques for detecting DGA, specially in the analysis of the n-gram frequency distribution of domain names. An n-gram is defined as a contiguous sequence of n items from a given sequence of text. It is possible to use greater values for n than 1. In the simpler form, when $n = 1$, the single character frequency distribution is generated. The assumption is that DGA domains will have a different n-gram distribution than normal domain names. An example of n-gram based DGA detection are [12, 9] that compares uni-grams and bi-grams distribution using the Kullback-Leibler (K-L) divergence.

Several other DGA detection approaches extended the idea of using the information provided by domain names properties (including n-gram distributions) to train a machine learning classifier such as Random Forest[1] or Linear Regression[12]. Recently, to avoid the need of designing the right set of features for training machine learning classifiers, some authors explored the application of Deep Learning (DL) techniques. In particular the application of Long-Short-Term-Memory (**LSTM**) networks [2]. When applied to the text analysis problem, the internal design of LSTM cell is capable to capture combinations of characters that are important to discriminating DGA domains from non-DGA domains. This flexible architecture generalizes manual feature extraction like n-grams, but instead learns dependencies of one or multiple characters, whether in succession or with arbitrary separation.

Despite the good results reported, LSTM networks have proved to be difficult to train under some particular cases [6]. The aforementioned issue together with the considerable time required during training could be the major obstacle for the massive adoption of LSTM networks in DGA detection. The fact is that DGA techniques change over the time and the periodical retraining of the network becomes mandatory. However, if the presence of long-term dependency patterns in DGA domain names is ignored, it is possible to apply another well-known Deep Learning technique: Convolutional Neural Networks (**CNN**). CNN are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers. CNN have been successfully applied in a many practical applications mainly related to image and video recognition. When applied to text analysis the convolution is applied over one dimension and denoted as 1D-CNN. The main advantage of 1D-CNN are they can be trained much faster than LSTM (up to 9X times faster than LSTM) and similarly to LSTM, 1D-CNN are capable of learning representations for groups of characters without being explicitly told about the existence of such groups. However, they can't deal with patterns with arbitrary separation.

In the present work we focus on analyzing the DGA detection performance of a 1D-CNN. Similarly to [11], we are interested in evaluating the performance of a network with a minimal architecture complexity. Hence, the considered network architecture consists of just a minimal extra layers in addition to the Convolutional Layer. Evaluation is conducted on a dataset containing 51 different DGA families as well as normal domain names from the Alexa corpus and the Bambenek feeds. Two different DGA schemes, including the recent word-based, are

Table 1. The complete network architecture including the corresponding output dimensions and activation function used in each layer

Layer (type)	Activation Function
input (Input Layer)	-
embedding (Embedding)	-
conv1d (Conv1D)	relu
dense_1 (Dense)	relu
dense_2 (Dense)	sigmoid

included in the dataset. Word-based DGA consists of concatenating a sequence of words from one or more wordlists, resulting in less randomly appealing domain name and thus more difficult to detect [7].

The hypothesis presented in this work is that despite the 1D-CNN limitations, they can learn the common properties from the different DGA generation schemes and the resulting detection performance is within the range required for real-world scenarios.

The main contributions of the present article are:

- An analysis of the advantages and limitations of a simple 1D-CNN learning model for detecting DGA.
- The detailed evaluation of 1D-CNN on a extended dataset that includes domain names from 51 different real malware DGA following different generation schemed as well as normal domains from two different sources.

The rest of this paper is organized as follows: Section 2 describes the network architecture, Section 3 details the experiments design and results while Section 4 discusses the importance of the results, and finally in Section 5, we presents the conclusions.

2 The Neural Network Architecture

The Neural Network Architecture model is a 1D Convolutional Neural Network (CNN). This CNN is composed of three main layers. The first one is an *Embedding* layer, then there is a *Conv1D* layer, and finaly a Dense fully connected layer. The first two layers are the most relevant components of the architecture regarding the problem of detecting DGA domains. Both layers are responsible for learning the feature representation in order to feed the third Dense and fully connected layer. Beside the three layers previously described, the complete Neural Network Architecture includes some other layers for dealing with the dimensions output of the *Conv1D* layer as well as layers for representing the input domain and the output probability. A detail of the complete architecture together with the used activation functions is shown in Table 1, whereas the three main layers are described in the following subsections.

2.1 Embedding Layer

A character embedding consists in projecting l-length sequences of input characters into a sequence of vectors R^{lxd}, where l has to be determined from the information provided by the sequences in the training set and d is a free parameter indicating the dimension of the resulting matrix [11]. By using an *Embedding* layer in the architecture, the neural network learns in an efficient manner the optimal set of features that represent the input data.

2.2 1D Convolutional Layer

The *Conv1D* layer refers to a convolutional network layer over one dimension. For the DGA detection problem, such dimension consists of the length of the domain name sequence. The convolutional layer is composed of a set of convolutional filters that are applied to different portions of the domain name. A visual example of the feature extraction process for a 1D Conv layer is shown in Fig. 1. The figure depicts a 1D convolutional layer constructing 256 filters (features) ($nf = 256$), with a window (kernel) size of 4 ($ks = 4$) and a stride length value of 1 ($sl = 1$). The layer selects from groups (also referred as patches) of 4 characters to apply the convolutional filters, and continues shifting one character at a time (stride value) applying the same convolutions filter over the rest of the sequence. Consequently, the neural network generates 4-grams features. These features represent the discriminative power of these group of letters in the domain names.

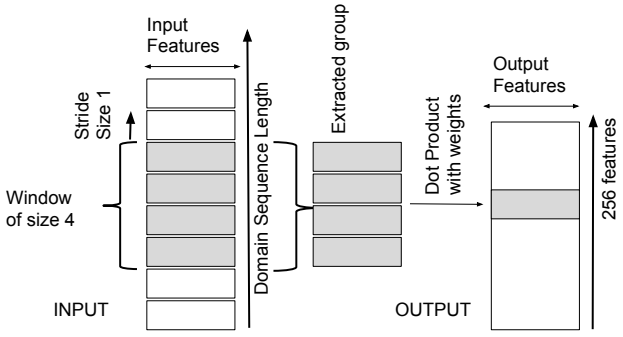


Fig. 1. Feature extraction process of the 1D convolutional layer

By applying the same filter all over the sequence the required computation time is considerable reduced when compared with traditional Multilayer Perceptron layers. Additionally, since a convolutional kernel independently operates on each 4-gram it is possible to go over the entire input layer concurrently. This paralellization and its consequent low computing time is one of the major benefits of using convolutional networks instead of other deep learning approaches usually used for text processing such as Long Short Term Memory (LSTM) [3, 10, 11].

2.3 Dense Layers

The features extracted by both previous layers are be used by traditional Multilayer Perceptron Network (MLP) to output the probability of a given domain belongs to DGA or Normal class. MLP is composed of two layers: A first fully connected layer of size *hn* (*Dense* layer) connected to a second *Dense* layer of size 1 used for actually giving the probability output about the considered domain.

3 Experimental Design

The analysis focused on the detection performance considering the DGA and normal domains, as well as the detection performance on the different malware families included in the dataset. Several standard performance metrics for network detection evaluation were used. These metrics correspond to True Positive Rate (**TPR**) and False Positive Rate (**FPR**). TPR is computed as the ratio between the number of correctly detected DGA domains and the total number of DGA domains. Whereas FPR is computed as the ratio between the number of normal domains that are incorrectly classified as DGA and the total number of normal domains.

The evaluation of the 1D-CNN DGA detection method was carried out following the usual machine learning methodology. A dataset containing both DGA and normal domains was split in a 70:30 ratio. To guarantee the independence of the results, the 70% of the datasetset was used for tuning the 1D-CNN hyperparameters (training set). Whereas the remaining 30% (testing set) was used for testing the performance of the 1D-CNN DGA detection model on unseen domains.

3.1 Dataset Description

The 1D-CNN detection method was evaluated on a dataset containing both DGA and normal domain names. The normal domain names were taken from the Alexa top one million domains. An additional 3,161 normal domains were included in the dataset, provided by the Bambenek Consulting feed. The later group is particularly interesting since it consists of suspicious domain names that were not generated by DGA. Therefore, the total amount of domains normal in the dataset is 1,003,161 . DGA domains were obtained from the repositories of DGA domains of Andrey Abakumov [3] and John Bambenek [4]. The total amount of DGA domains is 1,915,335, and they correspond to 51 different malware families. The list of malware families and the amount of domains selected can be seen in Table 2.

The DGA generation scheme followed by the malware families includes the simple arithmetical (A) and the recent word based (W) schemes. Under the arithmetic scheme, the algorithm usually calculates a sequence of values that

[3] https://github.com/andrewaeva/DGA
[4] http://osint.bambenekconsulting.com/feeds/

Table 2. Episode frequency and generation scheme for the DGA Malware Families in the dataset used for training and testing the 1D-CNN for DGA detection method. (A) stands for Arithmetic generation scheme while (W) for word-based

Family	Scheme	Freq.	Family	Scheme	Freq.	Family	Scheme	Freq.
bamital	(A)	904	cryptolocker	(A)	112,809	padcrypt	(A)	1,920
p2p	(A)	4,000	proslikefan	(A)	100	murofet	(A)	49,199
bedep	(A)	706	dircrypt	(A)	570	necurs	(A)	81,920
post	(A)	220,000	dyre	(A)	26,993	newgoz	(A)	1,666
chinad	(A)	256	fobber	(A)	600	nymaim	(A)	20,225
conficker	(A)	99,996	gameover	(A)	12,000	pushdo	(A)	94,278
corebot	(A)	840	geodo	(A)	1,920	pykspa	(A)	25,727
goz	(A)	1,667	hesperbot	(A)	192	qadars	(A)	1,600
kraken	(A)	9,660	locky	(A)	9,028	qakbot	(A)	60,000
ramdo	(A)	102,000	ramnit	(A)	91,978	ranbyus	(A)	23,167
rovnix	(A)	53,632	shiotob	(A)	12,521	symmi	(A)	4,448
shifu	(A)	2,554	virut	(A)	11,994	sisron	(A)	60
zeus	(A)	1,000	vawtrak	(A)	300	simda	(A)	28,339
tinba	(A)	193,912	tempedreve	(A)	225	pykspa_v1	(A)	18
pykspav2F	(A)	800	pykspav2R	(A)	200	banjori	(W)	439218
suppobox	(W)	8185	matsnu	(W)	100127	volatile	(W)	996
beebone	(W)	210	cryptowall	(W)	94	madmax	(A)	2

have a direct ASCII representation usable for a domain name. On the other hand, word-based consists of concatenating a sequence of words from one or more wordlists.

3.2 Hyper Parameters Tuning

Adjusting the network hyper-parameters is perhaps the most difficult task in the application of neural networks. For the proposed 1D-CNN architecture several hyper parameters needed to be adjusted. Among all the possible hyper-parameters, we particularly focused on finding the optimal values related to the *Embedding*, *Conv1D* and *Dense* layers. A traditional grid search was conducted through a specified subset on the training set. For a robust estimation, the evaluation of each parameter combination was carried out using a k-fold cross validation with $k = 10$ folds. The The 1D-CNN was trained using the back propagation algorithm [8] considering the *Adaptive Moment Estimation* optimizer [4]. The 1D-CNN training was carried out during 10 epochs.

Table 3 shows the parameter combinations with the better performance detection in terms of the F1-Score.

Table 3. Best Hyper-parameters subset. For space reason only the first parameter combination is shown. The parameters combination with the higher average F1-Score was chosen for all the remaining experiments.

avg. F1	sd	nf	ks	sl	d	l	hn	parameters
0.9797	**0.0050**	**256**	**4**	**1**	**100**	**45**	**512**	5,612,705

3.3 Evaluation on Unseen Domains

The present section details the results after evaluating the 1D-CNN with the hyper-parameters selected from previous section. In this case, complete training set was used for training the 1D-CNN and then tested on the remaining 30% of the dataset. The metrics described in subsection 3 were calculated on testing set with a decision boundary threshold set to 0.90. In terms of the two considered classes (i.e Normal or DGA), the resulting TPR value was around 97% of total DGA present in the dataset while the FPR was 0.7% of total normal domains.

Regarding the FPR discriminated by normal domains types, the 1D-CNN DGA detection algorithm has correctly detected as normal almost the 100% of total Alexa domains. However, for the case of the Bambenek domains, the false positive rate increased to 16%.

In Fig. 2 we show the TPR per DGA malware family and its DGA generation scheme (red for arithmetic and blue for word based). The diameter of the circle around each point provides a visual idea about the absolute frequency of DNS requests that belong to that malware family in the training set. In general, most of the DGA were detected with TPR values close to 0.75 no matter the DGA generation scheme used. The only three exceptions were `Cryptowall` with TPR value of 0.16, `Virut` with 0.56 and `Suppobox` malware with 0.47.

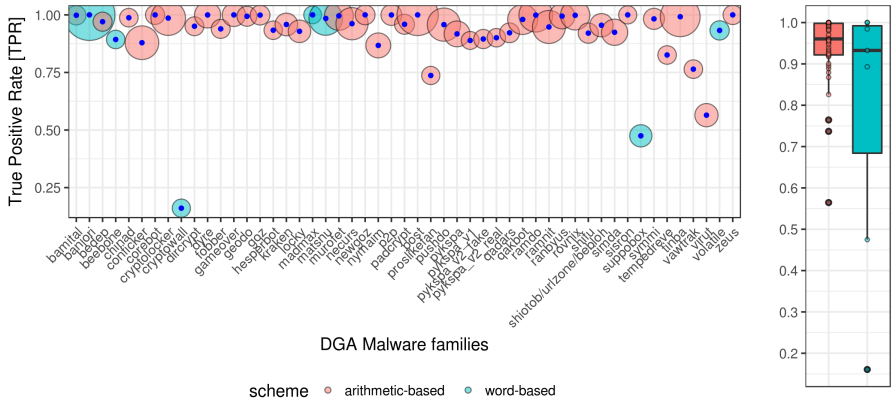


Fig. 2. True Positive Rate per DGA malware family

4 Discussion

It seems that the 1D-CNN detection method was capable of extracting common patterns present in arithmetic-based DGA generation algorithms of the different malware families. The previous claim could be an explanation of the good results observed in malware families with very low episode frequency. That is the case of families such as `chinad` and `bamital`, which with less than 1000 examples showed a TPR greater than 90%.

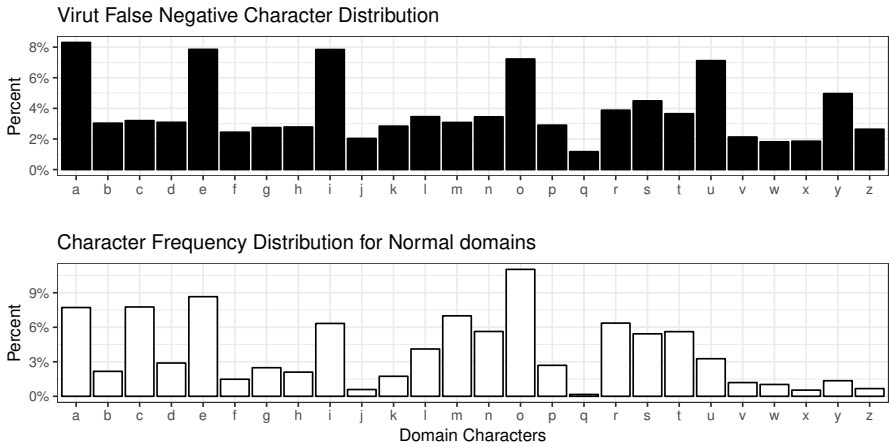


Fig. 3. Character Frequency Distribution for not detected DGA malware

The situation was different for the word-based DGA schemes. As can be seen in the boxplot located at the right of Fig. 2, the 75% of TPR for arithmetic-based DGA were concentrated between 1 and 0.92. On the other hand, in the case of word-based, the 75% of TPR values were between 1 and 0.58. It seems clear that TPR for malware families using the arithmetic-based generation scheme were significantly better than word-based. Since the word-based generation scheme aim at imitating the look of the normal domains, such poor results are expected.

There was, however, some unexpected results that deserved a deeper analysis. In particular, results showed for the `Matsnu` malware family, a word-based DGA with a 0.92 TPR and the case of `Virut`, an arithmetic-based DGA with a 0.57 TPR.

In the case of `Virut`, the poor performance can be explained by the the Frequency Character Distribution (**FCD**) of the generated DGA domains. The Fig. 3 shows the FCD for the `Virut` DGA malware (in black) and compare them with FCD for normal domains (in white). Notice that for ease of comparison only characters present in DGA domains were included in the normal FCD.

Despite not being a word-based DGA, `Virut` vowels frequency is considerable higher when compared with the rest of the characters. Moreover, such vowel frequency in DGA domains shows similarities with the normal FCD. The similarities with normal FCD could difficult the discrimination process carried out by the 1D-CNN detection method. Such process is even more complicated since `Virut` a small number of episodes included in the dataset.

In the case of `Matsnu`, the high TPR could not be explained by the FCD, since, as expected, `Matsnu` FCD shared similarities with normal FCD. After inspecting several aspects of the DGA, we found that domain name length was the discriminative feature between normal and `Matsnu` domain names. Domains names generated by `Matsnu` were considerable larger than normal domains. In

Fig. 4 we compare the Character Length Distribution (**CLD**) of `Matsnu` and normal domains. As can be seen, the `Matsnu` domains are significant larger than normal domains. On the other hand, the `suppobox` malware (shown also in Fig. 4) has a CLD close to normal domains. Since not only CLD but also FCD are similar to normal domains, `suppobox` domains become very difficult to detect.

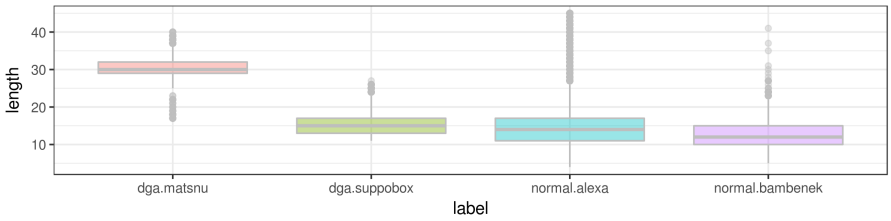


Fig. 4. character length distribution of `Matsnu`, *Suppobox* and normal domains names

5 Concluding Remarks

In the present work, we explored the viability of 1D-CNN for lexicographical DGA detection. A network with a minimal architecture complexity was evaluated on a dataset containing domains names generated by 51 different malware families as well as normal domains. Malware families included two different DGA generation scheme: (1) the most common arithmetic-based and (2) the recent and more difficult to detect Word-based. The dataset was properly split in training and testing sets. A hyper-parameters grid search was conducted on training set and the best resulting model was then evaluated on the testing set.

Despite its simple architecture, the resulting 1D-CNN model correctly detected more than 97% of total DGA domains with a FPR around 0.7%. Such results make it suitable for real-life networks. Inspecting the malware family results, we observed that the 75% arithmetic-based DGA showed TPR values between 1 and 0.92. Such values were observed even in malware families with low frequency episode, confirming the hypothesis that a 1D-CNN can learn the common properties from the different DGA generation families, at least under the arithmetic-based generation scheme. On the other hand, in the case of word-based, the 75% of TPR values were between 1 and 0.58. Such high variability in the TPR results responds to the high TPR for malware such as `Matsnu` and the low TPR for malware such as `suppobox` and `cryptowall`. By analyzing the FCD and CLD we detected than even tough both shared the FCD with normal domains, `Matsnu` showed a significant increment in the length of the domains, a situation that makes it easily detectable. Different was the case of malware such as `suppobox`, that shared both FCD and LCD with normal domains, and consequently was very difficult to detect.

To sum up, despite the good results shown by 1D-CNN, we observed that it could be very difficult to detect word-based DGA that share the CLD and FCD with normal domains. Therefore, as the word-based generation scheme become more precise in imitating normal domains the detection could be extremely hard.

Moreover, it is possible that such particular cases be outside the capabilities of detection methods following a lexicographical approach. Consequently, an alternative strategy could be necessary.

Acknowledgments

The authors would like to thank the financial support received by CVUT and UNCuyo during this work. In particular the founding provided by the Czech TACR project no. TH02010990 and the PICT 1435 granted by ANPCyT.

References

1. Ahluwalia, A., Traore, I., Ganame, K., Agarwal, N.: Detecting broad length algorithmically generated domains. In: Traore, I., Woungang, I., Awad, A. (eds.) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. pp. 19–34. Springer International Publishing, Cham (2017)
2. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016), http://www.deeplearningbook.org
3. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. 9(8), 1735–1780 (Nov 1997)
4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR abs/1412.6980 (2014)
5. Kührer, M., Rossow, C., Holz, T.: Paint it black: Evaluating the effectiveness of malware blacklists. In: Stavrou, A., Bos, H., Portokalidis, G. (eds.) Research in Attacks, Intrusions and Defenses. Springer International Publishing (2014)
6. Pascanu, R., Mikolov, T., Bengio, Y.: Understanding the exploding gradient problem. CoRR abs/1211.5063 (2012), http://arxiv.org/abs/1211.5063
7. Plohmann, D., Yakdan, K., Klatt, M., Bader, J., Gerhards-Padilla, E.: A comprehensive measurement study of domain generating malware. In: 25th USENIX Security Symposium (USENIX Security 16). pp. 263–278. USENIX Association, Austin, TX (2016)
8. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Neurocomputing: Foundations of research. chap. Learning Representations by Back-propagating Errors, pp. 696–699. MIT Press (1988)
9. Schiavoni, S., Maggi, F., Cavallaro, L., Zanero, S.: vol. 8550, pp. 192–211. Springer (7 2014)
10. Torres, P., Catania, C., Garcia, S., Garino, C.G.: An analysis of recurrent neural networks for botnet detection behavior. In: 2016 IEEE Biennial Congress of Argentina (ARGENCON). pp. 1–6 (June 2016)
11. Woodbridge, J., Anderson, H.S., Ahuja, A., Grant, D.: Predicting Domain Generation Algorithms with Long Short-Term Memory Networks (2016), http://arxiv.org/abs/1611.00791
12. Yadav, S., Reddy, A.K.K., Narasimha Reddy, A.L., Ranjan, S.: Detecting algorithmically generated domain-flux attacks with DNS traffic analysis. IEEE/ACM Transactions on Networking 20(5), 1663–1677 (2012)

Análisis, desarrollo e implementación de sistema que permite detectar páginas fraudulentas

Santiago Hernan Duran Paredes[1], Jorge Eterovic[1], Luis Torres, Iris Sattolo[1].

[1]Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales,
Universidad de Morón.
Cabildo 134 - CP (1708) - Morón - Prov. de Bs. As. Tel: 5627-2000
sam.paredes@hotmail.com, jeterovic@hotmail.com, torreslu@ar.ibm.com,
iris.sattolo@gmail.com

Resumen. La seguridad informática requiere una actualización constante para enfrentar una amplia gama de amenazas y evitar que estos hechos se lleven a cabo. El phishing es una actividad ilícita, cuyo principal objetivo es robar datos. Uno de los métodos que se usa para evitar esta actividad ilícita es mediante campañas de prevención, pero las mismos se continúan presentando, generando malestar en la sociedad. Se propone elaborar una aplicación, que permita a los usuarios reconocer si están ante una página segura o no y así evitar proporcionar datos personales, a la página de dudosa procedencia. Deberá indicar que nivel de seguridad presenta la página, ya sea fuerte, moderada, baja o pobre, informando los métodos que usa el certificado digital, y dando puntaje a los mismos, indicando que tan seguro es.

Keywords: phishing, certificado digital, seguridad informática, sistema.

1 Introducción

Este trabajo pretende contribuir al problema de phishing mediante una aplicación que permita analizar certificados de seguridad de páginas web. Se ha realizado un análisis comparativo de ciertas páginas web que analizan certificados digitales informando y analizando dicho certificado.

Se llevó a cabo un análisis de la problemática actual de dicha actividad ilícita y se ha realizado una investigación exploratoria documental respecto a definiciones de phishing, certificado digital, los métodos de cifrado de datos y la actualidad de Argentina en los ataques de phishing y qué medidas se toman habitualmente para contrarrestar este mal.

A pocos días del inicio del año 2018, Bitglass (compañía que ofrece soluciones de seguridad en el servicio de nube) ha recopilado algunas de las predicciones que estiman se van a dar en 2018 en el campo de la seguridad. Las nuevas formas de phishing, el malware avanzado, los errores humanos, y las multas por la entrada en vigor del reglamento de protección de datos y la no adecuación de las compañías, son a juicio de esta empresa los elementos que mayor impacto tendrán en este ámbito [1].

En 2017, los ataques de todos los tipos de phishing financiero – ataques contra bancos, sistemas de pago y tiendas online – crecieron un 1.2, 4.3 y 0.8 puntos porcentuales respectivamente [2].

Hartmann refiere que por phishing se entiende que "son varias las metodologías para cometer fraude bancario, pero el phishing es sin dudas el más notable: se monta una escena o instancia de comunicación premeditadamente engañosa donde algún elemento persuasivo lleva al cliente, enceguecido, a entregar sus datos personales al hacker" [3].

Talens-Oliag define el certificado digital como, "Para los usuarios proporcionan un mecanismo para verificar la autenticidad de programas y documentos obtenidos a través de la red, el envío de correo encriptado y/o firmado digitalmente, el control de acceso a recursos, etc." [4].

En si el certificado digital es una clase de archivo, informativo que su principal actividad es comunicar al usuario que está visitando dicha página web, es confiable, segura y que cumple con los protocolos básicos de seguridad, permitiendo al visitante, confiar plenamente en dicha página, sin preocuparse que la misma sea fraudulenta.

Talens-Oliag hace referencia a la criptografía de dos formas:

- Vista en términos sociales, es la ciencia de hacer que el costo de adquirir o alterar información de modo impropio sea mayor que el posible valor obtenido al hacerlo.
- Vista en términos más formales, es la práctica y el estudio de técnicas de *cifrado* y *descifrado* de información, es decir, de técnicas para codificar un mensaje haciéndolo ininteligible (*cifrado*) y recuperar el mensaje original a partir de esa versión ininteligible (*descifrado*) [5].

González Pérez, señalo en un estudio reciente que **la Argentina está entre los países del mundo que más mensajes de phishing recibe**. A la vez, expertos en seguridad informática señalan que las técnicas de engaño de esta modalidad delictiva se refinan cada vez más [6].

Un relevamiento (que realizó la empresa de seguridad Kaspersky Lab) que abarcó el tercer trimestre de 2017 ubicó a la Argentina como el **séptimo país del planeta con mayor cantidad de ataques de phishing recibidos**, en un ranking que encabeza Brasil y que tiene entre sus primeros diez puestos a Australia, Nueva Zelanda, China, Francia, Perú, Canadá, Qatar y Georgia [6].

Se ha detectado que hubo un incremento gradual de phishing no solo en este país (Argentina) sino en todas partes del mundo, a pesar de los esfuerzos de realizar campañas de concientización, esto continúa creciendo.

Se citan a continuación 6 formas de proteger a los usuarios de los ataques, intentan minimizar los efectos negativos de un ataque phishing y si es posible de impedirlo:

1. **No compartas tu nombre de usuario y contraseña con nadie**, incluso si la persona afirma ser un empleado del banco.
2. **Escoge un proveedor de correo electrónico que ofrezca autenticación en dos fases.** También es buena idea que tenga filtros de spam, malware y suplantación de identidad porque te mostrará un mensaje de alerta si algo parece sospechoso.
3. **Utiliza únicamente tus datos de acceso en la aplicación oficial del banco** (por ejemplo, el enlace de la app N26), nunca te descargues la aplicación de ningún otro sitio. Si eres cliente de N26, nunca utilices tus

datos de acceso en otro dominio que no sea https://app.n26.com o https://my.n26.com

4. **No hagas clic en los enlaces si recibes un correo electrónico que te pide que realices una acción que tú no has iniciado** (restablecer la contraseña, validar tu cuenta, etc.)
5. Comprueba siempre el enlace antes de hacer clic en él. **Coloca el ratón encima para obtener una vista previa de la URL** y fíjate cuidadosamente si hay faltas de ortografía u otras irregularidades.
6. Los sitios web de los bancos siempre usan HTTPS. **Si no ves el icono del candado verde en el navegador o no ves el prefijo «https» antes de la URL del sitio, \ es probable que el sitio no sea seguro [7]**.

Dado este contexto, este artículo presenta los servicios considerados para el análisis comparativo (Sección 2). Presenta una aplicación que permita analizar certificados de seguridad de páginas web (Sección 3) y un caso de estudio para la validación de la propuesta de la solución (Sección 4). Se formulan conclusiones y futuras líneas de trabajo (Sección 5).

2 Servicios considerados

En esta sección se presentan servicios que se han contemplado para el estudio comparativo: Qualys SSL Labs [8] (sección 2.1), Redalia [9] (sección 2.2), SSL Checker [10] (sección 2.3), ImmuniWeb® SSLScan [11] (sección 2.4). Por último, se presenta el resultado del análisis realizado (sección 2.5).

2.1 Qualys SSL Labs

Este servicio en línea gratuito realiza un análisis profundo de la configuración SSL de cualquier servidor de la web. Analiza el certificado como por ejemplo a quien provee el mismo, que métodos de encriptación usa, da un puntaje total en cuanto a que tan fuerte es el certificado, etc. Da bastante información que para un usuario común puede llegar a ser confusa y que no es necesaria para el objetivo de este trabajo. [8]

2.2 Redalia

La página Redalia que ofrece el servicio de análisis, ayuda a verificar que el certificado SSL del sitio web está correctamente instalado, y en caso contrario te muestra los problemas que presenta y la manera de solucionarlos. Pero no hace un análisis total de cada componente del certificado digital, sin indicar un resultado en su totalidad o de cada componente del certificado, informa las propiedades del certificado como una mera información. [9]

2.3 SSL Checker

El sitio SSL Checker realiza un análisis del certificado, indicando si cumple con cada parte del certificado, asignándole una puntuación por separado, pero no hace un análisis general en cuanto al certificado. Lo destacable del sitio se puede consultar las reseñas, que realizan los usuarios de dicha página, de la empresa que certifico la página que se está analizando [10].

2.4 ImmuniWeb® SSLScan

ImmuniWeb® SSLScan es un producto gratuito disponible en línea, proporcionado y operado por High-Tech Bridge.

Con el objetivo de permitir a cualquier persona evaluar cuán seguro y confiable es su conexión SSL / TLS con un servidor (en cualquier puerto), el servicio realiza cinco pruebas distintas:

- Prueba el cumplimiento de los requisitos de las PCI DSS;
- Prueba de conformidad con la guía de HIPAA;
- Prueba de conformidad con las Pautas de NIST;
- Prueba las vulnerabilidades y debilidades más recientes de SSL / TLS;
- Prueba contenido inseguro de terceros que pueda exponer la privacidad del usuario [11].

Este sitio es completo en cuanto al análisis del certificado de seguridad dando un análisis individual y total de la página analizada, también analiza contenido de terceros, que pueden ser imágenes, JavaScript o CSS.

2.5 Dimensiones consideradas para el análisis

En esta sección se plantean las dimensiones que se han considerado para el análisis de las distintas aplicaciones que examinan el certificado digital. Ellas son:

- Puntaje final: una vez que se analizó todo el certificado digital, se debe informar un puntaje total indicando el nivel de seguridad que representa la página web.
- PFS: Perfect Forward Secrecy, cuando una conexión cifrada utiliza PFS, significa que las claves de sesión que se generan por el servidor son efímeras, incluso alguien que tenga el acceso a la clave secreta no puede utilizar la clave de sesión relevante que descifra cualquier sesión HTTPS. Por lo tanto, los datos cifrados interceptados están protegidos de miradas indiscretas en el futuro, incluso si la clave secreta del sitio web se ve comprometida posteriormente [12].
- TLS versión: Permite confiar información personal a sitios web, ya que los datos se ocultan a través de métodos criptográficos mientras se navega en sitios seguros[13].
- Key Exchange: El intercambio de clave encriptada es un protocolo, o conjunto de reglas, que permite a dos partes compartir una contraseña común

para comunicarse en una red insegura sin exponer esta contraseña. Este protocolo fue desarrollado originalmente por Steven Bellovin y Michal Merritt de Laboratorios AT&T Bell, quienes produjeron un trabajo seminal en el tema [14].

- Bulk Cipher: es un algoritmo de cifrado simétrico que se utiliza para cifrar y descifrar grandes cantidades de datos [15].
- MAC: los datos a enviar, junto con una clave secreta, se utiliza para generar un código de autenticación de mensaje. Los datos más el código se transmiten al receptor deseado. EL recepto realiza mediante el mismo algoritmo de clave secreta, un cálculo sobre los datos, utilizando la misma clave secreta para generar un nuevo código de autenticación de mensaje y compara el resultado con ese cálculo [16].

Tabla 1. Cuadro comparativo de métodos criptográficos que usan los analizadores de certificados digitales

	Qualys SSL Labs	**Redalia**	**SSL Checker**	**ImmuniWeb® SSLScan**
Puntaje final	X			X
PFS				X
TLS Versión	X			X
Key Exchange	X			
Bulk Cipher	X	X	X	X
MAC				

Los resultados de los análisis obtenidos a través de la tabla comparativa (Tabla 1) permiten formular las siguientes conclusiones: ninguna de las páginas pudo analizar todos los requisitos en su totalidad.

Por ejemplo, Redalia y SSL Checker en su totalidad de las dimensiones consideradas, no realiza análisis porque lo que hace es mostrar los datos que cuenta el certificado, pero no hace análisis total y de cada parte del mismo.

A diferencia de sus predecesores los sitios Qualys SSL Labs y ImmuniWeb® SSLScan, abarcaron casi todas las dimensiones excepto para el primer sitio PFS y el segundo Key Exchange y ambos coinciden que no pudieron analizar la MAC.

Como modalidad de todas estas páginas, se observó que no cuentan con un servicio automatizado a la hora de analizar los certificados web, sino que se debe ingresar el link del sitio a mano.

A partir del análisis realizado en el cual se detectaron las áreas de vacancia mencionadas, se propone el desarrollo de un sistema que considere las dimensiones analizadas.

(sección 3).

3 Diseño de aplicación de detección de páginas fraudulentas

La aplicación desarrollada (llamada SSLvalidations) es un addons (ayuda a modificar y personalizar la experiencia de la navegación agregando funcionalidades a Firefox) que va embebido en Firefox, y el mismo cuenta con un icono que, a medida que se va navegando, va indicando si la página cumple con los certificados de seguridad que se vieron en la sección anterior. A su vez, informa los métodos criptográficos que usa el certificado.

El sistema va analizando el certificando, pidiendo determinados valores para luego utilizarlos en el análisis.

En la figura 1 se presenta un diagrama de clase en el cual se visualizan las llamadas que el sistema va realizando al certificado, solicitando determinados parámetros, que serán utilizados para su posterior análisis.

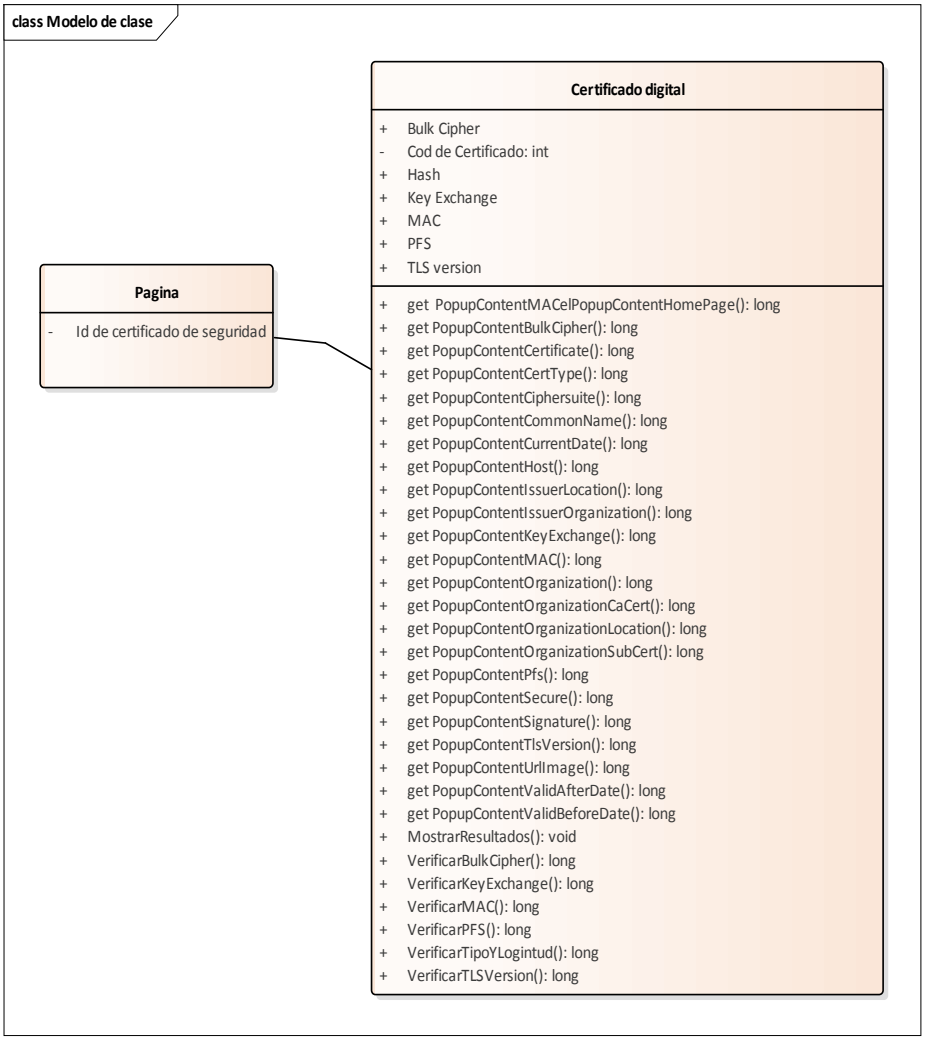
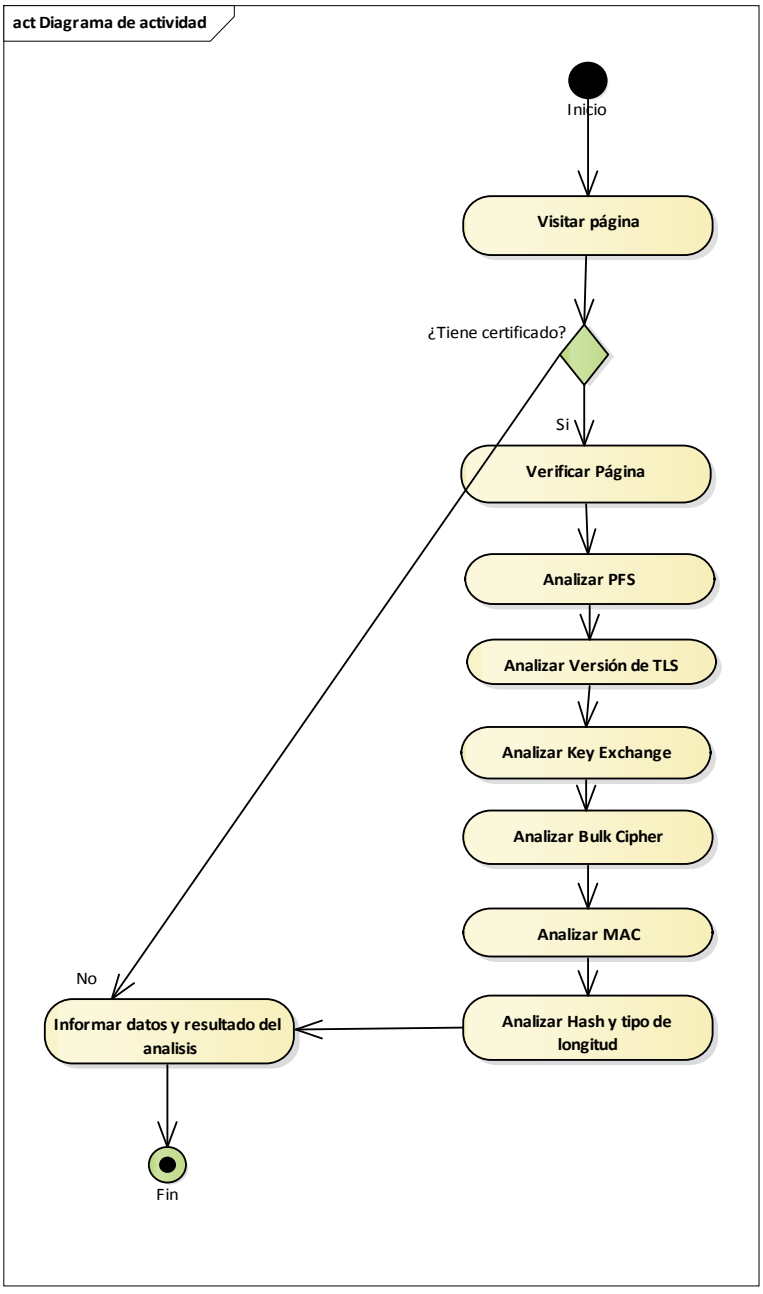


Fig. 2. Diagrama de modelo de clase de SSL validation

En la figura 2 se grafica lo que el sistema va a analizar de cada componente del certificado de seguridad.



Fig, 2. Diagrama de actividad del sistema .

4 Caso de estudio para validación de desarrollo

El caso de estudio seleccionado para validar la aplicación que se diseñó consiste en consultar una página web que tenga un certificado de seguridad no muy confiable. El link se tomó del sitio "anti-phishing", que cuenta con una base de datos de las páginas denunciadas como fraudulentas.

La selección de este caso de estudio es para demostrar que la aplicación funciona correctamente para todo sitio.

Agregar que es la figura 3

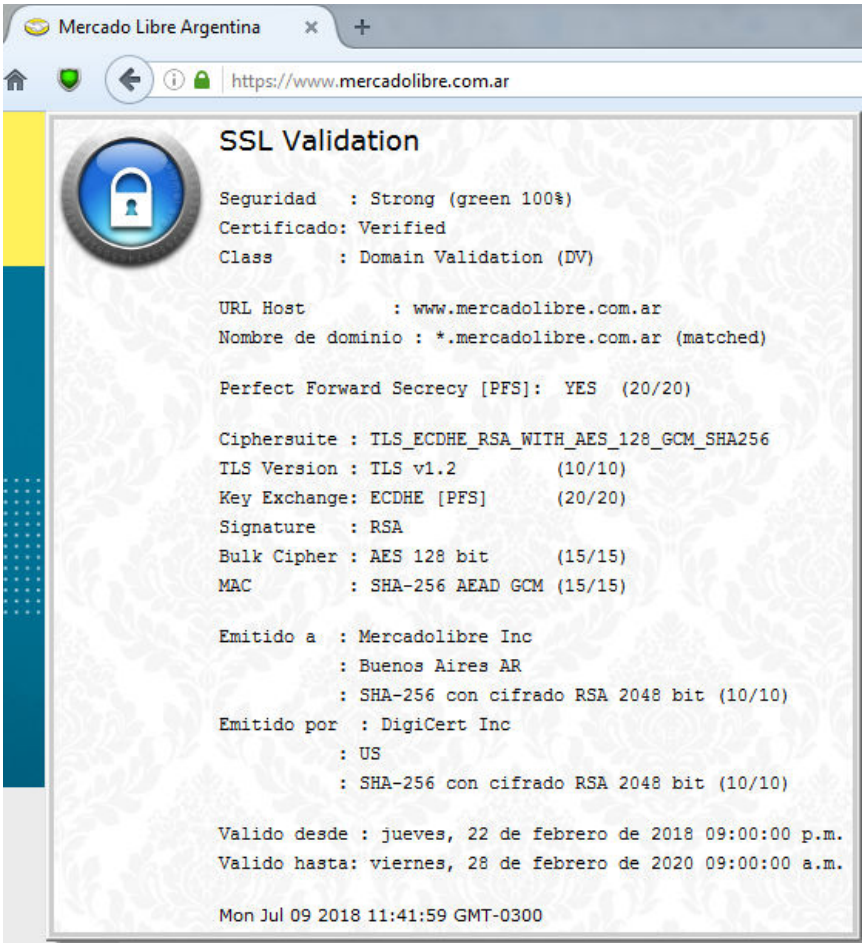


Fig. 3. Análisis y datos que contiene el certificado digital del sitio www.mercadolibre.com.ar

5 Conclusiones y futuras líneas de trabajo

Se arribó a la conclusión que el addons, funcionó de forma correcta indicando con un ícono de color rojo que la página tiene un bajo nivel de seguridad, e informando con que métodos criptográficos contaba el mismo. La conclusión es que no es confiable, dejando a decisión del usuario si confía los datos a dicha página o no.
Se ha presentado una revisión sistemática de páginas que realizan análisis, dando una puntuación final del análisis del certificado digital.

Se ha logrado de forma exitosa la construcción de una aplicación compuesta por un analizador de métodos criptográficos.

Como futuras líneas de trabajo se identifican:

- Ampliar la capacidad de analizar los certificados, de acuerdo con las nuevas medidas de seguridad informática que se apliquen, acorde a los nuevos estándares de seguridad para los certificados.
- Poder migrar el programa a otros browsers, ya sea Chrome, Internet Explorer y Safari.
- Actualizar el programa de acuerdo con las futuras versiones de Firefox que vayan apareciendo.

Referencias

1. Redacción de It Digital Security, en 2018 los esquemas de phishing serán mucho más sofisticados, It digiral security, miercoles 20 de diciembre de 2018 http://www.itdigitalsecurity.es/actualidad/2017/12/en-2018-los-esquemas-de-phishing-seran-mucho-mas-sofisticados
2. Redacción Byte TI, la mitad de los ataques de phishing son financiero, revista bye - España, lunes 26 de marzo de 2018. https://www.revistabyte.es/actualidad-byte/ataques-phishing-financiero/
3. Hartmann, Phishing bancario, la manipulación sigilosa y entradora para obtener datos, Diario Clarin - Argentina, domingo 19 de enero de 2018. https://www.clarin.com/suplementos/zona/phishing-bancario-manipulacion-sigilosa-entradora-obtener-datos_0_BkxH7iAEM.html
4. Talens-Oliag, Sergio, Introducción a los certificados digitales https://www.uv.es/~sto/articulos/BEI-2003-11/certificados_digitales.pdf
5. Talens-Oliag, Sergio, Introducción a la criptografia. https://www.uv.es/sto/articulos/BEI-2003-04/criptologia.html#id2448661
6. Pérez González Leo, Argentina, entre los países que más phishing reciben en el mundo, Diario Clarin - Argentina, martes 7 de noviembre de 2017 https://www.clarin.com/sociedad/argentina-paises-phishing-reciben-mundo_0_SkrEtz1kM.html
7. Redacción equipo de seguridad de n26. Revista n26 - España https://mag-es.n26.com/c%C3%B3mo-prevenir-un-ataque-de-phishing-en-tu-cuenta-bancaria-de17a79db18a
8. qualys ssl labs, SSl Server Test, https://www.ssllabs.com/ssltest/
9. Redalia, https://www.redalia.es/herramientas/comprobar-ssl/
10. SSL Checker, https://www.sslshopper.com/
11. ImmuniWeb® SSLScan, https://www.htbridge.com/ssl/

12. Parket Higgins, Pushing for Perfect Forward Secrecy, an Important Web Privacy Protection, Electronic Frontier Foundation, miercoles 28 de agosto de 2013 https://www.eff.org/deeplinks/2013/08/pushing-perfect-forward-secrecy-important-web-privacy-protection
13. Dante Odín Ramírez López, Carmina Cecilia Espinosa Madrigal, El cifrado web (SSL/TLS), Universidad Nacional Autónoma de México https://revista.seguridad.unam.mx/numero-10/el-cifrado-web-ssltls
14. David Dunning,¿Como funciona el "intercambio de clave encriptada"?, Diarion, La voz de Houston, Estados Unidos. https://pyme.lavoztx.com/cmo-funciona-el-intercambio-de-clave-encriptada-5765.html
15. IBM, Bulk ciphers https://www.ibm.com/support/knowledgecenter/en/SSB23S_1.1.0.13/gtps7/bulkcip.html
16. Departamento de Ingeneria de Sistemas Telematicos - ETSIT - UPM - España https://www.dit.upm.es/~pepe/401/2911.htm#!-alone

Track Gobierno Digital y Ciudades Inteligentes (TGDCI)

Coordinadores

Elsa Estevez (UNS) - Ariel Pasini (UNLP)

Temas

Administración de cambios en instituciones públicas
Administración de conocimiento en gobierno
Administración de interesados
Administración de registros públicos digitales
Arquitecturas empresariales gubernamentales
Asociaciones público-privadas
Automatización de procesos administrativos gubernamentales
Calidad de datos abiertos
Censura y restricción de datos abiertos
Ciberseguridad en gobierno
Co-creación de datos abiertos para beneficios sociales
Compartir información de gobierno
Computación en la nube gubernamental
Diseño y entrega de servicios públicos digitales
Estándares de metadatos para datos abiertos
Infraestructura tecnológica para datos abiertos
Innovación con datos abiertos
Integración de procesos administrativos gubernamentales
Marcos de interoperabilidad
Portales de información de gobierno
Protección de datos personales
Reingeniería de procesos administrativos gubernamentales
Seguridad de la infraestructura de información de gobierno
Servicios compartidos
Sistemas y servicios de energía inteligente
Sistemas y servicios de movilidad inteligente
Sistemas y servicios para ciudades inteligentes
Sistemas y servicios para gestión del agua
Sistemas y servicios para protección del medio ambiente
Soluciones para participación ciudadana
Tableros de control para gestión municipal

Modelos de evaluación de gobiernos abiertos, aplicado a los municipios de la provincia de Buenos Aires

Ariel Pasini[1] , Juan Santiago Preisegger[2] , Patricia Pesado[3]

Instituto de Investigación en Informática LIDI (III-LIDI)*
Facultad de Informática – Universidad Nacional de La Plata 50 y 120 La Plata Buenos Aires
*Centro Asociado Comisión de Investigaciones Científicas de la Pcia. de Bs. As. (CIC)

{apasini,jspreisegger,ppesado}@lidi.info.unlp.edu.ar

Abstract.
Con el aumento de políticas relacionadas al gobierno digital y los ciudadanos más exigentes con sus gobiernos, se generó un nuevo paradigma de la gestión pública, que se dio a llamar *gobierno abierto*, la cual se sustenta en tres principios básicos: *Transparencia*, *Colaboración* y *Participación*. Estos se reflejan en múltiples estrategias diferentes, en función del país, provincia o ciudad en la que se adopta, pero con dos ejes comunes: La *apertura de datos públicos* y *la apertura de procesos*. En el último tiempo se impulsaron diversas iniciativas a diferentes niveles de gobierno, para que todos los ciudadanos puedan acceder a la información pública. En ese contexto, este documento propone un modelo de evaluación del estado de avance en dicho paradigma, en un grupo preseleccionado de municipios de la Provincia de Buenos Aires, a través de las distintas herramientas relacionadas al gobierno abierto que ofrecen en sus sitios web a sus ciudadanos para el mejoramiento de sus servicios.

Keywords: Ingeniería de Software, Gobierno Abierto, Datos Abiertos, Portales de Gobierno.

1 Introducción

En la actualidad, la sociedad posee exigencias cada vez más elevadas para con sus gobernantes y sus entes de gobierno. Entre estas exigencias se incluye la transparencia y el manejo eficiente de los bienes públicos, la inclusión en la toma de decisiones y la colaboración de la sociedad. En función de estos requisitos y, con la ayuda de nuevas tecnologías, se generó una nueva forma gobierno, incluyendo más al ciudadano, permitiéndole generar aportes a las políticas públicas y participar en la toma de decisiones del gobierno [1]. A esta se la denomino *Gobierno Abierto* y se la define como "… *una plataforma tecnológica institucional que convierta los datos gubernamentales en datos abiertos para permitir su uso, protección y colaboración por parte de los ciudadanos en los procesos de decisión pública, rendición de cuentas*

[1] ORCID: 0000-0002-4752-7112
[2] ORCID: 0000-0001-5282-3269
[3] ORCID: 0000-0003-0000-3482

y mejoramiento de los servicios públicos." [2]. Esta se apoya en tres principios bien definidos: **Transparencia**, **Colaboración** y **Participación**. [3][4] A estos principios se los puede abordar desde distintas perspectivas y su aplicación varía en base al interés real que posean los gobernantes en fomentar el concepto de gobierno abierto, esto hace muy complejo definir un modelo que permita realizar la evaluación de los mismos que sea extrapolable a cualquier estructura gubernamental.

Tomando como base al ecosistema planteado por a Álvaro V. Ramírez Alujas y Nicolás Dassen [5], se generó un modelo de evaluación que permitió analizar los sitios web de un conjunto de municipios de la provincia de Buenos Aires y calificar el estado de apertura de los datos y sus procesos.

El modelo consta de ocho atributos (2 para transparencia, 3 para participación y 3 para colaboración) que pueden tomar valores discretos pre-definidos en función de la información relevada en los sitios web de los municipios.

La provincia de Buenos Aires posee 135 municipios, en un principio se tomó una muestra en función a su población. Este criterio solo permitía evaluar las grandes ciudades y no era representativa de la provincia, para lograr una muestra más homogénea se tomaron los cinco municipios con mayor población de cada sección electoral. Luego se aplicó el modelo de evaluación propuesto a los 36 municipios seleccionados para determinar el avance de la provincia respecto a los tres principios del Gobierno Abierto.

En la sección 2 se presentan conceptos generales de Gobierno Abierto. En la siguiente sección se describe el modelo de evaluación. En la cuarta la aplicación del modelo a los municipios seleccionados y el análisis de los resultados obtenidos y por último se presentan las conclusiones obtenidas.

2 Conceptos de gobierno abierto

Los portales de gobierno, últimamente, se han utilizado como una herramienta para la difusión de información financiera en vías de generar una imagen de transparencia gubernamental. Esta práctica, comúnmente se la confunde con el concepto de "*gobierno abierto".* Pero la transparencia no es solo una característica de los gobiernos abiertos, sino que también es la consecuencia de políticas de gobierno abiertas que generan información a disposición de la ciudadanía. [6] [7]

2.1 Gobierno abierto

Para definir el concepto de "*gobierno abierto*", primero debemos analizar el concepto de "*gobierno*", que se define como la *acción y efecto de gobernar,* lo que deriva en administración de recursos públicos según diversas leyes o normativas que alcancen el gobierno en cuestión. Siguiendo este enfoque el concepto de "*abierto*" lo asociamos a la liberación de la información relacionada a administración de los recursos públicos. Esta liberación que se espera del concepto, se plantea en dos planos, por un lado los datos y por otro los proceso de gobierno. Es decir como un sistema complejo

de procesos tecnológicos para gestionar la información por parte del gobierno y la comunidad.

En [2] Sandoval lo define: "*El gobierno abierto debe entenderse como una plataforma tecnológica institucional que convierta los datos gubernamentales en datos abiertos para permitir su uso, protección y colaboración por parte de los ciudadanos en los procesos de decisión pública, rendición de cuentas y mejoramiento de los servicios públicos."*

De la definición se desprende que el *gobierno abierto* no es solo "datos" en un sitio web para que promocionar la transparencia del gobierno, sino que debe proveer tecnologías de información y comunicación (incluidas las redes sociales) para garantizar la protección de los datos y promover la participación por parte de la comunidad. [6][7][8]

2.2 Principios del gobierno abierto

En [5] se presenta una nueva aproximación de los principios propuestos por Obama en [9].

Transparencia: *Un gobierno transparente proporciona información sobre lo que está haciendo, sobre sus planes de actuación, sus fuentes de datos y sobre aquello de lo que puede responsabilizárselo frente a la sociedad. Ello fomenta y promueve la rendición de cuentas de la administración ante la ciudadanía y un permanente control social.*
Participación: *Un gobierno participativo promueve el derecho de la ciudadanía a colaborar activamente en la formulación de políticas públicas y facilita el camino para que las administraciones públicas se beneficien del conocimiento, de las ideas y de la experiencia de los ciudadanos. Promueve la creación de nuevos espacios de encuentro que favorezcan el protagonismo y la implicación de los ciudadanos en los asuntos públicos.*
Colaboración: *Un gobierno colaborativo compromete e implica a los ciudadanos y demás agentes sociales en el esfuerzo por trabajar conjuntamente para resolver los problemas nacionales. Ello supone la cooperación y acciones coordinadas no sólo con la ciudadanía, sino con las empresas, asociaciones y demás agentes. Asimismo, potencia el trabajo combinado dentro de las propias administraciones, y entre ellas y sus funcionarios, de manera transversal*.

La aplicación de estos conceptos, se basa en dos ejes fundaméntales:

1. ***Apertura de los datos****:* utilización de datos públicos (*transparencia*) para la innovación, generación de nuevas aplicaciones transformando a los gobiernos en gestores de servicios (permitiendo a la comunidad generar nuevos servicios que agregan valor).
2. ***Apertura de los procesos***: facilitando la comunicaron para la toma de decisiones abiertas (*participación*), conocimiento y experiencia de los ciudadanos en el diseño de políticas públicas (*colaboración*).

Los mismos autores [5] plantean un ecosistema entre los tres principios para el desarrollo de un gobierno abierto, el mismo se puede ver en la Figura 1 a continuación.

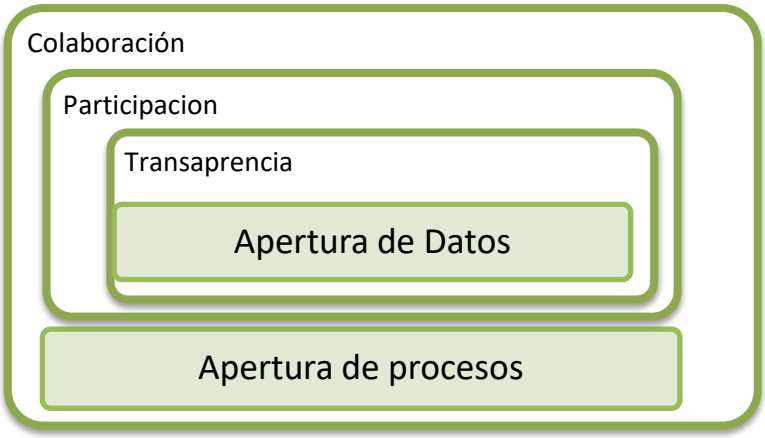


Fig.1. Vinculación entre los tres principios de Gobierno Abierto

2.3 Portal de gobierno abierto

Un *portal* de es un sitio web que ofrece al usuario, de forma fácil e integrada, el acceso a una serie de recursos y de servicios relacionados a un tema en particular. Un "*Portal de Gobierno abierto*" es un sitio web que ofrece a los ciudadanos una serie de recursos y servicios relacionados con la apertura de datos. Es una parte fundamental en el proceso de apertura de datos. Son plataformas digitales que sirven para almacenar, compartir, conectar y visualizar bases de datos [10].

2.4 Catálogo de datos

El catálogo de datos es la parte central de un portal de datos abiertos y contiene un listado de todas las tablas publicadas con una descripción del contenido de la base de datos, el nombre de la agencia responsable, la frecuencia de actualización, el número de veces que se ha visitado, la información técnica para conectarla con aplicaciones informáticas y un espacio para los comentarios de los usuarios. Algunos portales incorporan también un área de valoración de la calidad de la base de datos. [10]

3 Modelo de evaluación del gobierno abierto

Con el objetivo de hacer un relevamiento de los municipios de la provincia de Buenos Aires, se definió un modelo basado en los tres principios mencionados anteriormente que se puede observar en la Tabla 1. Para cada uno, se establece un conjunto de atributos y un conjunto de valores que pueden tomar. La ***transparencia*** se relaciona con la posibilidad de acceder a los datos y el nivel de automatización del

medio que los pone a disposición, en consecuencia, se definió el atributo *Nivel de visualización de datos y Nivel de presentación de la información.* En el caso de la ***participación*** se propuso medir la posibilidad de contar con la participación ciudadana en propuestas concretas, en la participación en el diseño de propuestas políticas y los debates públicos, para lo cual se definieron los atributos de: *Nivel participación ciudadana, Nivel de participación en el diseño de políticas públicas y Nivel de debates públicos.*

Tabla 1. Atributos y valores del modelo de evalaucion

	Atributo	*Descripción*	*Valores*	
Transparencia	Nivel de visualización de datos - NVD	Categoriza el formato de los datos.	0- 1- 2- 3- 4- 5-	No hay datos disponibles. Disponibles en la Web, en cualquier formato (PDF). Estructurados propietario, (Excel). En formato no propietario, (CSV). Utilizar URI (Uniform Resource Identifier). Vincularlos a otros datos para proporcionar un contexto (Junar).
	Nivel de presentación de la información – NPI	Categoriza la forma en la cual se presentan los datos.	0- 1- 2-	No posee un portal/catálogo de datos abiertos; Posee un catálogo de datos abiertos. Posee un portal de datos abiertos.
Participación	Nivel participación ciudadana - NPC	Categoriza la posibilidad de intervenir en una toma de decisión (por ejemplo presupuesto participativo)	0- 1- 2-	No permite participación. Solo posee información sobre posibilidades de participar en la toma de decisiones (participación presencial). Permite participas de forma remota.
	Nivel de participación en el diseño de políticas públicas - NPDPP	Categoriza las políticas que fueron diseñados en base a los aportes ciudadanos	0- 1- 2-	No hay evidencia de la participación ciudadana en el diseño de políticas públicas. Hay evidencia de que alguna política pública se originó desde los aportes ciudadanos. Hay evidencia de que alguna política pública se originó desde aportes y participación ciudadanos.
	Nivel de debates públicos NDP	Categoriza la apertura del gobierno a debatir políticas con la ciudadanía.	0- 1- 2-	No posee información de los debates públicos. Posee información de los debates públicos (presenciales). Posee un foro de discusión para realizar un debate público.
Colaboración	Nivel de Interacción - NI	Categoriza la interacción con el ciudadano.	0- 1- 2- 3-	No posee información de los servicios. Posee una guía de trámites o información sobre los servicios. Permite realizar operaciones básicas de comunicación. Permite realizar operaciones básicas de transacción.
	Nivel de Vinculación con PyMEs – NVP	Categoriza la vinculación con la PyMEs desde las perspectivas de la asistencia a PyMEs/ licitaciones/ bolsas de empleo privado	0- 1- 2- 3-	No posee vinculación con PyMEs Posee solo información dirigida a PyMEs (asistencia a PyMEs/ licitaciones/ bolsa de empleo) Permite inscripciones online a los: planes de asistencia a PyMEs, licitaciones y/o bolsa de empleo Permite gestionar todo el proceso de asistencia a PyMEs, las licitaciones y/o bolsa de empleo de forma online
	Nivel de aporte ciudadano - NAC	Categoriza el tratamiento de: Sugerencias, Observaciones, Requerimientos y Reclamos realizados por los ciudadanos	0- 1- 2- 3-	No permite aportes. Permite ingresar aportes por parte de los ciudadanos. Permite ingresar aportes por parte de los ciudadanos y realizar el seguimiento del mismo. A partir del aporte ciudadano se generó un plan político.

Para la ***colaboración*** se tuvo en cuenta la participación de la ciudadanía y la vinculación con organizaciones privadas, en particular PyMEs, que pueden tener un desarrollo comercial en el municipio. Como principal entrada de la participación se analizó la interacción de un ciudadano en el portal de municipio, las sugerencias, observaciones, requisitos y reclamos que hace la comunidad al municipio y si hay evidencia que alguno de estos se convirtió en una política del organismo, además dentro

de la colaboración se incluyó la vinculación de las PyMEs, si existen propuestas de asistencia a las PyMEs por parte del municipio, si hay información sobre la licitaciones y ofertas laborales de las empresas a la comunidad. Para lo cual se definieron los atributos de *Nivel de Interacción*, *Nivel de Vinculación con PyMEs y Nivel de aporte ciudadano*.

Para llegar a un valor de *transparencia (T)*, se denota una dependencia muy fuerte entre los dos atributos definidos, en el caso que el valor de NPI sea 0 y el NVD sea 0 significa que no hay datos disponibles y no poseen portal o catalogo por lo tanto la transparencia es 0. En el caso que NPI sea 1, es decir que hay un catálogo de datos abierto, el valor de NVD lo podemos diferenciar entre si los datos son procesable o no. Por último en el caso que NPI sea 2, quiere decir que obligatoriamente tengo un nivel 4-5 de NVD. Esto se puede observar en la Tabla 2.

Tabla 2. Tranparencia

T	*NPI*	*NVD*
0	0	0
1	0	1-5
2	1	1-3
3	1	4-5
4	2	4-5

Tabla 3. Participacion

P	*NPC*	*NPDPP*	*NDP*
0	0	-	-
1	1	0-1	0-1
2	1	1-2	1
3	2	1-2	0-1
4	2	1-2	2

Tabla 4. Colaboracion

C	*NI*	*NVP*	*NAC*
0	0	-	-
1	1	1	-
2	1	2	1-2
3	2	2-3	2-3
4	3	3	3

En el caso de la *participación (P)* Si el valor de NPC es 0, no importa el valor de las otras dos variables ya que no se considera que haya participación. Si el valor de NPC es 1 (participación presencial) o 2 (participación remota) se analiza si además hay participación o aportes en el diseño de políticas públicas (NPDPP) y, por último si se aplica la metodología de debate público. En este caso el nivel más alto está caracterizado por aquellos municipios que brindan el máximo nivel de participación a sus ciudadanos permitiéndoles generar debates públicos mediante un foro, generando políticas públicas desde sus aportes y votándolas de manera remota. Esto se puede observar en la Tabla 3.

Para la *colaboración (C)* se observa el nivel de interacción del ciudadano (NIC), si el valor es 0 ya descartamos los otros valores, ya que no se considera que haya colaboración. En el caso de registrar interacción por parte del ciudadano se analizan los valores del Nivel de Vinculación con PyMEs (NVP) y el Nivel de aporte ciudadano (NAC), los cuales miden la colaboración del gobierno con empresas del sector privado y con los ciudadanos de a pie. En este caso el nivel más alto está caracterizado por aquellos municipios que permiten operaciones de transacción con las partes, permitiéndoles generar un plan político a través de sus aportes y tomando parte activa en la generación y vinculación de empleos con las PyMEs. Esto se puede observar en la Tabla 4.

4 Evaluación de municipios

4.1 Criterios de selección de municipios.

La Provincia de Buenos Aires posee una población de 15.625.000 de habitantes lo que representa casi un 40 % de la población de Argentina (censo 2010). Posee una superficie de 307.571 km, que se divide en 135 municipios con autonomía de gobierno. Con una densidad poblacional entre los municipios que varía entre 1 y 9167 Habitantes por Km^2. La Provincia se divide en ocho secciones electorales. Para tomar una muestra representativa de los municipios de la Provincia, e tomaron para cada sección electoral los 5 municipios con mayor porcentaje de población respecto a la población total de Buenos Aires.

En la Tabla 5, que se encuentra a continuación, se puede observar los municipios seleccionados.

Tabla 5. Municipios seleccionados

Sección	*Municipios*
1	General San Martín, Merlo, Moreno, San Miguel y Tigre
2	Baradero, Pergamino, San Nicolás, San Pedro y Zárate
3	Almirante Brown, La Matanza, Lanús, Lomas de Zamora y Quilmes.
4	Chacabuco, Chivilcoy, Junín, Nueve de Julio y Trenque Lauquen.
5	Balcarce, General Pueyrredón, La Costa, Necochea y Tandil.
6	Bahía Blanca, Coronel de Marina Leonardo Rosales, Coronel Suárez, Tres Arroyos y Villarino
7	Azul, Bolívar, Olavarría, Saladillo y Veinticinco de Mayo
8	La Plata

A estos se les evalúo el sitio web en función de modelo expuesto anteriormente.

4.2 Aplicación del modelo

Se evaluaron los sitios web de los 36 municipios, recorriendo las diferentes secciones observando si cumplían con los atributos planteados en el modelo.

Se procesaron los datos, obteniendo en *transparencia* un 56% (20 municipios) que se encuentra en nivel 1, es decir, posen datos abiertos en cualquier formato, pero carecen de un catálogo o portal que los presente de forma estructurada. Le siguió un 17% sin datos abierto. Como contraparte hay un 14% (5 municipios) que posee portales de datos abierto, ubicándose en la vanguardia. En cuanto a la *participación* un 78% que no permite participación. Seguido por un 14% solo posee información sobre mecanismo de participación. El nivel 4 lo integra solamente un 6% (2 municipios). Por ultimo en el mayor grado de participación se encuentra un 3% (1 municipio). Para la *colaboración* el 31% (11 municipios) no presenta ningún mecanismo de colaboración. Un 3% posee cierta colaboración con el sector privado. Un 11% (4 municipios) posee colaboración con los ciudadanos y los privados en un nivel aceptable. Por último, un

56% (entre nivel 3 y 4) permiten mecanismos de colaboración con ambos sectores mediante comunicación y transacciones con estos.

En la Tabla 6, que se encuentra a continuación, pueden observarse los valores individualizados.

Tabla 6. Resultado de la evaluación

Nivel	*Transparencia*		*Participación*		*Colaboración*	
0	6	17%	28	78%	11	31%
1	20	56%	5	14%	1	3%
2	4	11%	0	0%	4	11%
3	1	3%	2	6%	11	31%
4	5	14%	1	3%	9	25%

Como se puede observar, la aplicación del Gobierno Abierto es un área de particular interés para los gobiernos y la sociedad, que se está empezando a implementar en la provincia. Sin embargo, cabe destacar que el puntapié inicial se realiza desde la *transparencia*, donde se observa que gran parte de los municipios relevados está tomando medidas para consolidar este principio. Respecto a la *Colaboración* se puede observar el incremento respecto a la falta de aplicación de este principio, como así también, el avance en su consolidación, ya que cuando se lo aplica es en los niveles superiores. Como contraposición está el caso de la *Participación*, en este principio es en el que los municipios están menos avanzados o presentan mayor desinterés. Se puede observar que cerca del 80% de los municipios no aplican ninguna medida para la participación ciudadana y, tan solo el 9% de los mismos la aplica en los dos puntos más altos de la escala.

En la Figura 2, que se encuentra a continuación, se pueden observar los resultados con mayor descripción.

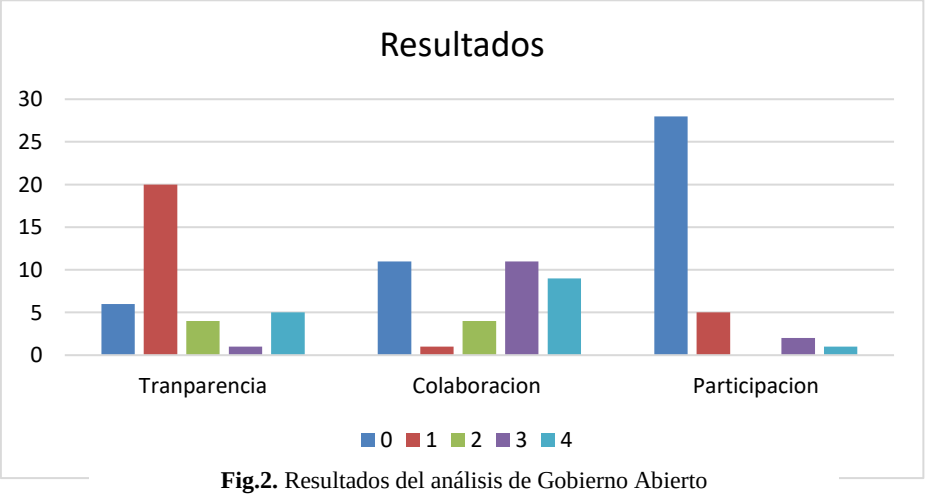


Fig.2. Resultados del análisis de Gobierno Abierto

5 Conclusiones

A lo largo del artículo se introdujeron los conceptos básicos de gobierno abierto, haciendo hincapié en los principios de *transparencia*, *participación* y *colaboración*.

En función de esos principios se diseñó un modelo de evaluación compuesto por ocho atributos.

Se tomaron como caso de estudio 36 municipios de la provincia de Buenos Aires, seleccionados por su población respecto a la sección electoral a la que pertenecían.

El relevamiento de los municipios brindo la posibilidad de analizar el estado en el que se encuentra la provincia de Buenos Aires, respecto a la aplicación de los tres principios de Gobierno Abierto, concluyendo que se encuentra abordando el tema de diversas formas, inicialmente por las acciones relacionadas a la transparencia, luego abordando la Colaboración y Participación. Los municipios más avanzados brindan servicios de transferencia a sus ciudadanos integrándolos en la toma de decisiones y generando impacto positivo en la sociedad y su calidad de vida.

6 Agradecimientos

Este trabajo de investigación fue financiado por el Proyecto PIT-AP “Herramientas para el desarrollo y la entrega de servicios públicos digitales de acción social para municipios bonaerenses”.

7 Referencias

1. Soon Ae Chun, Stuart Shulman, Rodrigo Sandoval and Eduard Hovy. (2010). Government 2.0: Making connections between citizens, data and government. City University of New York, College of Staten Island
2. J. Ramón Gil-García, J. Ignacio Criado y Juan Carlos Téllez. Tecnologías de Información y Comunicación en la Administración Pública: Conceptos, Enfoques, Aplicaciones y Resultados. Infotec
3. César Calderón, Sebastián Lorenzo. (2010). OPEN GOVERNMENT: Gobierno Abierto. Algón Editores
4. Alejandra Naser, Álvaro Ramírez-Alujas. (2017). Plan de gobierno abierto Una hoja de ruta para los Gobiernos de la región, Libros de la CEPAL.
5. Nicolás Dassen y Juan Cruz Vieyra. (2012). Gobierno abierto y transparencia focalizada: Tendencias y desafíos para América Latina y el Caribe. BID
6. Álvaro V. Ramírez-Alujas. El Gobierno Abierto Y Los Desafíos Tecnológicos En Latinoamérica. Goberna, América Latina - Instituto Universitario de Investigación Ortega y Gasset
7. César Nicandro Cruz-Rubio. (2015). ¿Qué es (y que no es) gobierno abierto? Una discusión conceptual. Eunomía. Revista en Cultura de la Legalidad.
8. Sánchez González, José Juan. (2015). La participación ciudadana como instrumento del gobierno abierto. Revista Espacios - Universidad Autónoma del Estado de México.

9. Barack Obama. (2009). Transparency and Open Government: Memorandum for the Heads of Executive Departments and Agencies.
10. Antonio Moneo. https://blogs.iadb.org/abierto-al-publico/2014/11/20/que-es-un-portal-de-datos-abiertos-y-para-que-sirve/. Portal del BID (2014)

Gestor de contenidos orientado a portales de organismos gubernamentales

Ariel Pasini [1] , Rocío Muñoz [2] , Patricia Pesado [3]

Instituto de Investigación en Informática LIDI (III-LIDI)*
Facultad de Informática – Universidad Nacional de La Plata 50 y 120 La Plata Buenos Aires
*Centro Asociado Comisión de Investigaciones Científicas de la Pcia. de Bs. As. (CIC)

{apasini,rmunoz,ppesado}@lidi.info.unlp.edu.ar

Abstract. Se presenta una herramienta para la generación de portales web para diferentes organismos gubernamentales. La misma permite brindar, a un conjunto de receptores, la cartera de servicios que el organismo pone a su disposición. Se establece un formato que debe respetarse en los portales, teniendo en cuenta factores tales como la presencia institucional y el contenido de los servicios brindados. La herramienta presentada permite que usuarios sin conocimientos de programación web puedan crear, modificar y personalizar portales para las diferentes organizaciones. Para este trabajo se tomaron tres casos de estudio: un prototipo de una municipalidad, uno de una institución educativa pública y uno de una secretaría específica que forma parte de una municipalidad.

Keywords: Gobierno electrónico, Gestor de contenidos, Portales gubernamentales, Organismos gubernamentales.

1 Introducción

Desde la aparición de internet los gobiernos han ido ganando posiciones en la web, desde la simple presencia con información estática hasta complejos portales que permiten realizar operaciones, participar en las decisiones y construir medidas de gobierno en conjunto.

Un sitio web es un conjunto de páginas asociadas por un eje de información, la cual puede ser presentada utilizando texto simple o diferentes formatos multimedia. Generalmente se divide a los sitios web en estáticos y dinámicos, dependiendo de la forma que se genera la información que presentan. Los portales web [1] tienen la característica de reunir accesos a diferentes sitios web de temas relacionados de forma estructurada, funcionando como punto de entrada principal a los diferentes servicios que pueden dar los diferentes sitios web en cuestión, proveyendo contenido informativo y transaccional de mayor calidad y ajustándose sobre todo a las necesidades de los usuarios.

Un gestor de contenidos [2] es una aplicación web que permite la creación, administración y publicación de contenido en una página web. Permite realizar todas

[1] ORCID: 0000-0002-4752-7112
[2] ORCID: 0000-0001-6521-0568
[3] ORCID: 0000-0003-0000-3482

estas acciones sin tener conocimientos avanzados, mediante la utilización de plantillas fácilmente personalizables. No es necesario tener conocimientos de programación.

Una de las características que permiten los gestores de contenido es la estandarización de la información. Una vez definido un formato/plantilla, los elementos que se agreguen mantendrán la estructura documental de la plantilla seleccionada.

Generar un gestor de contenidos para portales gubernamentales facilita a todas las dependencias la prestación de sus servicios públicos, respetando la presencia institucional del organismo y permitiendo exponer los servicios que ofrece de una forma consistente y estructurada. Facilita al usuario o receptor del servicio encontrar la información necesaria para acceder a los mismos de forma sencilla.

CMS-Gov es un gestor de contenido que permite generar portales orientados a instituciones gubernamentales, definiendo la imagen institucional, la administración de los receptores de los servicios y la descripción estandarizada de los mismos. De forma tal que el usuario, al ingresar al portal, se encuentra en el contexto de la organización (imagen institucional), puede ver los servicios que se proveen, filtrarlos de acuerdo a los receptores y, para cada uno de estos servicios, ver los datos de forma estandarizada.

Con la herramienta se generaron prototipos de tres portales de diferentes organismos gubernamentales como ejemplo de aplicación.

En la sección dos se encuentran definiciones sobre portales gubernamentales, la importancia de la presencia institucional y de la estructuración de la información que brindan sobre los servicios. En la tercera, se describen las características del CMS-Gov, su estructura y forma de configuración. Luego un conjunto de prototipos de aplicación en diferentes organismos gubernamentales, y finalmente las conclusiones del presente trabajo.

2 Portales gubernamentales.

Un organismo gubernamental es una institución estatal cuya administración está a cargo del gobierno de turno. Su finalidad es brindar ***servicios públicos*** que resulta necesario para la ciudadanía. Los portales de gobierno facilitan a los diferentes receptores el acceso a los ***servicios públicos*** que presta.

Un "***servicio público***" se define como: "*Actividad llevada a cabo por la Administración o, bajo un cierto control y regulación de esta, por una organización, especializada o no, y destinada a satisfacer necesidades de la colectividad*.". Los receptores de los servicios se clasifican en cinco grupos: 1- *Ciudadanos*, 2- *Empresas*, 3- *Gobiernos*, 4-*Visitantes*, 5-*Empleados*. Cada receptor tendrá un grupo de servicios disponibles, para simplificar su uso, los portales deben estandarizar la estructura, ubicación y contenido de los mismos. Además de normalizar el contenido, se debe respetar la estética y presentación con el fin de mantener la imagen de la organización que el portal representa.

Un sitio web o portal de una organización gubernamental [3-4] es una herramienta de importancia estratégica para la misma y para las autoridades y funcionarios que la conforman, así como un canal efectivo y estratégico de prestación de servicios y de comunicación externa e interna.

Un portal bien planteado, que contenga información idónea, actualizada y que cuente con un diseño de secciones adecuado, le dará a la organización la posibilidad de generar una buena imagen y será valorado positivamente por los ciudadanos. El portal no debería contener información acerca de las actividades personales del titular de la dependencia, ni sobre sus gustos y preferencias. Tampoco deberá mostrar material de algún partido o corriente política.

2.1 Presencia Institucional de los portales gubernamentales.

Logo/Imagen.

En una organización gubernamental es muy importante tener una identificación de la misma, algo que permita diferenciarla de otras organizaciones y que la defina. Es necesario contar con un logo o una imagen que represente a la organización unívocamente en el portal. La fuerza y poder que proporciona el logo a una organización es tanta que solo con verlo sabemos de quién estamos hablando.

Una imagen corporativa se crea para ser atractiva al público, de modo que provoque un interés en los consumidores. Puede estar compuesta por uno o más elementos, que de manera conjunta o independiente cumplen una misma función: permitir a los usuarios reconocer a la organización.

Colores/Estilos.

Una de los aspectos más importantes de una organización gubernamental es la importancia de la elección de sus colores. El color es el que va a representar a la organización, está presente en el logo, los carteles, los folletos, etc. Los colores son la forma de comunicación no verbal más poderosa de la que se dispone. Es una carta de presentación, ya que produce un impacto inicial en los clientes o consumidores.

La elección de las tonalidades es muy importante, se deben evitar tonalidades estridentes, desordenadas y distorsionadas. Las tonalidades deben ser armoniosas, se debe elegir una gama de colores que combinen y contrasten entre sí, que faciliten una rápida lectura y a la vez reflejen la identidad corporativa que se quiere proyectar.

En un portal gubernamental la tipografía elegida es muy importante. La tipografía es la ciencia de arreglar las letras de manera tal que el texto escrito sea agradable y fácil de leer. Incluye varios factores, como por ejemplo la fuente, el tamaño, el color, etc.

El tamaño de la fuente debe ser lo suficientemente acorde como para que sea fácil de leer tanto en una computadora de escritorio como en un dispositivo móvil. El color de la fuente va a depender, en gran medida, de los colores elegidos para el resto del portal y también de los colores originales de la organización. Los sitios web de las organizaciones deben ser representaciones en Internet de las mismas.

Se debe tener en cuenta, además, la importancia de delimitar correctamente los títulos, párrafos, espacios entre ellos, etc. Sin una estructura bien definida el texto se ve poco amigable y resulta incómodo de leer para los usuarios.

Accesibilidad/Portabilidad.

La gran mayoría del poder cerebral está enfocado en la vista. Los humanos somos criaturas visuales, es por eso que a la hora de crear un portal web es muy importante el efecto que se tiene sobre el usuario.

La consistencia es primordial. Muchos de los diseños visuales afectan al usuario a nivel subconsciente. Mantener la simetría, diseños, fuentes, colores, espaciado entre líneas, etc. ayuda a que los usuarios se sientan cómodos y se queden. Los cambios esporádicos en la tipografía, íconos, ubicación de elementos y patrones establecidos brindan la impresión de desorganización y ociosidad, mientras que la consistencia luce y da la sensación de una página más profesional y de confianza.

Es necesario tener en cuenta el concepto de aprendizaje intuitivo, es decir, otorgarle al usuario una interfaz sencilla que puedan aprender con rapidez y poco esfuerzo, brindar la sensación al usuario de que ya lo ha utilizado anteriormente.

Teniendo en cuenta la usabilidad del portal, un aspecto que se debe considerar es el vinculado al número de clics que se deben dar hasta llegar a la información que se busca. Existe una regla, llamada la regla de los tres clics, que sugiere que todo usuario debería poder acceder a toda la información de la página en tan solo tres clics desde la página de inicio para evitar llegar a una situación de frustración.

Se debe tener en cuenta, también, que el diseño web del portal permita que personas con alguna discapacidad puedan percibir, entender, navegar e interactuar con el mismo. Sin embargo, la accesibilidad web beneficia no solo a personas con discapacidades sino también a, por ejemplo, personas de edad avanzada que han visto mermadas sus habilidades a consecuencia de la edad, personas con alguna incapacidad transitoria como puede ser un brazo roto, e incluso también a organizaciones o personas sin discapacidad. La web es un recurso importante para diferentes aspectos de la vida y es necesario que sea accesible para proporcionar un acceso equitativo e igualdad de oportunidades a todas las personas.

Resulta necesario que el acceso a la web sea posible desde cualquier lugar, con cualquier navegador, en cualquier momento y utilizando cualquier dispositivo, ya sea móvil o de escritorio.

2.2 Contenido de los portales gubernamentales.

Los portales gubernamentales deben estar orientados a los receptores de los servicios, y deben contener tres aspectos importantes: 1) Información, 2) Integración e 3) Interacción. [5]

- Información: es necesario adecuar el portal a partir de la información para asegurar que los receptores obtengan lo que buscan de forma rápida y sencilla. Se debe organizar el portal a partir de secciones generales complementadas con secciones específicas que entreguen información para cada receptor. Los documentos publicados deben estar en secciones acordes que faciliten el acceso a los ciudadanos, para lo cual es recomendable asegurar secciones básicas que se mantengan en todas las páginas del portal. El lenguaje del portal debe estar acorde a los receptores, para que sea fácil de leer y entender por los interesados.
- Integración: las secciones del portal deben mantener coherencia en cuanto a su contenido y diseño. Es importante diferenciar cada sección (receptores, navegación, información, etc.) y mantener uniforme la distribución de los componentes en cada una de ellas. Se debe mantener, además, la “imagen visual del organismo”.

- Interacción: los portales deben proponer una interacción con los prestadores de los servicios, la interacción puede ir desde un contacto telefónico o mail hasta la realización de trámites online.

Receptores.

Los receptores de los servicios varían para cada organismo gubernamental. Por ejemplo, en un municipio podemos encontrar: *Ciudadanos*, *Empresas*, *Gobiernos*, *Empleados*. Mientras que en el ámbito de un gobierno de una universidad pública los receptores serán los *Alumnos, Docentes, No docentes, Graduados*, etc.

Esta clasificación permite agrupar los servicios del portal en función a las necesidades de los interesados.

Catálogo de servicios.

El catálogo de servicios lista los servicios disponibles por ente gubernamental. Por ejemplo, en el caso de una municipalidad, se pueden dar servicios para la habilitación de un comercio, la tramitación de una tasa municipal, etc. En el caso de una universidad podrían brindarse servicios para consultar las mesas de exámenes, inscribirse a materias, consultar cursos de postgrado, etc.

Descripción del servicio.

Los servicios son la base de un portal, se debe estandarizar la información de los servicios a fin de brindar una visión estructurada del portal. Para cada servicio se definen: 1) Nombre, 2) Descripción, 3) Requisitos, 4) Forma de acceso, 5) Contacto y 6) Forma de pago. Por ejemplo, una municipalidad brinda datos para obtener una licencia de conducir, donde se definen:

Nombre del servicio: "Licencia de conducir"

Descripción: "Se realizan trámites para el otorgamiento de los distintos tipos de licencias (originales, renovaciones, duplicaciones, ampliaciones, cambios de datos y domicilio)"

Requisitos: Los vecinos que tengan que renovar su licencia de conducir deben sacar el turno para realizar el trámite, dentro de los 30 días anteriores al vencimiento de la misma. Al momento de su turno, se deberá presentar la documentación requerida sin excepción caso contrario se procederá a la anulación del mismo.

Forma de acceso: "Ingresando a la página www.municipalidad.gov.ar/licencia-de-conducir usted puede solicitar su turno para trámites relacionados con licencias de conducir"

Contacto: "El horario de atención al público es de lunes a viernes de 8 a 14 horas en el edificio A de la Municipalidad. Teléfono: 0811-111-1111 int. 1. Mail: licencias@municipalidad.gov.ar"

Forma de pago: "Usted deberá abonar un costo total de $500 para el trámite de la licencia de conducir"

3 CMS-Gov

CMS-Gov es un gestor de contenidos especialmente diseñado para organizaciones gubernamentales. Permite configurar la presencia institucional, los receptores y los servicios públicos [6]. Inicialmente el organismo configura su portal, definiendo todos los ítems relacionados con la imagen visual del mismo: incorpora su logo, los colores representativos, tipografía, etc. Luego define la estructura documental del contenido a presentar en el portal, enfocado en los conceptos de servicios públicos y receptores de los mismos.

3.1 Estructura de CMS-Gov

El modelo CMS-Gov define dos secciones, como puede verse en la Figura 1: 1) Presencia institucional y 2) Contenido.

- Presencia institucional: permite darle al portal la entidad representativa de la organización gubernamental.
- Contenido: contiene la definición de los receptores de los servicios, y de los servicios públicos que entrega la organización, cada servicio se vincula a uno o varios receptores. Los servicios cuentan con cinco secciones para establecer: Descripción, Forma de Acceso, Requisitos, Forma de pago y Contacto.

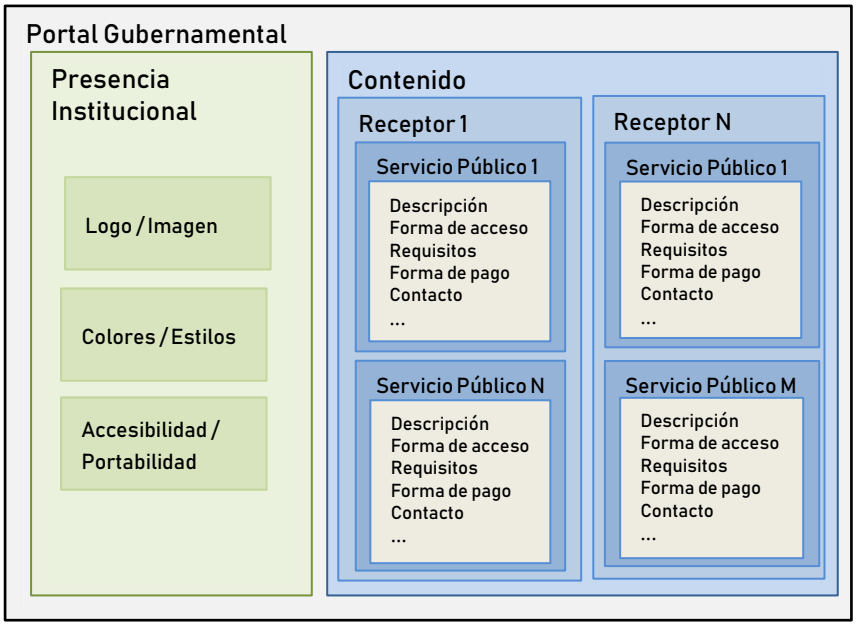


Fig. 1. Estructura de CMS-Gov

CMS-Gov está pensado para poder ser reutilizado por diferentes áreas dentro de una misma organización gubernamental, es decir que si un conjunto de servicios es brindado por un área particular de la organización, esta puede generar su propio portal de forma independiente y dicho portal ser vinculado desde el portal de la organización.

3.2 Configuración del Portal

Presencia Institucional.

Logo / Imagen: es importante para un organismo gubernamental que al momento de crear su portal se le brinde la posibilidad de agregar un logo o una imagen que permita identificarlo. Con CMS-Gov es posible seleccionar un logo o una imagen desde un archivo de su dispositivo local. Al momento de elegir un logo para el portal es

recomendable que el mismo no tenga fondo, para que pueda acomodarse sin inconvenientes al resto de la plantilla.

Colores / Estilos: el siguiente paso en la configuración del portal es la elección de los colores y los estilos de la página. CMS-Gov permite definir un conjunto de atributos relacionados con este punto. Una de las características que se puede definir es el color de fondo del portal. Es aconsejable que el mismo tenga cierta correspondencia con el logo y los colores de la organización. Como se puede ver en la Figura 2 CMS-Gov permite seleccionar los colores de acuerdo al modelo de color RGB (Red Green Blue, en español Rojo Verde Azul). También es posible definir la tipografía que va a tener el portal, incluyendo aspectos como tamaño, color y fuente de los títulos, los párrafos, los menúes, etc.

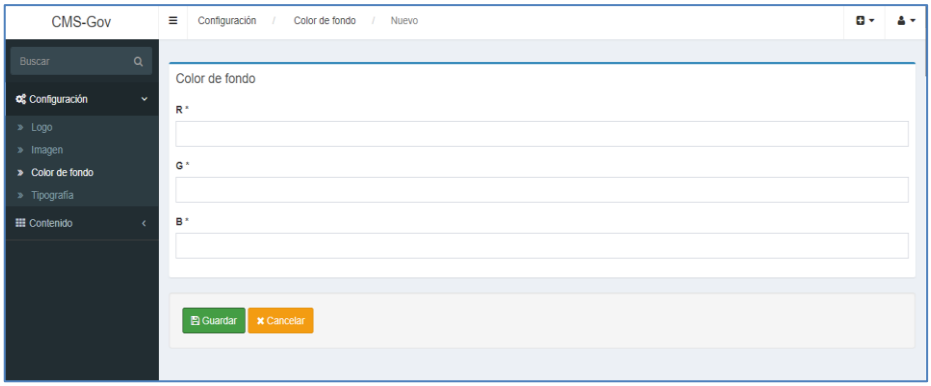


Fig. 2. Configuración de los colores del portal

Accesibilidad / Portabilidad: CMS-Gov brinda el aseguramiento de que el portal creado por la organización cumplirá con las características de portabilidad y accesibilidad definidas anteriormente.

Contenido.

Receptores: el portal está dirigido una comunidad que utiliza los servicios. Esta comunidad es variable en función de los fines del organismo. Por ejemplo, un municipio dirige servicios a ciudadanos que viven en ese municipio (Residentes), a empresas que tributan en el municipio, a otras secretarias del municipio, etc. En el caso de una Universidad los servicios son dirigidos a los alumnos, docentes, no docentes, etc. CMS-Gov permite definir los receptores y genera filtros automáticos para los servicios que se asocien a cada uno de los mismos.

Servicios públicos: el principal objetivo del portal es acercar servicios a los consumidores, estos servicios tienen que estar organizados de forma que sean fáciles de encontrar por los interesados. Una vez definidos los receptores, se procede a la carga de nuevos servicios que se mostrarán en el portal. Al ingresar un servicio, como se puede observar en la Figura 3, se debe indicar: el nombre del servicio, un nombre corto con el cual se presentará al mismo en la lista de servicios y quiénes serán los receptores de dicho servicio.

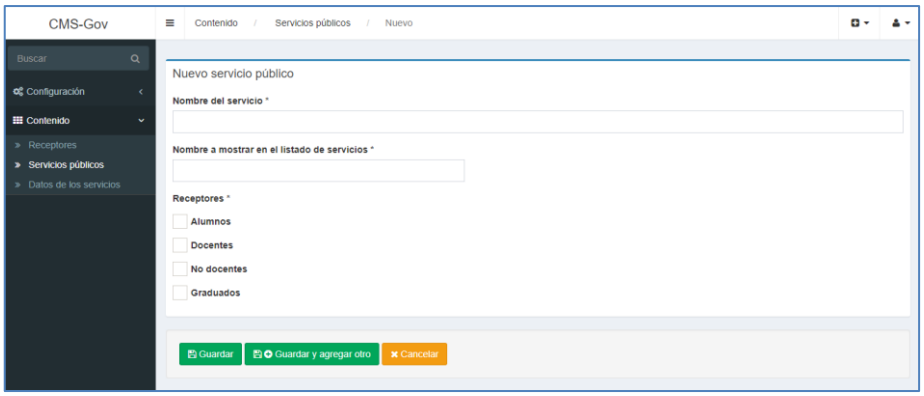


Fig. 3. Creación de un nuevo servicio

Datos de los servicios: CMS-Gov propone cinco apartados de datos correspondientes a los servicios, *Descripción*, *Forma de Acceso*, *Requisitos*, *Forma de Pago y Contacto*. Las organizaciones pueden agregar, modificar o quitar apartados en el momento de la configuración, de forma que sean acordes a los servicios que brindan. Los apartados definidos serán mantenidos para todos los servicios agregados, con el fin de mantener la consistencia, de acceso y visual, entre estos. Una vez configurada la estructura de los datos de los servicios es posible iniciar la carga de los mismos. Como se puede observar en la Figura 4, al seleccionar un servicio la organización puede decidir cuál o cuáles de sus apartados desea agregar y CMS-Gov le permitirá completar los datos de los mismos.

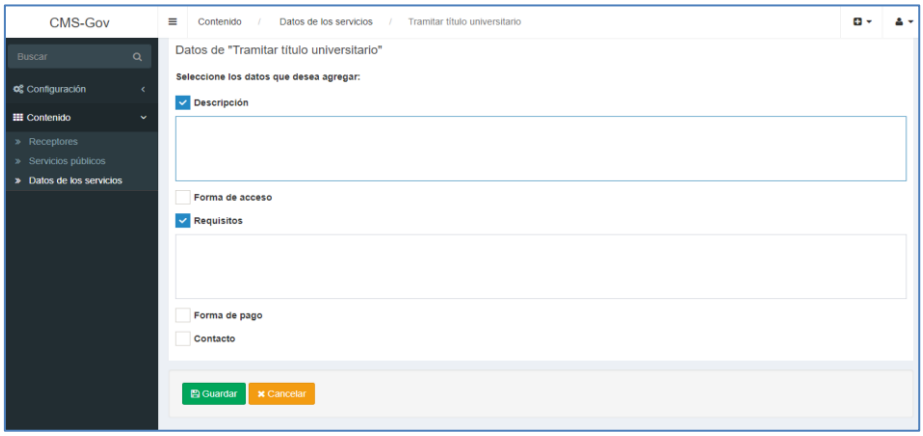


Fig. 4. Configuración de los datos de un servicio

4 Casos de estudio

Se realizaron tres casos de estudio con diferentes prototipos de organizaciones gubernamentales distintas y se aplicó CMS-Gov para generar, para cada una de ellas, un portal que se adecúe a las necesidades de las mismas.

Caso a) Prototipo de municipalidad que brinda servicios a *Ciudadanos*, *Empresas*, *Gobiernos* y *Empleados* y define, para cada uno de estos receptores, una cartera de servicios específica. En la Figura 5 se pueden observar los receptores anteriormente mencionados y el conjunto de servicios ofrecido para los *Empleados*.



Fig. 5. Prototipo Municipalidad

Caso b) Prototipo de institución educativa pública que otorga diversos servicios a un conjunto de receptores, tales como, *Alumnos*, *Docentes*, *No docentes* y *Graduados*. De cada uno de los servicios que definidos se desea mostrar los apartados de *Descripción*, *Forma de Acceso* y *Requisitos*. El resto de los apartados sugeridos por CMS-Gov no son de importancia para la organización. En la Figura 6 se muestra la pantalla del portal del caso b) accedida desde un teléfono móvil.



Fig. 6. Prototipo Universidad (móvil)

Caso c) Prototipo de secretaría perteneciente a la municipalidad del caso a). Se trata de una secretaría de salud que brinda sus servicios a *Ciudadanos*, *Empresas* y *Hospitales*. Si bien mantiene el logo y los colores de la municipalidad, la secretaría como área específica redefine un nuevo portal adaptándolo a sus necesidades. En la Figura 7 vemos que para esta secretaría se define una imagen representativa, los receptores que anteriormente veíamos en un menú superior ahora los podemos observar en el mismo menú lateral y que la única información que tendrán los servicios será la *Descripción* y la *Forma de Acceso*.



Fig. 7. Prototipo Secretaria de Salud

5 Conclusiones

Un gestor de contenidos para portales de organismos gubernamentales permite la administración y publicación de accesos a servicios públicos a los ciudadanos de forma rápida y sencilla.

Se presentó CMS-Gov, un gestor de contenidos que permite generar portales orientados a servicios, que diferentes organismos ponen a disposición de la comunidad. La herramienta permite realizar la configuración de la *presencia institucional* para que el portal mantenga consistencia con el sitio web de la organización y permite la administración de los accesos a los *servicios públicos* que brinda, manteniendo la información estandarizada para cada servicio, de una forma sencilla. No requiere tener conocimientos de programación de sitios web y, además, facilita el acceso a los servicios a los receptores definidos por la propia organización.

Se generaron tres prototipos, uno orientado a una municipalidad, uno orientado a una institución educativa pública, y otro a una secretaría que forma parte de una estructura interna de una municipalidad.

6 Agradecimientos

Este trabajo de investigación fue financiado por el Proyecto PIT-AP "Herramientas para el desarrollo y la entrega de servicios públicos digitales de acción social para municipios bonaerenses".

7 Referencias

1. T. Aitkenhead (2005). "Web Portals in Government Service" en A. Tatnall "Web Portals, the New Gateways to Internet Information and Services". IGI Global 2005.
2. M. Pérez-Montoro (2006). "Gestión del conocimiento, gestión documental y gestión de contenidos" en J. Tramullas "Tendencias en documentación digital". TREA, pp. 110-134.
3. D. E. Luna Reyes (2016). "Sitios Web y Portales de Gobierno", en J. R. Gil-García, J. I. Criado y J. C. Téllez "Tecnologías de Información y Comunicación en la Administración Pública: Conceptos, Enfoques, Aplicaciones y Resultados". INFOTEC, pp. 161-186.
4. L. F. Luna Reyes, J. R. Gil-García, R. Sandoval Almazán (2010). "Reflexiones sobre la evaluación de los portales de gobierno en Internet" en http://www.redalyc.org/articulo.oa?id=67613199005. Fecha de consulta: Julio de 2018.
5. J. C. Pasco, R. Roy, M. Borja (2013). "Portales web como herramienta de gobierno electrónico". ProGobernabilidad.
6. A. Fath-Allah, L. Cheikhi, R. E. Al-Qutaish y A. Idri (2014), "Towards a best practice based e-government portals maturity model". Proceedings of the International Conference on Computer Science and Information Systems, pp. 101-107.

Caso de estudio sobre GDPR aplicado en Sistemas de Gestión Académica

Lía Hebe Molinari[1], María Alejandra Sebastián[1], Nadia Estefanía Vázquez[1]
[1]Universidad Nacional de La Plata - Facultad de Informática
50 y 120 - La Plata, Argentina
lmolinari@info.unlp.edu.ar, asebastian@linti.unlp.edu.ar , nvazquez@linti.unlp.edu.ar

Resumen: El reglamento europeo para la protección de los datos (General Data Privacy Regulation, GDPR), fue publicado en Mayo de 2016 y comenzó a aplicarse a partir del 25 de Mayo de 2018. Como toda nueva normativa su puesta en vigor exige su cumplimiento, e inevitablemente interpela a las organizaciones en cuanto a los cambios formales, administrativos y culturales que conlleva; cuándo y cómo se debe aplicar, qué nuevos roles incorpora en la estructura organizacional, cómo y a quién se informa de dichos cambios son temas que se deben resolver. Este artículo aborda un caso de estudio en el ámbito académico nacional argentino poniendo en evidencia cómo una regulación europea tiene incidencia en repositorios locales.

Palabras Clave: Reglamento General de Protección de Datos. GDPR. Datos Privados. Derechos fundamentales. Protección de datos personales.

1 Introducción

Uno de los mayores retos con los que nos hemos enfrentado en los últimos años es la regulación de los datos que circulan en la red, es decir, el establecimiento de mecanismos para la protección integral de los datos personales de cualquier ciudadano que haya quedado almacenado en un archivo, registro, banco de datos u otro medio externo, ya sea público o privado [1].

En Europa, el 25 de Mayo de 2018 entró en vigencia el nuevo reglamento 2016/679 del Parlamento Europeo y del Consejo (en inglés, Global Data Privacy Regulation, conocido como GDPR) [2], que dispone que el flujo de datos personales hacia países terceros podrá hacerse libremente, siempre y cuando esos países cuenten con legislación que garantice, a juicio de la Comisión Europea, un nivel de protección adecuado. Tras la finalización de la reforma de la normativa de protección de datos, la Unión Europea (UE) asume que la protección y el intercambio de datos personales no son mutuamente excluyentes, y concluye que un sistema sólido de protección de datos facilita la circulación de los mismos en cualquier ámbito (educativo, económico, etc.) y/o país. Por esto, la Comisión al Parlamento Europeo y al Consejo, ha emitido el comunicado “Intercambio y protección de los datos personales en un mundo globalizado”, donde proporciona una serie de mecanismos para permitir las transferencias internacionales de datos. El objetivo es garantizar que, cuando se

transfieran datos personales de ciudadanos europeos a terceros países, se mantenga el mismo nivel de protección con respecto a los mismos.
Este reglamento exige a los países que están por fuera de la UE, analizar y establecer medidas sobre datos inherentes a ciudadanos europeos que puedan alojarse en sus repositorios y a reveer sus legislaciones en el contexto de la protección de los datos personales.
En paralelo a la aprobación del GDPR, en Argentina se han producido cambios normativos. A comienzos del 2000, la Dirección Nacional de Protección de Datos Personales elaboró un anteproyecto de reforma de la legislación, sobre la cual recientemente se produjo un importante avance cualitativo alineado con los estándares internacionales vigentes, al otorgar desde fines de 2017 el carácter de autoridad de aplicación en materia de protección de datos personales a un organismo autárquico y con autonomía funcional como es la Agencia de Acceso a la Información Pública (AAIP). La AAIP es el organismo nacional que tiene como función el garantizar el cumplimiento de la Ley de Acceso a la Información Pública, la protección de los datos personales y el funcionamiento del Registro Nacional "No Llame".

2 Reglamento General de Protección de Datos (GDPR)

GDPR ha reforzado los derechos relativos a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de éstos, añadiendo nuevos elementos que mejoran la capacidad de decisión y control de los ciudadanos sobre sus propios datos.
El reglamento busca proteger los datos y la forma en la que las organizaciones los procesan, almacenan y destruyen cuando ya no son requeridos. Establece normas más estrictas, como el autocontrol o control individual, o cómo las compañías pueden usar la información que está directa y personalmente relacionada con los ciudadanos. Estas normas rigen además lo que sucede si se viola el acceso a los datos personales de un individuo y las secuelas (penalidades) que las organizaciones pueden padecer en tal caso [3].
En el tiempo que regía la Directiva de Protección de datos EU, cuando una entidad sufría una filtración no estaba obligada a informar de ella. Sin embargo debido a la cantidad de situaciones que han puesto en evidencia los escasos controles orientados a proteger nuestra información más privada, las entidades se vieron en el compromiso de dar las explicaciones necesarias, especialmente acerca de la forma en que la filtración afectó directamente a sus clientes.
El GDPR va a tener un impacto significativo para las organizaciones y su forma de administrar los datos, con sanciones para aquellas entidades que provoquen una violación que pueden llegar a una multa que involucre hasta un 4% de los ingresos globales.

2.1 Alcance del GDPR

Por medio de esta regulación, la UE quiere dar más control sobre los datos personales y la utilización de las mismas, y a las entidades un entorno jurídico más simple y más claro para operar, haciendo que la ley de protección de datos sea idéntica en todo el mercado.
La aplicación de la reglamentación se extiende a [4]:

- Organizaciones con presencia física en al menos algún país miembro de la Unión Europea.
- Organizaciones que procesan o almacenan datos sobre individuos que residen en la Unión Europea.
- Organizaciones que utilizan servicios de terceros que procesan o almacenan información sobre individuos que residen en la Unión Europea.

2.2 ¿Qué se entiende por datos personales?

La era digital que transitamos plantea diariamente desafíos complejos relacionados con la recopilación y uso de la información personal en áreas muy distintas pero que a su vez se entrecruzan, como la economía, las políticas, las telecomunicaciones, la salud. Desde la perspectiva legal, los datos y la información personal online siempre se han enfrentado al desafío de garantizar a los consumidores-usuarios la misma protección y seguridad jurídica que en un mercado físico. El significado del concepto de dato personal no es uniforme. A continuación se detallan definiciones de "datos personales" de distintas legislaciones [5]:

- La Ley 25.326 de Protección de los datos personales de Argentina, en el Art.2 define a los datos personales como *"Información de cualquier tipo referida a personas físicas o de existencia ideal determinadas o determinables"*.

- El GDPR en el Art.4 del Reglamento expresa que: «***"datos personales"**: toda información sobre una persona física identificada o identificable; se considerará identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un número de identificación o uno o varios elementos específicos, característicos de su identidad física, fisiológica, psíquica, económica, cultural o social*».

Es decir, si la persona es identificable, toda la información que se tenga sobre ella serán datos personales**.** En la actualidad, una persona está directamente identificada por su "nombre y apellido". Aunque en ocasiones el nombre y apellido no son suficientes para identificar a una persona y deberán complementarse con otros datos, como una dirección, una foto de perfil o un número de cobertura social. En esos casos en que los identificadores que disponemos no permiten a nadie individualizar a una persona determinada, pero al combinarlos con otros datos (tanto si tenemos conocimiento de ellos como si no) es posible distinguir a la persona de otras, entenderemos que la persona es *indirectamente identificable*. También puede

considerarse a una persona identificada cuando se disponga de otros identificadores que nos permitirán individualizar a una persona, crear un perfil y atribuirle decisiones. Las definiciones presentadas se limitan a la protección de las personas físicas vivas, excluyendo a las personas fallecidas. El reglamento expresa en el Art 27: *"El presente Reglamento no se aplica a la protección de datos personales de personas fallecidas. Los Estados miembros son competentes para establecer normas relativas al tratamiento de los datos personales de éstas"*.
De este análisis surge una clasificación de datos que se debe tener en cuenta para la protección de la misma:

1. Básicos: toda información relativa a una persona.
2. Especiales: los datos especiales son los que poseen contenido de origen étnico o racial, datos genéticos o biométricos; convicciones religiosas o filosóficas, opiniones políticas, afiliación sindical; datos de Salud, vida sexual u origen sexual. (art.9 GDPR)
3. Penales: datos relativos a condenas y delitos penales o medidas de seguridad afines.

2.3 Derechos

La regulación otorga a los involucrados el control sobre qué información pueden utilizar que esté directamente vinculada con ellos y les proporciona ocho derechos específicos. Estos derechos adquieren el nombre de "*Derechos de Sujetos de Datos*". Esto significa que No son entidades pasivas que no tienen más opción que aceptar lo que sea que suceda con sus datos personales, sino que son propietarios independientes de sus datos y determinan cómo quieren que sean usados.
Los derechos que proporciona la misma son: derechos de acceso (art.12 y art.15 GDPR); de rectificación (art.12 y art.16 GDPR); derecho de oposición (art.12 y art.21 GDPR); derecho de supresión ("al olvido") (art.12 y art.17 GDPR); derecho a la limitación del tratamiento (art.12 y art.18 GDPR); derecho a la portabilidad (art.12 y art.20 GDPR); derecho a no ser objeto de decisiones individuales automatizadas (art.12 y art.22 GDPR) y derecho de información: (art.12, art.13 y art.14 GDPR).

2.4. Tratamiento y Consentimiento

Uno de los cambios más importantes que impone la regulación fuertemente es el consentimiento del usuario sobre sus datos, definiendo en el artículo al término "consentimiento del interesado" como "*toda manifestación de voluntad libre, específica, informada e inequívoca por la que el interesado acepta, ya sea mediante una declaración o una clara acción afirmativa, el tratamiento de datos personales que le conciernen.*" (art.4 GDPR).

2.4.1 Aplicación de los Principios del tratamiento

El reglamento dedica el capítulo 2 a los "Principios" y especifica que deben ser aplicados a toda la información relativa a una persona y los datos que lo identifiquen.
El tratamiento debe ser lícito, es decir legítimo, para ello debe definirse un contrato con el interesado, o celebrar un consentimiento explícito para fines específicos. Si los datos se obtienen de un acceso público, entonces su procedencia debe ser de una fuente pública o expresamente el interesado los ha hecho públicos.

El sujeto obligado a verificar el cumplimiento de todos estos requisitos será el Responsable del Tratamiento que además deberá ser capaz de demostrar de forma activa que cumple con ellos: «responsabilidad proactiva».

2. 4.2 Consentimiento y Consentimiento Explícito

El consentimiento es toda manifestación de voluntad libre, específica, informada e inequívoca por la que el interesado acepta el tratamiento de los datos.
A partir de la vigencia del GDPR, la forma que más se utilizaba era utilizando una clara acción afirmativa como, por ejemplo, marcar una casilla de una web, el movimiento físico del teléfono, grabaciones de voz. Ahora el tratamiento sólo será lícito si un individuo dio el consentimiento para uno o varios fines específicos (art.6.1 GDPR). Y el consentimiento deberá ser explícito para el tratamiento de categorías especiales de datos (art.9.2ª GDPR).
La diferencia del consentimiento explícito con el general radica en que el primero no debe dejar lugar a la libre interpretación, siendo confeccionado de manera precisa y clara, y los responsables de los datos tendrán que asegurarse de la obtención de dicho consentimiento de manera indiscutible. Tal y como ha especificado la oficina británica (ICO), "la declaración para obtener el consentimiento explícito debe especificar la naturaleza de los datos a recopilar, los detalles de la decisión automatizada y sus efectos o los detalles de los datos que se van a transferir y los riesgos de tal transferencia".

2.5 El enfoque de riesgo

La transformación tecnológica que se ha estado desarrollando, y sigue aún, en el ámbito del tratamiento de la información personal producto de la entrada en vigencia del GDPR, ha producido que las entidades públicas o privadas deban estar preparadas para adoptar las medidas técnicas y organizativas adecuadas al uso de los datos "no anonimizados" no solo para aquellas pertenecientes a la UE sino que a todas que salvaguarden información de ciudadanos europeos. Como piezas claves para que convivir con la nueva legislación, las entidades deberían analizar los riesgos que conllevan la utilización de datos personales de sus usuarios, y por consiguiente satisfacer la necesidad de creación de una figura profesional responsable del buen uso de esos datos [6].
Para las entidades que traten datos con un nivel de riesgo elevado, deberán establecer las medidas que sean necesarias para garantizar los derechos y libertades de las personas [7].

3 Caso de estudio: Sistemas de gestión académica y cumplimiento GDPR

EL GDPR se aplica al Espacio Económico Europeo (EEE), que incluye todos los países de la Unión Europea, más Islandia, Liechtenstein y Noruega. Cuando los datos personales se transfieren fuera del EEE, la protección que ofrece el GDPR debe viajar con los datos. Esto significa que, para exportar datos al extranjero, las entidades deben garantizar que existen determinadas garantías.

El GDPR ofrece un conjunto variado de mecanismos para transferir datos a terceros países. Dichas transferencias están permitidas en los casos siguientes:

1. la UE considera adecuadas las protecciones del país, o
2. la entidad que trata los datos, por ejemplo, toma las medidas necesarias para ofrecer garantías adecuadas, como incluir cláusulas específicas en el contrato celebrado con el importador de fuera de la UE de los datos personales, o
3. la entidad, por ejemplo, se basa en motivos específicos («excepciones»), para la transferencia como el consentimiento del interesado.

Sin embargo, para poder realizar transferencias internacionales sin requerir ninguna autorización específica, es cuando la Comisión de la UE decida el nivel de adecuación en relación con el tercer país o un territorio o un sector específico del mismo, o con una organización internacional.

La decisión de adecuación es una resolución adoptada por la Comisión que garantiza que la transferencia internacional de datos posee un nivel de protección suficiente. La Comisión podrá derogar, modificar o suspender cualquier decisión de adecuación sin efecto retroactivo. En la actualidad, Argentina se encuentra dentro de los países que tienen una decisión de adecuación.

3.1 Planteo del Problema

Al contemplar esta situación y como el GDPR impactaría en Argentina, nuestro foco se dirigió a lo que sucede dentro del ámbito educativo y las herramientas de soporte que se utilizan para la gestión de los datos.

La cuestión surge a partir de que la Constitución Nacional Argentina establece que cualquier extranjero que consiga la residencia puede estudiar en las universidades públicas, con los mismos derechos que un ciudadano nacido en el país.

En los últimos años, el ingreso y egreso de alumnos extranjeros en las universidades argentinas incluyendo alumnos que son ciudadanos de países de la UE. En consecuencia, como se describió en la sección 2 del artículo, se debe realizar un análisis de los sistemas de gestión ya que almacena información de ciudadanos de la UE, y verificar que se cumplen con todas los puntos detallados en el GDPR, siempre y cuando no se contradiga con la normativa Argentina, el Habeas Data[5].

3.2 Análisis del problema

3.2.1 Funcionalidad del Sistema de Gestión Académica

Un Sistema de Gestión Académica (SGA) es una herramienta integrada que permite registrar y administrar todas las actividades académicas de la universidad y sus dependencias académicas, desde que un alumno se inscribe hasta que egresa. El objetivo del SGA es la administración de las tareas académicas en forma óptima y segura, con la finalidad de obtener información consistente para los niveles operativos y directivos.

El SGA no es un sistema en sí mismo, sino que es un conjunto de diferentes aplicaciones que interactúan entre sí, y que comparten la información que almacena cada una.

Las principales prestaciones que cuenta el sistema son:

- Almacena la totalidad de la oferta educativa de la institución: carreras, certificaciones, competencias, títulos, etc. / Flexibilidad de planes de estudio.
- Presenta una organización por módulos relacionados a la gestión: matrícula, asistencia, cursadas, egresos, etc.
- Almacena los registros de directa relación con la vida del alumno dentro de la institución, desde la postulación del aspirante hasta el circuito de egreso del mismo.
- Almacena los Registros del docente, desde sus antecedentes hasta asistencia en las distintas instancias.
- Gestión de múltiples perfiles de datos (información) y funcionales (sobre operaciones).
- Personalización de reportes, operaciones y módulos.

Como se puede notar, las funcionalidades del SGA requieren que se registre mucha información sensible y privada de las personas que intervienen en los diferentes procesos administrativos.
En el análisis de los tipos de datos que se almacenan de los diferentes actores que utilizan el sistema y dado un perfil básico, habitualmente un SGA registra datos del tipo:

- Datos personales
- Datos familiares
- Datos acerca de cobertura social
- Datos secundarios
- Fuente de financiamiento de estudios
- Situación laboral
- Deporte
- Idioma y nivel
- Discapacidad

Teniendo en cuenta toda esta información y la exigencia de la regulación sobre la obligación de informar a las personas interesadas sobre el tratamiento de sus datos, un SGA debería, en una primera instancia, verificar si con respecto al tipo de información que almacena, cumple con los ítems detallados:

1. Información Básica:
    a. Se notifica el responsable de tratamiento, es decir a la persona encargada de la gestión de los datos y de coordinar la administración del almacenamiento de los mismos?
    b. Se realiza una descripción sencilla de los fines del tratamiento?
    c. Se notifica los plazos y el criterio de conservación de los datos?
    d. Se hace referencia sobre los derechos que tiene los involucrados?

2. Información detallada:
    a. Se notifica la posibilidad de destinatarios de los datos personales?
    b. Existe un interés legítimo en el responsable?
    c. Se notifica políticas para decisiones automatizadas y/o perfiles?

d. Se notifica el derecho al retirar el consentimiento prestado?

En ambos aspectos esta información debe proporcionarse en el momento en el que se soliciten los datos.
Además un SGA debería informar si:

1. se recogen los datos personales con fines determinados,
2. los datos personales se mantienen exactos,
3. se informa del derecho a solicitar la rectificación o supresión de sus datos.

En todos estos casos, el sistema estará obteniendo datos personales de los usuarios, por lo que, en primera medida es necesario que inserte un aviso legal para dar cumplimiento al deber de información y consentimiento. Es necesario que los usuarios **acepten expresamente** dicho aviso.

3.2.2 Sistemas de gestión académica universitaria y el Consentimiento.

Dado que el GDPR pone en énfasis el consentimiento de los involucrados, y que los SGA registran mucha información sensible, debería realizarlos siguientes preguntas:

1. ¿El SGA gestiona el concepto de consentimiento de forma clara e independiente compatible con el GDPR?
2. ¿Hay algún registro o manera en la que se puede demostrar que el usuario (alumno, docente, etc...) dio su consentimiento para el tratamiento?
3. ¿El SGA solicita el consentimiento de forma inteligible y de fácil acceso para aquel que va a registrarse en el sistema?
4. ¿Se utiliza un lenguaje claro y sencillo en el consentimiento?
5. ¿Se le informa previamente en el consentimiento al usuario la información (o datos personales) que será almacenada recabar?
6. ¿El SGA implementa algún mecanismo que permita prescindir del consentimiento con la misma facilidad que se acepta?
7. ¿El SGA ofrece los medios para retirar el consentimiento en cualquier momento?
8. ¿Se recaba el libre consentimiento?

3.2.3 Recomendaciones para cumplimiento del GDPR en SGA

Más allá de las medidas de seguridad que el SGA implemente, es importante realizar una evaluación acerca del cumplimiento de la normativa, de manera que se asegure de adoptar una serie de medidas de seguridad, de carácter técnico y organizativo para proteger todos los datos.
Con el lineamiento en la regulación un SGA debería:

- Antes de implementar nuevas medidas de seguridad o reforzarlas, aquellas personas que tengan el rol para tratar los datos personales deben llevar a

cabo una evaluación de los riesgos para garantizar un nivel de seguridad adecuado.

- Si se produce una quiebra de seguridad que afecte los datos personales de los usuarios, poniendo en riesgo sus derechos y libertades, con carácter general, deberán contar con un mecanismo para comunicar tal quiebre se seguridad y los efectos que produjo o que pueden llegar a producir.
- Si el SGA va a realizar tratamientos que supongan de alto riesgo para los derechos y libertades de los usuarios, antes de empezar a ejecutarlo, deberá realizar una evaluación de impacto relativa a la protección de datos.
- El SGA deberá tener en cuenta la 'Privacidad por diseño' y 'Privacidad por defecto'.
- Deberá incluir medidas para asegurar la confidencialidad, integridad, disponibilidad y resiliencia permanentes de los sistemas y servicios de tratamiento de los datos
- Deberá contar con medidas de seguridad para evitar pérdida, destrucción o daño accidental
- El SGA en todo proceso de análisis deberá tener en cuenta los riesgos que presenta el tratamiento como consecuencia de su destrucción, pérdida o alteración accidental o ilícita que son transmitidos, conservados o tratados, o la comunicación o acceso no autorizados a dichos datos para evaluar el nivel de seguridad aplicado.
- Deberá tener un procedimiento para que los encargados del tratamiento notifiquen las brechas al responsable en el momento en que tengan conocimiento de ellas

4 Conclusiones y trabajos futuros

GDPR regula la forma en que las entidades tratan y gestionan los datos personales; y al mismo tiempo provee mecanismos para la transferencia de datos transfronterizos. En materia de protección de datos, de acuerdo a esta legislación europea, todo lugar que guarde información sensible sobre ciudadanos europeos debe ajustarse a la misma. GDPR representa una nueva oportunidad para mejorar la gestión de los datos personales y, posteriormente, aumente la confianza de todas las personas que intervienen.

Como trabajo futuro se propone analizar GDPR en otro ámbitos, teniendo un principal cuidado en ámbitos donde se registren datos de menores considerando que la regulación plantea un mayor cuidado de los mismos.

En el ámbito de la gestión pública son varias las dependencias que están analizando el cumplimiento de GDPR. En este caso hay mucho por hacer definiendo una metodología que incluya el proceso de análisis de riesgo, tratamiento de los datos, consentimiento, análisis del puesto de PDO, análisis del tratamiento de los datos y mejoras de diseño en el sistema.

Referencias

[1] Fuga de datos a nivel mundial: Riesgos y errores comunes de los empleados. https://www.cisco.com/web/offer/em/pdfs_innovators/LATAM/data_mist_sp.pdf

[2] Dictamen 4/2007 sobre el concepto de datos personales http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_es.pdf

[3] Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo. http://www.boe.es/buscar/doc.php?id=DOUE-L-2016-80807

[4] Comunicación de la comisión al parlamento europeo y al consejo.https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52017DC0007&from=ES

[5] Ley de Protección de los Datos Personales. http://www.jus.gob.ar/media/3210629/anteproyecto_de_ley_de_protecci_n_de_los_datos_personales.pdf

[6] Guía práctica de Análisis de riesgos en los tratamientos de datos personales sujetos al RGPD https://www.aepd.es/media/guias/guia-analisis-de-riesgos-rgpd.pdf

[7] Guía práctica para las Evaluaciones de Impacto en la Protección de los datos sujetas al GDPR https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf

[8] Principios Nacionales e Internacionales en el marco de la Protección de Datos Personales. http://44jaiio.sadio.org.ar/sites/default/files/sid227-249.pdf

[9] PROTECCIÓN DE DATOS Luis Felipe López Álvarez. http://www.isaca.org/chapters7/Madrid/Events/Eventos/Documents/20180320%20AGRC18/20180320%20ISACAMadrid%20AGRC18%20Luis%20Felipe%20Lopez.pdf

Monitoreo de unidades de transporte público de Ciudad del Este

Marcelo Velázquez, Katia Andrea Ayala Díaz, Jorge Luis Arrúa Ginés

Universidad Nacional del Este, Facultad Politecnica. Ciudad del Este, Paraguay
marcelovelazquez@fpune.edu.py
http://www.fpune.edu.py

Resumen El objetivo de este trabajo fue el desarrollo de un sistema basado en tecnologías libres, capaz de monitorear el recorrido de unidades de transporte haciendo uso de una aplicación para smartphones Android. El sistema propone una arquitectura cliente-servidor empotrado que consiste en un nodo recolector de datos, ubicado en el interior de la unidad a ser monitoreada y un servidor, el cual alojará las bases de datos y lógica de funcionamiento, con acceso a red. El enfoque del proyecto se centra en ofrecer un sistema cuya implantación presenta una alternativa viable, desde el punto de vista económico, por la selección de materiales utilizados, y desde el punto de vista social, al proporcionar un sistema accesible, adecuado a las necesidades del plantel administrativo beneficiado. A través de métodos de verificación como pruebas de funcionamiento y el estudio de percepción a través de la implementación de encuestas, se pudo observar que el proyecto presenta resultados satisfactorios en los entornos estudiados.

Keywords: public transportation, geolocation, accessibility

1. Introducción

El crecimiento acelerado de Ciudad del Este como centro de comercio, generó numerosos cambios no previstos dentro de la estructura y las organizaciones que operan dentro de la ciudad. Uno de los sectores más afectados, sin duda, fue el del transporte público. Su uso ha aumentado exponencialmente en los últimos años, por lo que la gestión de los recursos que implica el transporte de pasajeros se ha tornado caótico. La utilización de un sistema de control manual, fácilmente adulterable por su dependencia en el factor humano, que lleva operativo desde la década de los años 60 contribuye al desorden generalizado que caracteriza a la prestación de este tipo de servicios.

Teniendo esto en cuenta, se plantea la necesidad de proveer un sistema tecnológico, para una mejor gestión de flujo de unidades de empresa de transporte público desde el sector empresarial. Se plantea el desarrollo y diseño de un prototipo de sistema para un servicio de control de frecuencia de ómnibus de pasajeros del sistema de transporte de la ciudad, con el objetivo de hacerlo como alternativa de costo más económico que las propuestas comercialmente disponibles y

orientado hacia comportamientos regularmente asociados con los habitantes de esta zona.

La finalidad en la implantación de este sistema es de contar con las coordenadas exactas de las unidades con el menor retardo posible, a fin de facilitar la toma de decisiones para la liberación de nuevas unidades de transporte.

Este trabajo se propone como continuación de la tesis denominada Rastreo de Transporte Público interurbano de Ciudad del Este aplicando plataformas de código abierto"[1], orientada al usuario del transporte público. Su objetivo era el desarrollo de un sistema basado en tecnologías libres, capaz de monitorear el recorrido de unidades de transporte haciendo uso de una aplicación para smartphones Android.

2. Objetivos

2.1. Objetivo general

- Desarrollar un sistema de monitoreo de frecuencia de unidades de transporte público de Ciudad del Este.

2.2. Objetivos específicos

1. Diseñar componentes del nodo de recolección y transmisión de datos haciendo uso de una plataforma smartphone.
2. Desarrollar un algoritmo para recepción de datos de sensores.
3. Diseñar arquitectura cliente-servidor para transmitir y procesar datos en línea.
4. Desarrollar interfaz gráfica que posibilite interactividad con la capa de presentación del sistema.
5. Someter a pruebas que garanticen la usabilidad del sistema.

3. Antecedentes

A nivel internacional, podemos citar trabajos que guardan relación al tema, como:

1. "Designing a smart transport system application for South Indian traffic scenarios. A modern approach towards digitalizing the transport systems"propone un sistema de localización de vehículos del transporte público de ciudades del sur de India, con el fin de desarrollar una aplicación para smartphone capaz de rastrear la proximidad de los mismos, sus horarios y la implementación de un sistema de conteo interno. El enfoque del proyecto se centra en el incentivo del uso del transporte público en detrimento del transporte propio, como forma de paliar la crisis energética de combustibles fósiles [2].

2. "Analysis of an automated bus tracking system for metropolitan using IoT" propone un sistema de monitoreo capaz de procesar datos como la velocidad y localización del bus, y mediante la integración de un módulo de lectura de tarjetas infrarrojas, facilitar el uso del sistema de personas con dificultades para la lectura, por medio de un sistema de voz con los detalles de la clave contenida en la tarjeta [3].
3. "Real-time bus location and arrival information system" propone un sistema de localización basado en gps, pero con soporte colaborativo de los pasajeros del bus, de manera a mejorar la exactitud de los datos en rutas de areas rurales y de problemas inpredicibles del tráfico, haciendo esta información pública a los demás potenciales pasajeros [4].
4. "A smart cost effective public transportation system: An ingenious location tracking of public transit vehicles" propone un modelo de costo eficiente comparado con alternativas basadas en gps, haciendo uso de comunicaciones por radiofrecuencia, donde se permite localizar en tiempo real la posición de un vehículo y la muestra de los resultados en una interfaz web diseñada para el efecto [5].
5. "Real-time on-campus public transportation monitoring system" propone un sistema de monitoreo de unidades del trasnsporte publico basado en google maps y tecnologias GPS, obteniendo como resultado una interfaz gráfica en formato web capaz de hacer el monitoreo de unidades de forma sencilla [6].
6. "Structural model of passenger counting and public transport tracking system of smart city" propone un estudio de simulación de sistemas discretos para permitir la invesigación de la estructura de conteo de pasajeros y monitoreo de trayectoria [7].
7. "Performance of tracking public transport in heterogeneous networks" propone el estudio del comportamiento de tranmisión de datos haciendo uso de distintas formas de conectividad dentro de una red ad hoc vehicular, demostrando cual alternativa ofrece el mejor rendimiento en el esquema de conectividad [8].

4. Materiales y métodos

4.1. Metodología

El presente proyecto reúne las condiciones metodológicas de una investigación tecnológica con enfoque cualitativo. La metodología de desarrollo utilizada fue la de programación ágil, priorizando iteraciones con valor agregado de funcionalidad en corto tiempo en vez de grandes iteraciones con periodos mas largos de desarrollo.

4.2. Muestra

La población objetivo del proyecto fue definida de forma a responder a las necesidades de un grupo de personas que ejercen la fiscalización en las empresas

de transporte. Tomando en cuenta que, a mes de mayo del año 2018, en Ciudad del Este se hallan en funcionamiento 10 empresas de transporte público, se consideran 30 personas que desempeñan la función de fiscalizador.

4.3. Diseño

La arquitectura del sistema plantea la siguiente estructura:

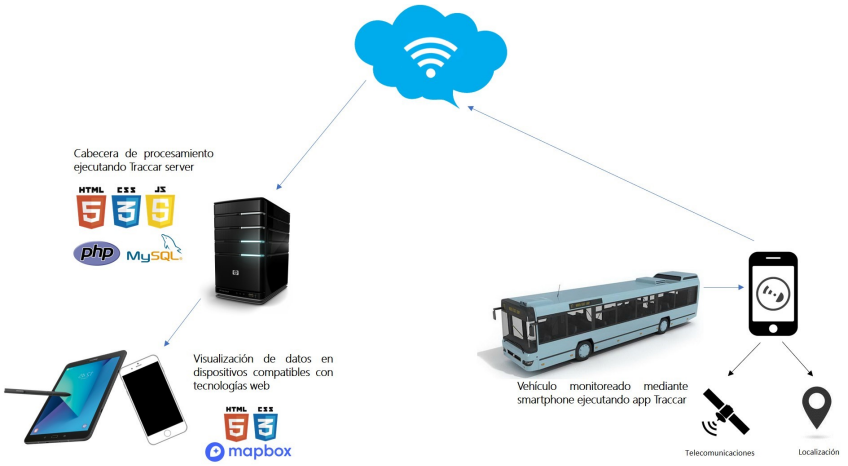


Figura 1. Diagrama funcional

Los dispositivos colectores instalados en la unidad a ser rastreada están compuestos por smartphones que ejecutan la plataforma Android como sistema operativo. Este requerimiento se basa en el uso de la aplicación cliente Traccar, que, utilizando los sensores propios del teléfono, es capaz de proporcionar información detallada de la geolocalización de este.

Fueron realizadas numerosas pruebas de funcionalidad, a fin de observar el rendimiento de la aplicación en teléfonos de distintas gamas, siendo finalmente seleccionados dispositivos que contaran al menos con 1GB de memoria RAM y la versión 4.4 Kit Kat del sistema operativo Android, a fin de tener una experiencia de uso satisfactoria.

El servidor de recepción de datos se compone de un ordenador ejecutando la plataforma Ubuntu Linux con una instancia del programa Traccar. El envío de paquetes a este servidor se realiza a través de la configuración de un parámetro dentro de la interfaz de usuario del programa instalado en el smartphone, donde se especifica la dirección IP y el puerto abierto para este propósito. Ante la dificultad propuesta por los trámites para la obtención de una IP pública, se procede al uso de un generador de dominios dinámicos basados en tecnología No-IP.

El hosting de la base de datos tipo MySQL y los datos necesarios para el funcionamiento de la interfaz web del sistema fueron alojados en el mismo entorno servidor descripto anteriormente.

El rastreo se realiza haciendo refrescos constantes de la unidad rastreada entre dos estaciones cabecera determinadas anteriormente, y en la interfaz del sistema es posible observar detalles como:

- La ubicación en el mapa, utilizando la API de Leaflet (Figura 2).
- Información como el horario de salida de la unidad, el potencial horario de llegada y el itinerario que cubre la unidad, además de su número identificador y una imagen de ésta.
- Se define también una página a la que se puede acceder con el botón Mas información, donde se tienen otros datos como el nombre del chofer al mando de esa unidad

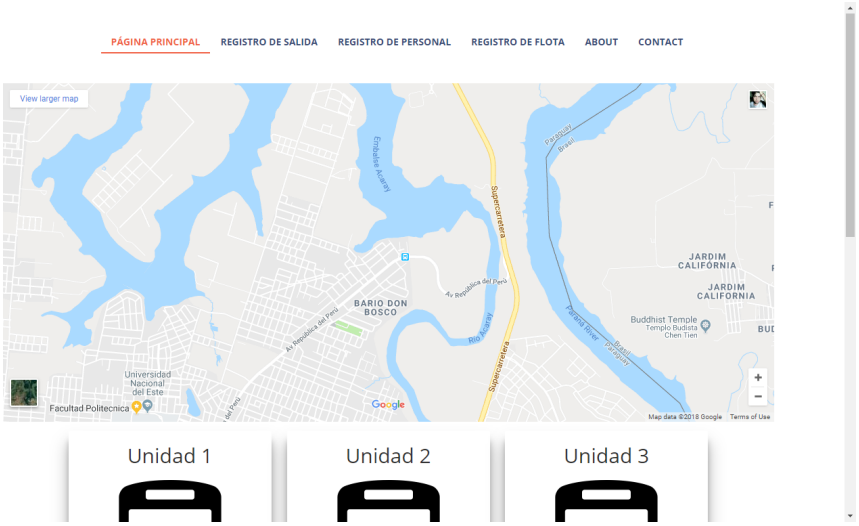


Figura 2. Pantalla Principal

5. Experimentos y resultados

5.1. Diseño de pruebas

Las pruebas de funcionamiento del sistema se basaron en el enfoque social del sistema y la integridad de los datos utilizados en mismo. Se desarrollaron de la siguiente forma:

- Se realizó un lista de verificación de distintos escenarios de inserción de datos. Se modelaron casos de uso de las funciones mas resaltantes, determinando el flujo correcto de ejecución y dos flujos alternativos que representaban interacciones con resultados no satisfactorios. Se realizaban pruebas en los tres escenarios, determinando el índice de ocurrencia de estos. A través de los hallazgos realizados en las pruebas, se puede verificar en la evaluación la robustez del sistema en cuanto a este tipo de situaciones, lo que garantiza la integridad y confiabilidad de los datos almacenados en base de datos.
- El estudio de usabilidad según perfil de usuarios potenciales del sistema se realizó definiendo el perfil del cliente potencial de este sistema, de manera a tener de forma delimitada las características de las personas que harán uso del sistema. Se tuvieron en cuenta dos parámetros, el nivel de involucramiento tecnológico y el tipo de lenguaje de estos. Esto sirvió como base para modelar la interfaz gráfica. Luego de la realización de pruebas con personas que cumplían con características similares al perfil definido, sus percepciones de uso del sistema fueron recolectadas en un formulario en el formato de encuesta de escala de Likert, lo que permitió, a través de un formulario de veinte preguntas, analizando la dispersión en el espectro positivo o negativo con cinco grados, conocer la respuesta del usuario potencial al sistema propuesto.

5.2. Resultados y discusión

Las pruebas realizadas al sistema presentaron los siguientes resultados:

1. El sistema es lo suficientemente robusto para evitar la inserción de datos que podrían comprometer la integridad de los datos almacenados en el mismo. La reacciones tomadas en las primeras revisiones de este, ayudaron a fortalecer las versiones preliminares del sistema. Se tomaron medidas como la implementación del parámetro *pattern* sobre el atributo *input* del formulario de rellenado de información, de tal forma a limitar la combinación de caracteres insertables en el campo de acuerdo al diccionario de datos establecido. Además, se implementaron soluciones basadas en texto a procesos de redireccionamiento entre módulos del sistema, para facilitar la navegación de los usuarios, evitando ambigüedades.
2. Para evaluar la fiabilidad del test de actitudes y de autonomía mediante la construcción de una escala ordinal de Likert, se solicitó a las personas compatibles con el perfil la evaluación de ítem con valores en una escala de 5

posibles respuestas. Para ello se construyeron preguntas utilizando la técnica par-impar aleatorizada, en dirección negativa a positiva en todos los ítems. Además, para la validación de los test, se realizó una prueba piloto a una muestra aleatoria simple de 75 personas, quienes sirvieron como evaluadoras. A ellas se les aplicó el test de conocimientos y actitudes, cuyos resultados permitieron un análisis descriptivo del puntaje obtenido para observar tendencias o variaciones de las preguntas.
El 86 % de la población mostró una respuesta favorable al sistema, el 13 % demostró una actitud neutra y el 1 % demostró actitud negativa hacia el mismo (Figura 3).

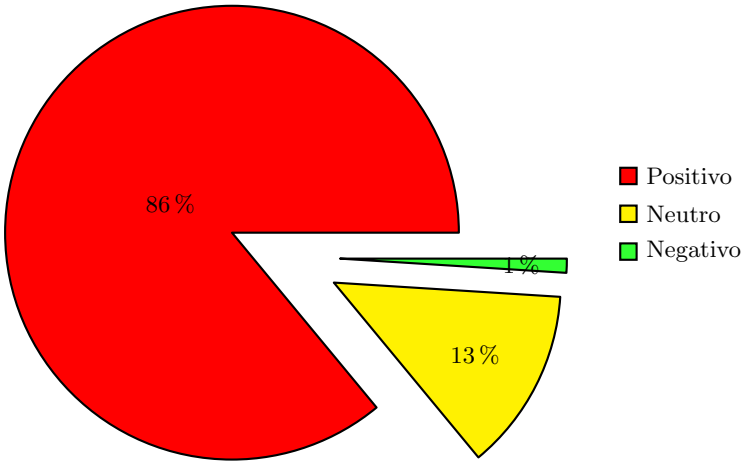


Figura 3. Promedio de respuestas a encuestados

En promedio la calificación final del sistema fue de 4.33, lo que puede considerarse como favorable dentro de una escala numérica de Likert de 1 al 5 (Figura 4).

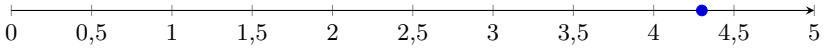


Figura 4. Calificación del sistema según escala de percepción de Likert

3. La utilización del software cliente Traccar presenta un funcionamiento estable y garantiza el correcto funcionamiento de colección de datos en equipos

que cumplan con las siguientes características mínimas: Sistema operativo Android en su versión 4 en adelante, Procesador de al menos 2 núcleos a una velocidad de 1GHz y 1GB de memoria RAM.

4. El uso de las API de rellenado de formularios nativo de HTML5 garantiza, a través de la implementación de reglas de completado, la integridad de los datos que sean insertados en la base de datos por parte del usuario final del sistema.
5. La utilización de una plataforma web basada en desarrollo de interfaz responsiva posibilita el uso del sistema en cualquier dispositivo con conexión a internet, sea este un ordenador o un smartphone.

6. Conclusiones

A través de las pruebas realizadas, pudo observarse el correcto funcionamiento del sistema como alternativa para el monitoreo de unidades de transporte público del área de Ciudad del Este. Pueden nombrarse entre los factores más importantes para su diferenciación con respecto a alternativas presentes en el mercado, la utilización de materiales de bajo costo que permiten la utilización de equipos con los cuales se dispone, sin hacer una inversión elevada,la accesibilidad del sistema, manifestada a través del énfasis del proyecto en la facilidad en el uso de las tecnologías de la información y la intergración de comportamientos propios de la región en que opera.

Esto implica que se ha logrado lo siguiente:

- Se desarrollaron diseños de componentes del nodo de recolección y transmisión de datos, haciendo uso de una aplicación instalada en un smartphone android.
- Se desarrolló un algoritmo para recepción de datos de sensores haciendo uso de tecnologías web, para garantizar la disponibilidad de la aplicación en dispositivos compatibles.
- Se diseñó una arquitectura cliente-servidor para transmitir y procesar datos en línea.
- Se desarrolló interfaz gráfica para el sistema, haciendo especial énfasis en la accesibilidad y permitiendo la mayor autonomía posible para los potenciales usuarios.
- Se sometió a pruebas que permitieron garantizar la usabilidad del sistema, de manera a modelar iteraciones subsecuentes del sistema en base a las respuestas obtenidas y la percepción del público.

6.1. Trabajos futuros

- Mecanismos para utilizar la infraestructura dispuesta para este proyecto como base para la implementación de un sistema de conteo de pasajeros en tiempo real.

- Utilización de la información interpolada de la localización, el horario y la cantidad de personas como datos de entrada a un sistema experto basado en inteligencia artificial, capaz de proponer horarios posibles de mayor circulación de unidades de transporte.
- Implementación de un sistema de transmisión de datos para estructuras de billetaje electrónico.
- Implementación de un módulo de participación ciudadana para contrastar los resultados del sistema con situaciones externas a las mediciones con sensores.

Referencias

1. Carolina Pereira and José Oviedo, *Rastreo de transporte público interurbano de Ciudad del Este aplicando plataformas de código abierto*, FPUNE, 2017.
2. Swetha Sridharan, R. Venkatesh Prasad, and S. Srinarayan, *Designing a smart transport system application for South Indian traffic scenarios — A modern approach towards digitalizing the transport systems*, 2017 2nd International Conference on Communication and Electronics Systems (ICCES), IEEE, oct 2017, pp. 1–6.
3. K. Gowri Subadra, J. Mumtaj Begum, and H. Dhivya, *Analysis of an automated bus tracking system for metropolitan using IoT*, 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), IEEE, mar 2017, pp. 1–5.
4. Benjamin Y.O. Low, Samsul Haimi Dahlan, and Mohd Helmy Abd Wahab, *Real-time bus location and arrival information system*, 2016 IEEE Conference on Wireless Sensors (ICWiSE), IEEE, oct 2016, pp. 50–53.
5. Tarun Kumar, Suraj Gupta, and Dharmender Singh Kushwaha, *A smart cost effective public transporation system: An ingenious location tracking of public transit vehicles*, 2017 5th International Symposium on Computational and Business Intelligence (ISCBI), IEEE, aug 2017, pp. 134–138.
6. Sarah Aimi Saad, Amirah 'Aisha Badrul Hisham, Mohamad Hafis Izran Ishak, Mohd Husaini Mohd Fauzi, Muhammad Ariff Baharudin, and Nurul Hawani Idris, *Real-time on-campus public transportation monitoring system*, 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA), IEEE, mar 2018, pp. 215–220.
7. Oleh Boreiko and Vasyl Teslyuk, *Structural model of passenger counting and public transport tracking system of smart city*, 2016 XII International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH), IEEE, apr 2016, pp. 124–126.
8. Nandini Prasad and S P Priyanka, *Performance of tracking public transport in heterogeneous networks*, 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), IEEE, nov 2015, pp. 357–362.

Sistema de medición de energía multiusuario de alta precisión

Agustín S. Isasmendi[1], Oscar E. Goñi[2], Leandro J. Aguierre[3]
[1]Fac. Ingeniería., UNICEN, isasmendi.agus@gmail.com
[2]Instituto INTIA, Fac. Cs. Exactas, UNICEN, oegoni@exa.unicen.edu.ar
[3]Redimec SRL, laguierre@redimec.com.ar

Resumen En la actualidad, existen aún paises con alto índice de marginalidad donde es difícil el acceso a los servicios que hoy parecen básicos. En este trabajo se describe la experiencia en el diseño y construcción de un medidor de energía multiusuario de alta precisión aplicable en situaciones donde la cantidad de suministro es limitada y por lo tanto debe ser racionalizada. El proyecto presenta características relacionadas con la reconfiguración, la flexibilidad y la precisión en la medición de manera tal de poder proveer un servicio de calidad a los usuarios de energía eleéctrica. Si bien el sistema se diseñó considerando los requerimientos específicos del pais Africano Kenia, sus características modulares y reconfigurables lo hacen también aplicables en el contexto de ciudades inteligentes.

1. Introducción

Desde hace ya varias décadas que los recursos del planeta se han visto seriamente comprometidos y hace tan sólo un par de años que han comenzado a tomarse acciones paliativas. Es por eso que la comunidad científica ha tomado a las energías renovables como tema impostergable en el presente y el futuro para su investigación. Pero no sólo generar energías limpias es parte del cambio. La concientización de usuarios ha tomado fuerte relevancia dentro de las ciencias sociales. Este trabajo se aborda desde una perspectiva técnica pero intenta dar soporte a la idea de analizar el impacto social de proveer energía eléctrica a poblaciones que no poseen este servicio proveniente de fuentes renovables ni no renovables.

Kenia, un país africano sub desarrollado, posee gran cantidad de habitantes que no acceden a estos servicios ya sea por razones de distancia o carencia de infraestructura. La solución que se está abordando en estas regiones es a través de mini redes aisladas [1]. Dependiendo de la cantidad y tipo de usuarios, el sistema de generación de cada mini red combina: energía solar, grupos electrógenos y almacenamiento de energía con baterías [2]. Sin embargo, la cantidad de energía generada es limitada y por lo tanto su administración requiere de control delicado y preciso.

El presente trabajo describe la experiencia del diseño y construcción de un sistema de medición de energía prepago, multiusuario y de alta precisión que se

adapta a los requerimientos de servicios de energías renovables en comunidades que aún no cuentan con este servicio. El trabajo surge como un acuerdo comercial entre una empresa local (Redimec SRL) y una estadounidense (Renewvia Energy Corp) que provee servicios de parques solares en Africa. Si bien el trabajo se encuentra en funcionamiento en Kenia, sus carácteristicas modulares permiten su aplicación en Argentina incentivado por una poítica de ahorro energético.

El trabajo se organiza de la siguiente manera: En la sección 2 se presentan los trabajos y productos relacionados. En la sección 3 se enumeran los requerimientos tanto generales como específicos que debe cumplir un sistema de medición de energía multisuario mientras que en la sección 4 se describe de manera global el diseño. La sección 5 presenta las decisiones tecnológicas en cuanto al desaarrollo de un caso concreto. Finalmente, la sección 6 muestra resultados sobre la performance y características del equipo y el software mientras que en la sección 7 se comentan conclusiones y comentarios adicionales.

El trabajo se presenta en un contexto de colaboración entre la Universidad Nacional del Centro de la Provincia de Buenos Aires y la empresa Redimec SRL.

Vale aclarar que en el presente informe se han omitido ciertos detalles de diseño e implementación por cuestiones de confidencialidad.

2. Trabajo y productos relacionados

Tradicionalmente, la medición de consumo eléctrico se realizó utilizando métodos mecánicos. Un conjunto de bobinas inductoras generan un movimiento que permite hacer girar un disco de aluminio de forma proporcional a la corriente que circula. Conocidas las relaciones de giro vs. corriente, es posible realizar la estimación proporcional del consumo. Si bien el principio de funcionamiento es simple, a lo largo de los años, la fricción genera desgaste y esto se traduce en pérdida de precisión lo que obliga a realizar una calibración periódica de los instrumentos.

La medición de corriente eléctrica no ha sido ajena a la revolución digital de los útimos años. El trabajo presentado en [3] es uno de los precursores en la medición digital de consumo eléctrico. Allí se presenta un medidor multiusuario que permite realizar mediciones para un edificio de departamentos. En ese entonces, un operario realizaba el registro de consumos en papel, pero todos en un mismo sitio. Años más tarde, con el auge de los DSP (Digital Signal Processors) los fabricantes electrónicos comenzaron a desarrollar chips para esta aplicación específica. Mediante la integración en el tiempo de parámetros como tensión y corriente con conversores analógicos digitales, es posible conocer de manera indirecta la potencia.

El trabajo presentado por [4] presenta una aplicación donde se utilizan múltiples circuitos integrados STC11F52XE en un medidor multiusuario de enería prepaga. La información producida por estos sistemas puede ser utilizada más allá de los fines de facturación. Con determinados parámetros, es posible realizar cálculos precisos de consumos por regiones dentro de ciudades o bien consumos por franjas horarias.

Ya con la aparición de sistemas orientados a IoT (Internet Of Things), este tipo de aplicaciones resultan cada vez más importantes en el contexto de ciudades inteligentes. En este caso en particular, el proveedor de suministro eléctrico tiene la capacidad de conocer la demanda sectorizada de la red en tiempo real. Con esta información es posible realizar ajustes precisos en corrientes provistas por redes de media y alta tensión de manera dinámica. Por otro lado, si se consideran políticas de ahorro de energía, es posible tasar el servicio con diferentes tarifas considerando la franja horaria o ubicación, o bien beneficiar a usuarios que logren ahorro enegético. Trabajos como el de [5] presentan nuevos deafíos que conciernen a la privacidad de los datos producidos por este tipo de medidores.

3. Requerimientos del sistema

Un listado de requerimientos que el sistema básico debe cumplir fueron extraidos de una narrativa y de diversas entrevistas con el proveedor del servicio eléctrico.

3.1. Requerimientos funcionales

- *Conectividad remota:* Debe proveerse un esquema de comuniación que permita enlazar grandes distancias, preferentemente, inalámbica.
- *Penalización al usuario por infracciones:* En este sentido, deben diferenciarse dos tipos de infracciones:
  - Tamper: Intento de fraude al método de medición de energía. Comúnmente su detección es posible mediante medición y comparación de corrientes en ambos conductores (fase y neutro).
  - Exceso de consumo: El usuario establece por contrato su consumo máximo con el objetivo de no perjudicar a otros usuarios, dañar al equipo de medición o exceder el límite de energía que la red puede proveer. Cada usuario establece sus límites establecidos por contrato y de manera independiente a los demás usuarios.

  El incumplimiento de alguna de las condiciones de establecidas por contrato provoca la interrupción automática del servicio eléctrico. Dado que los tiempos de corte así como la potencia demandada máxima son valores establecidos por contrato entre el usuario y la empresa distribuidora, el equipo desarrollado debe proveer la funcionalidad de configuración de estos parámetros. Debe de considerarse que no se trata de un método de seguridad personal (por ejemplo, disyuntor) sino de un mecanismo de protección del equipamiento y de toda la red.
- *Reporte automático del estado del sistema:* Se debe proveer instantáneas del sistema que permitan monitorear estados de las mediciones así como variables internas para adminisración del equipo de medición.
- *Persistencia de información histórica:* Dado que el sistema está orientado a emplazarse en en zonas alejadas, debe preverse que el sistema almacene todos los registros de estados detallados de cada usuario.

- *Reporte antivandálico:* Ante el intento de destrucción ya sea mediante golpes o bien aperturas del gabinete no autorizadas, el evento deberá ser almacenado y reportado de forma inmediata. Además, el sistema debeá alertar ante a la presencia de campos magnéticos intensos que intenten alterar los métodos de medición.
- *Resistencia física:* El elemento de medicón debe soportar condiciones ambientales desérticas, lo que involucra temperaturas que oscilan entre -5 C de noche y hasta 50 grados centígrados de día. La humedad depende de la temporada y varía entre 30 y 100 %.

3.2. Requerimientos no funcionales

- *Bajo costo de mantenimiento:* El costo asociado al consumo eléctrico del propio equipo así como el costo en la transmisión de datos, debe ser despreciable ante los valores facturados.
- *Diseño extensible:* El sistema debe permitir alterar la cantidad de usuarios sin impactar sobre el hardware ni el software. Se propone la producción de equipos con incremento en la cantidad de usuarios en múltiplos de 10.

3.3. Requerimientos eléctricos y mecánicos

La flexibilidad del sistema debe estar acompañada por un sistema eléctrico que permita su reconfiguración y adaptación a otras aplicaciones. Se han establecido los siguientes requerimientos eléctricos:

- Tensión nominal: Debe permitir configurar el instrumento de medición entre 90-220V
- Corriente máxima: Debe preveerse la adaptación del sistema en consumos que van desde los 5A hasta 25A.
- Frecuencia: Según red 50-60Hz
- Clase: El tipo de medidor propuesto debe calificar como clase 2

El la actualidad, una ciudad desarrollada con un tendido eléctrico convencional posee medidores que utilizan el método de medición mecánico, obteniendo una precisión del 5 % [6].

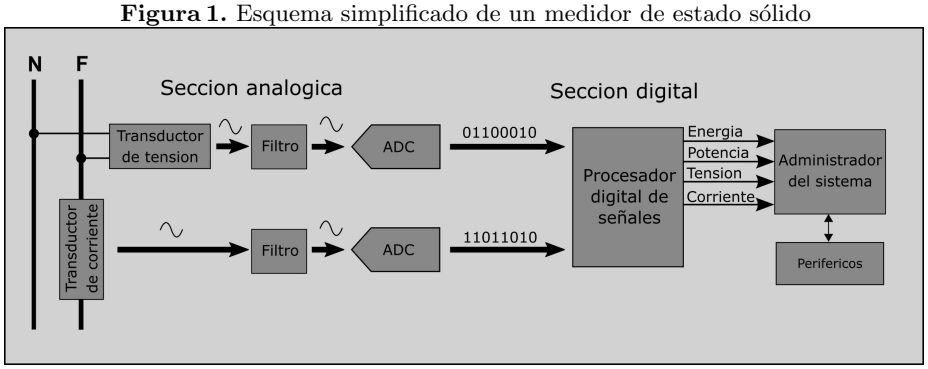
4. Diseño del sistema

El sistema fue diseñado utilizando un enfoque *bottom-up* de modo de poder aumentar la independencia de los subsistemas. Cada capa posee funcionalidades bien definidas. En esta sección primero se presenta cada parte o capa del sistema, luego se presentan las posibles topologías que se pueden obtener recombinando las partes y finalmente se presenta un caso de aplicación concreto.

4.1. Medidor mono usuario (MMU)

Se trata de un instrumento con capacidad para un único usuario y representa un productor indivisible de información. Cada medidor monousuario reporta estado instantaneo de consumo, tensión e indicadores de fraude. Este componente por sí solo representa una versión extendida de un medidor domiciliario convencional, sólo que en su versión de estado sólido.

Figura 1. Esquema simplificado de un medidor de estado sólido



4.2. Coordinador multi usuario (CMU)

Representa el agrupamiento y coordinaión de N medidores monousuario. Este dispositivo posee dos medios de almacenamiento, uno primario y otro secundario. El primario almacena los estados de tensión, corriente y potencia instantáneos, de manera tal que ante un corte de energía no se borren dichos valores. Se utilizó tecnogiía EEPROM y es posible extraer su contenido mediante sus pines en caso de destrucción total del dispositivo. El almacenamiento secundario contiene información histórica de las mismas variables pero a lo largo del tiempo. Cada conjunto de datos incluye, identificador de usuario y su conjunto de variables observadas junto con la fecha y hora de la muestra provistas por un RTC (Real-Time-Clock). Al contener información más detallada del consumo, es posible utilizar esta información para obtener estadísticas de modo de consumo (ej: Horarios) o bien como prueba ante hechos de fraude. Este instrumento, al poseer un comportamiento independiente, representa un producto por si mismo que podría utilizarse en el contexto de la inmótica en edificios de departamentos u hoteles. Provee conectividad EIA-485 para conexión con otros dispositivos o bien para su configuración y gestión (Por ejemplo: lectura mensual de consumos).

4.3. Gateway

Recibe este nombre ya que su principal objetivo es ser intermediario entre el exterior y los CMU. Se trata de un dispositivo que permite registrar, programar y configurar CMU de manera tal de proveer flexibilidad al sistema.

Con el objetivo de detectar intentos de sabotaje o aperturas no autorizadas del equipo, se agregan sensores (de apertura, magnéticos, temperatura, humedad y acelerómetros) y algoritmos para su detección. Además posee un conjunto de Entradas/Salidas de uso general configurables por el cliente a su necesidad. Este dispositivo representa el primer nivel de comunicación con el exterior. Mientras que para coordinar los módulos multiusuario utiliza una interfaz EIA-485, para la interación con otros sistemas provee interfaces Ethernet, PLC (Power Line Control), Serial y USB. La interfaz de usuario para la configuración de este equipo, así cómo los medidores multusuario y la descarga de datos in-situ, se realiza mediante un servidor web interno (interfaz ethernet) o bien mediante una aplicación Windows (USB, Serial). Al igual que los CMU poseen dos niveles de almacenamiento. El primario almacena configuraciones y variables del equipo así como de los CMU que administra. Además, permite registrar consumos máximos de modo tal de poder ser utilizado en un sistema con esquema prepago. El nivel secundario almacena instantaneas del nivel primario y funciona como cache para los datos almacenados en los CMU. Este instrumento, también desarrolla una actividad independiente y puede ser utilizado en edificios o comunidades que requieran un monitoreo *in-situ* o bien remoto siempre que exista una infraestructura de red.

4.4. Módulo de comunicaciones inalámbricas

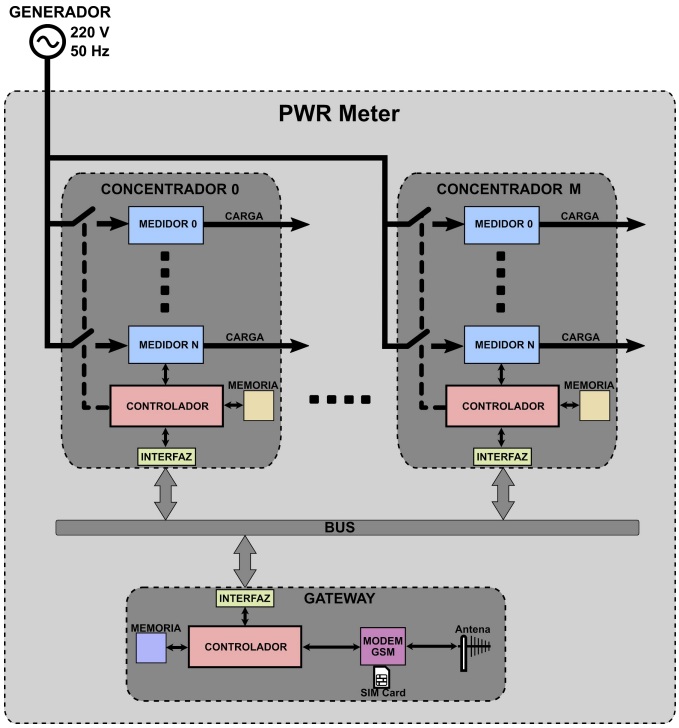
Este módulo permite la telemetría y telecontrol de cada consumidor de energía. Está basado en tenologías inalámbricas diversas de modo tal que se adecuen a redes eléctricas remotas o de difícil acceso. Entre las tecnologías soportadas se encuentran, módulo 3/4G, radio módem UHF e interfaz para teléfono satelital. Dado el abanico de posibles tecnologías, se implementó un protocolo estándard compatible con distintas tecnologías. Este módulo contiene un esquema de encriptación que brindan cierta seguridad a los datos.

4.5. Host de gestión y administración

Consiste básicamente en un servidor web que permite la gestión remota de CMUs. Posee tres sistemas principales:

- *Capa de comunicaciones:* gestiona las comunicaciones con cada CMU de manera abstracta, es decir, independientemente del medio de comunicación utilizado.
- *Capa de datos:* Registra datos de variables de cada MMU así como sus valores límites.

Figura 2. Esquema de la topología adoptada



- *Tarifador:* Gestiona los consumos y pagos de cada consumidor en particular. Posee la caracterísitca de prepago indicando la cantidad de energía máxima que puede consumir.
- *Interfaz de Gestión:* Permite el mantenimiento de los CMU de manera remota. Desde allí se agregan o eliminan usuarios, indicando la relación que tiene el cliente con la boca del CMU.
- *Alertas:* Permite visualizar las alertas producidas por el sistema. En el caso particular de intento de fraude, este módulo sólo reporta la situación ya que la acción de interrupción del servicio se realiza de manera automática. En el caso de un evento vandálico, la alerta debe evaluarse para informar a la autoridad de seguridad competente.

5. Tecnología utilizada

En el proyecto descripto, se utilizaron tecnologías que involucran tanto software embebido como aplicaciones ejecutándose en la nube. En esta sección se da una breve descripción del hardware involucrado así como algunos detalles de la arquitectura de software utilizada en cada nivel.

5.1. Coordinadores multiusuario

Se presenta una arquitectura hardware *Maestro-Esclavos* que involucra un microcontrolador de 32 bits (Maestro) encargado de coordinar el funcionamiento de, en este caso en particular, 10 medidores de estado sólido. Para este trabajo en paritcular se eligió un *N=10*. En particular, se eligió empíricamente esta granularidad de manera tal de cubrir una vereda convencional. Dado que la encuesta a cada medidor se realiza a un periodo constante y de velocidad (200 hz) considerable, sus registros deben ser muestreados en periodos acordes para evitar el desborde de los mismos. Es por eso que el periodo de muestreo se fijo considerando la mitad del tiempo en que los registros desborden a máximo consumo. La persistencia de datos se realiza en nivel primario con una memoria EEPROM de 256 Mbit. Considerando que la tecnología permite acceso aleatorio, las magnitudes de cada medidor específico son almacenadas en palabras de 32 bits y mediante un cálculo simple es posible acceder directamente a la tupla *Medidor-Magnitud*. Cada palabra admite dos dígitos de precisión. La memoria secundaria está implementada con tecnología Flash NOR, por lo cual las instantáneas son almacenadas en forma secuencial y circular, eliminando los registros más antiguos.

El software se ejecuta sobre un sistema operativo de tiempo real que permite, no sólo el muestreo de los múltiples esclavos en períodos precisos, sino que también es el encargado de la recuperación ante un desperfecto técnico o bien interrupción del suministro. Además gestiona el almacenamiento de instantáneas en memoria secundaria.

5.2. Gateway

Permite la gestión de hasta 20 CMU muestreados a intervalos de tiempo de 15 minutos. El hardware consta de un microcontrolador de 32 bits e idéntico almacenamiento que los CMU. Además provee una interfaz Ethernet de 100 mbps para remota de los datos o bien para acceder a la configuración del dispositivo mediante un servidor web interno.

Dado que la información almacenada por este dispositivo es consultada esporádicamente, tanto el refresco de registro de magnitudes de cada usuario así como el almacenamiento de instántaneas, se realiza a intervalos de 15 minutos permitiendo además pasar el equipo a modo de ahorro de energía. Todas estas características son posibles gracias a que el software embebido también ejecuta un sistema operativo de tiempo real.

6. Resultados

El proyecto se llevó a cabo en un período de 6 meses. En esta sección se detallan algunas métricas del sistema final obtenido.

6.1. Pruebas físicas

Las pruebas del sistéma anivandálico fueron llevadas a cabo aplicando sucesivos golpes sobre la estructura. Además se le aplicaron neodimios sobre la caja para detectar fraudes de consumo. Por otro lado, toda la electrónica fue sometida a estrés térmico variando de -40C a 85C controlando humedad y temperatura.

6.2. Desempenño MMU y CMU

El requerimiento de precisión inicial se fijó en el 2 %. Inicialmente, se construyó el equipo y luego se sometió una etapa de calibración donde todos los medidores fueron ensayados. El cuadro 6.2 muestra los resultados obtenidos antes y luego de la calibración. El equipo utilizado para las pruebas de calibración es el Schlumburger SM3050 [7].

Cuadro 1. Errores antes y después de la calibración

	Error en tensión	Error en corriente	Categoría
Sin calibrar	5.2 %	4.2 %	5
Calibración 1	0.3 %	0.4 %	0.5
Calibración 2	0.2 %	0.108 %	(>) 0.25

Respecto a la performance de los CMUs se sobredimensionaron los tiempos para el muestreo para evitar el desborde. El cuadro 6.2 presenta los tiempos utilizados, sobredimensionados y el tiempo mínimo en el que podría realizarse la muestra. Esto permitiría que el equipo realice mediciones de hasta más de 20 veces el consumo para el cual fue diseñado o bien para el doble de medidores multiusuario.

Cuadro 2. Tiempos de muestreo

	Periodo de desborde	Periodo de muestra	Periodo máximo
CMU	50 ms	5 ms	2 ms

El cuadro 6.2 resume la capacidad de persistencia del CMU. La capacidad de almacenamiento excede lo establecido, permitiendo a sistemas posteriores, aumentar la cantidad de consumidores.

Cuadro 3. Utilización de almacenamiento secundario

bytes/usuario	bytes/hora (30 usr)	capacidad
5	360	∼1 año

6.3. Desempeño server

El software de gestión, administración y tarifador se desplegó en una instancia de Amazon Web Services EC2 [8]. La desición de elegir esta tecnología se debe a las prestaciones que ésta presenta en lo referente a confiabilidad y disponibilidad. Además, dado que el sistema se implantará geográficamente distante, es necesario contar con una plataforma que permita elegir donde se ejecutará software. Este subsistema realiza sondeos periódicos por la salud de los dispositivos remotos mientras que realiza actualizaciones sobre facturación y estado de cuenta de cada uno de los consumidores que provee. En cuanto al volumen de datos transferidos desde y hacia el dispositivo, se realizan transmisiones periódicas (15 minutos) de paquetes de un kilobyte en ambos sentidos. Esto se debe a que se eligieron cuidadosamente las estructuras y tipos de datos que son utilizados para representar la información. Esto permite realizar un ahorro significativo en el paquete de datos elegido si la comunicación se realiza a través de la red GSM. El tiempo de transmisión de tal paquete es del orden del segundo, despreciable en una plataforma 3G.

7. Comentarios finales y Conclusiones

En el presente texto, se describe la experiencia en el diseño y consrtucción de un medidor de enería multiusuario. El dispositivo se diferencia con otros dispositivos desarrollados en su alta precisión, flexibilidad y escalabilidad.

Cada medidor de estado sólido en particular fue evaluado antes y después de un proceso de calibración. El equipo sin calibrar presentó una precisión idéntica a los medidores domiciliarios tradicionales (error del 5 %) mientras que el equipo calibrado, presentó precisiones de 10 y 20 veces mejores.

Si bien se sobredimensionaron parámetros de diseño (Por ejemplo: tiempo de refresco de medidores y almacenamiento histórico), esta situación permite al dispositivo estar preparado para prestar servicios a grandes consumidores o en diferente configuración con sólo ligeras modificaciones en el Hardware y nulas en el Software.

Luego de completar los trámites de exportación necesarios, el equipo fue enviado al cliente quien realizó una prueba de aceptación. Al momento de redacción de este texto, el dispositivo se encuentra en proceso de instalación en Kenia.

Referencias

1. J. Nkiriki and T. S. Ustun. Mini-grid policy directions for decentralized smart energy models in sub-saharan africa. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6, Sept 2017.
2. M. Byamukama, R. Akol, G. Bakkabulindi, B. Pehrson, R. Olsson, and J. Sansa-Otim. Energy storage options for environment monitoring wireless sensor networks in rural africa. In *2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018)*, pages 1–5, April 2018.

3. Liang Zhishan, Qiu Changtao, and Wang Jinli. A new type intelligent multi-user single phase kwh meter. In *[1992] Proceedings of the IEEE International Symposium on Industrial Electronics*, pages 551–554 vol.2, May 1992.
4. Wei Sun and Jian Zhi Tuo. Prepayment electric energy meter with multi-functional for multi-user based on stc11f52xe. 367:211–215, 08 2013.
5. J. Gmez-Vilardeb and D. Gndz. Smart meter privacy for multiple users in the presence of an alternative energy source. *IEEE Transactions on Information Forensics and Security*, 10(1):132–141, Jan 2015.
6. Instituto Nacional de Tecnología Industrial INTI. Medidores de energía elćtrica activa de inducción, clases 0,5; 1 y 2 (segn norma iec 521/88), June 2011.
7. Schlumberger. Sm 3050 user manual), June 1988.
8. Amazon. Amazon elastic compute cloud (amazon ec2), June 2018.

Detección de Averías Viales Mediante IoMT Aplicada a Smart Cities

Miguel Augusto Azar[1], Maria Antonia Tapia[1]

[1] Universidad Nacional de Jujuy, Jujuy Argentina
augusto.azar@gmail.com; maria.antonia.tapia@gmail.com

Abstract. El presente trabajo describe la primera fase del desarrollo de un sistema smartcity que permite registrar las posiciones globales y las dimensiones de las averías viales dentro de una ciudad utilizando una red de sensores móviles. Se aprovecha la posibilidad de acceder al acelerómetro y al GPS de los smartphones que emplean los conductores de vehículos. La magnitud de los baches se mide por medio de la intensidad de señal registrada por el acelerómetro mientras que el lugar donde se encuentra, lógicamente, se almacena según las coordenadas provistas por el GPS. Esta fase inicial es escalable y permitiría a los administradores municipales contar con información actualizada del estado de las calles para su correspondiente mantenimiento o reparación.

Keywords: Internet de las Cosas. Sensores. Smart City. Mobile. Android. IoT.

1 Introducción

IoT (por su nombre en inglés Internet of Things) consiste en una red de objetos conectados a Internet capaces de recopilar e intercambiar datos utilizando sensores integrados. Dichos sensores se utilizan en redes eléctricas, ferroviarias, sistemas de conductos de agua potable, distribución de gas, entre otras. A su vez, estas redes no necesariamente poseen sensores estáticos sino que pueden ser parte de sistemas en los que los sensores suministran información a medida que se desplazan dentro de un área específica. Estas redes de sensores inalámbricos (WSN, o Wireless Sensor Network) [5] se conectan a la nube y conforman lo que se conoce como Internet de las Cosas Móviles (IoMT).

El término “Internet de las Cosas” fue mencionado por primera vez por Kevin Ashton en 1999 [1]. Este concepto hace referencia a que las computadoras dependen de los humanos para obtener información. Aun así, estos últimos no lo hacen de manera dedicada, precisa u oportuna como lo hacen los sensores que se conectan a sistemas embebidos.

En [2] IoT se define como aquel sistema que se utiliza para cubrir diversos aspectos relacionados con la extensión de Internet y la Web en el ámbito físico, mediante el despliegue generalizado de dispositivos distribuidos espacialmente con capacidades de identificación, detección y/o actuación integradas.

Otros autores tales como [3], consideran que IoT es el nombre colectivo de los sensores interconectados, colocados en "cosas" dentro de las infraestructuras cibernéticas físicas fijas.

Actualmente, una gran cantidad de infraestructuras son físicamente dinámicas en donde los sensores pueden estar asociados a personas o cosas que se mueven permanentemente y proporcionan datos relativos al lugar en que se encuentran. Este concepto es complementado por [4] en el que se amplía el paradigma de Internet de las Cosas hacia Internet de las Cosas Móviles (IoMT). En base a esta idea y al aprovechamiento de la conectividad de Internet cada vez más extendida es posible desarrollar sistemas que proporcionen información dinámica en función del posicionamiento de las cosas y las personas.

Una de las problemáticas que algunas de las “cosas” móviles enfrentan a diario es el de las irregularidades de calles y rutas (Figura 1), las cuales provocan daños y molestias que podrían ser evitadas.



Fig. 1. Baches de gran magnitud en carreteras.

Las averías viales constituyen un gran problema tanto para quienes deben repararlas como para los vehículos involucrados. En la mayoría de los casos el mantenimiento de las calles y rutas obliga a deshabilitar algún tramo de las mismas ocasionando mayor congestión del tránsito que la habitual.

En cuanto a los vehículos, los daños producidos por las calles y rutas en mal estado (en casos extremos) pueden ir desde la rotura de amortiguadores y llantas, la expulsión del aire de las ruedas, hasta incluso la posibilidad de accidentes entre vehículos y transeúntes. En el mejor de los casos los daños en el vehículo pueden ser graduales y con el paso del tiempo deteriorar lentamente el tren delantero, reduciendo de este modo su vida útil.

En el presente trabajo se plantea la primera etapa del desarrollo de un sistema que permite identificar las averías viales dentro de una ciudad. El concepto consiste en el diseño de una app para smartphones que, una vez instalada, realiza la lectura de los impactos bruscos (provocados por las irregularidades del camino) que recibe un automóvil a lo largo de un trayecto dentro o fuera de una ciudad. La lectura de esos impactos es tomada a través del acelerómetro que informa sus mediciones permanentemente a una aplicación y ésta a su vez registra esos datos para

posteriormente ser analizados. Cada vez que el sensor proporciona un dato que supera determinado umbral la aplicación consulta la posición global actual (GPS) y registra tanto la intensidad del impacto como la latitud y longitud en donde se produjo.

2 Antecedentes

Existe, desde la incorporación del acelerómetro a los smartphones, una gran cantidad de investigaciones intentando abordar la problemática de las averías viales mediante el uso de aplicaciones para sistema operativo Android y en diversas configuraciones en cuanto a arquitectura.

En [6], los autores se concentran en diferentes algoritmos de detección de baches basados en los datos del acelerómetro de smartphones con sistema operativo Android y realizan una clasificación de baches (pequeños, grupo de baches y grandes). De este modo logran una identificación de los daños en calles en forma exitosa, de hasta un 90%.

En el trabajo presentado por [7], los autores se enfocan en un análisis técnico profundo y desarrollan un sistema de adquisición de datos que son procesados y sometidos al cálculo de un índice de rugosidad compuesto en base a un modelo matemático utilizado para extraer información de las condiciones de la superficie de una carretera.

Otra propuesta es la desarrollada por [8] en la que se diseña una combinación de hardware y software compuesta por acelerómetro, unidad de GPS, microcontrolador, módulo Zigbee [9] y servicio en la nube.

En [10] se emplea un conjunto hardware consistente en una unidad de GPS, acelerómetro, Arduino y un servidor que es actualizado mediante el uso de un módulo Bluetooth. El enfoque para detectar baches esta basado en funciones definidas en [6].

El trabajo presentado por [11] emplea una aplicación móvil que además del acelerómetro utiliza el giroscopio. A partir de los datos recolectados se entrenan modelos SVM (Support Vector Machines) logrando una precisión superior al 90%.

Otros trabajos similares a los ya citados son los estudiados por [12]-[14]. En [15] se describe de manera muy simplificada la publicación de una patente de un sistema embebido que emplea acelerómetro, GPS y microprocesador.

3 Arquitectura Propuesta

En esta primera fase del desarrollo, la investigación se enfocó en la implementación de una aplicación móvil destinada a recolectar la información acerca de las averías viales, su geo posición y el almacenamiento de los datos de aceleración, latitud y longitud. El conjunto de datos luego se somete a un análisis para posteriormente determinar los pasos a seguir en futuros trabajos.

En la Figura 2 se ilustra la arquitectura básica del sistema desarrollado la cual consiste en una aplicación implementada para el sistema operativo Android 6.0 Marshmallow.

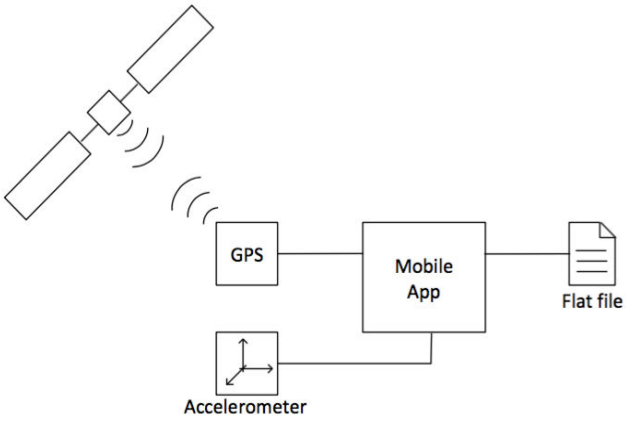


Fig. 2. Arquitectura desarrollada para la primera fase.

4 Implementación

Para el desarrollo de la app se utilizó el entorno de desarrollo integrado Android Studio 3.0 y el emulador de aplicaciones móviles Genymotion 2.12 debido a su versatilidad frente a emuladores similares. La aplicación fue instalada y probada exitosamente en un smartphone Samsung J7 Prime.

En la Figura 3 puede observarse la interfaz principal de la aplicación móvil en la que es posible iniciar la lectura de datos una vez que se fija el smartphone al interior del automóvil.



Fig. 3. Interfaz principal de la app desarrollada para la primera fase de recolección y análisis de datos.

En la Figura 4 se ilustran las pantallas del sensado de los ejes del acelerómetro, las magnitudes proporcionadas por el GPS y el registro en archivo plano.

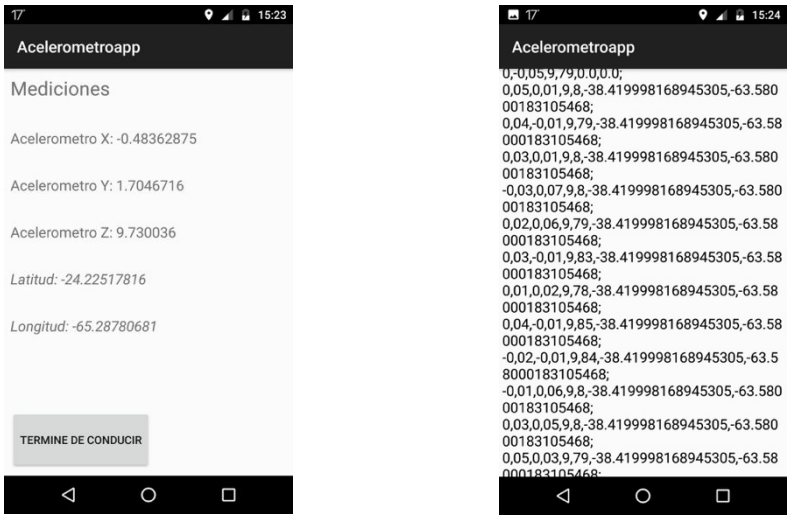


Fig. 4. Activities de lectura de acelerómetro y GPS (izq., monitorea la medición que se esta realizando) y los datos consultados una vez que finaliza el proceso de registro (der.).

5 Experimentación In Situ

Las pruebas se realizaron en una ruta poco transitada en un tramo de 2.6 km. La Figura 5 ilustra la zona analizada en una serie de mediciones a diferentes velocidades.

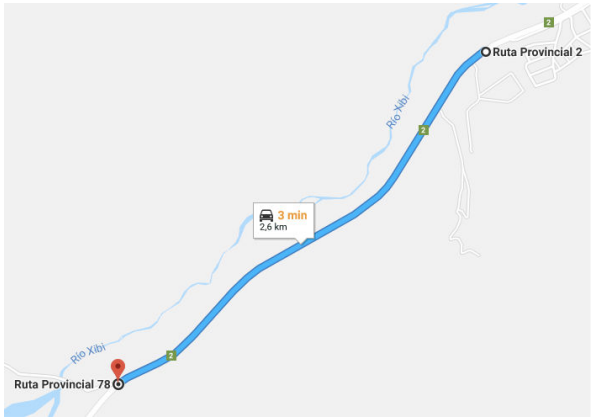


Fig. 5. Tramo de la carretera utilizada para la recolección de los datos de aceleración, latitud y longitud.

El análisis de los datos extraídos se dividió en dos partes en las que se evalúa el efecto producido por las irregularidades:

- en forma aislada
- en forma general

A su vez, el análisis (en forma aislada) de los efectos de la aceleración sobre el eje Z se evaluaron en función de las irregularidades de la superficie cuando poseen:

- elevación anormal (reductores de velocidad vial)
- hundimiento anormal (bache)

En el caso de las elevaciones anormales, el efecto en la aceleración del eje Z acompaña al movimiento de la elevación; Figura 6. El eje Z normalmente se mantiene alrededor de los 9.7 m/s^2, pero cuando el vehículo encuentra un obstáculo por encima de la superficie entonces la aceleración tiene una variación positiva que depende de la velocidad y de la magnitud del obstáculo.

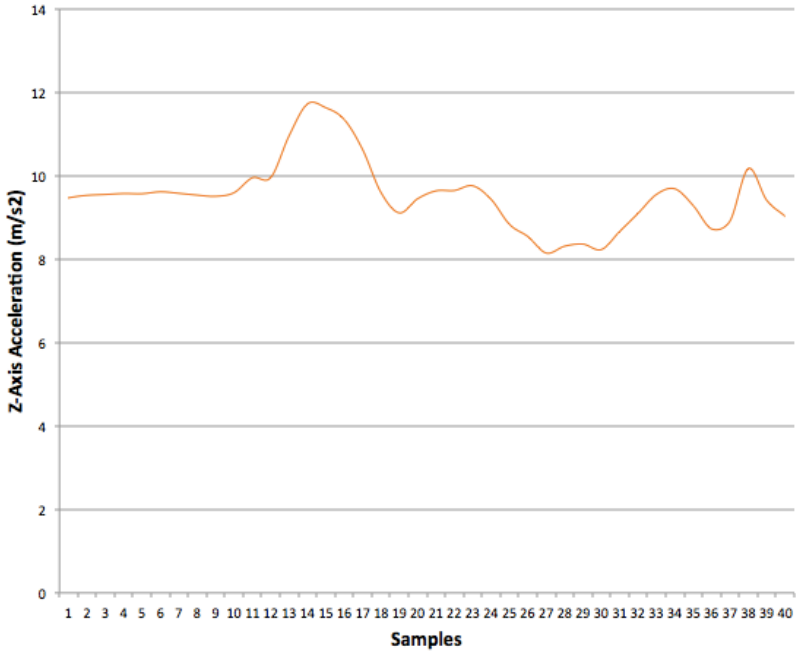


Fig. 6. Efecto sobre el eje Z provocado por un reductor de velocidad vial a baja velocidad (40 km/h).

En la Figura 6 puede observarse que luego de mantenerse la aceleración en 9.7 m/s^2 se eleva a un valor cercano a 12 m/s^2.

En el caso de la aceleración frente a los baches se produce el efecto contrario (Figura 7).

Al encontrarse el vehículo frente a un bache ocurre una variación negativa de la aceleración. En la Figura 7 se observa que la aceleración pasa de su valor normal (9.7 m/s^2) a -0.075 m/s^2. La variación positiva inicial se debe al movimiento de cabeceo que realiza el vehículo dado que la medición es tomada en el interior del mismo

(cerca de su baricentro). Una medición mas precisa se lograría si el acelerómetro fuese instalado sobre el eje de las ruedas delanteras o traseras.

En cuanto al análisis general, en la Tabla I se muestra un resumen de las mediciones promedio realizadas a velocidades de 40 y 80 km/h. Puede observarse que a mayor velocidad del automóvil mayor es la intensidad de la aceleración negativa. Mas específicamente, cuando el vehículo se desplaza a una velocidad de 80 km/h el impacto recibido tiene mayor intensidad debido a que la aceleración en el eje Z cae abruptamente a 1.05824 m/s^2.

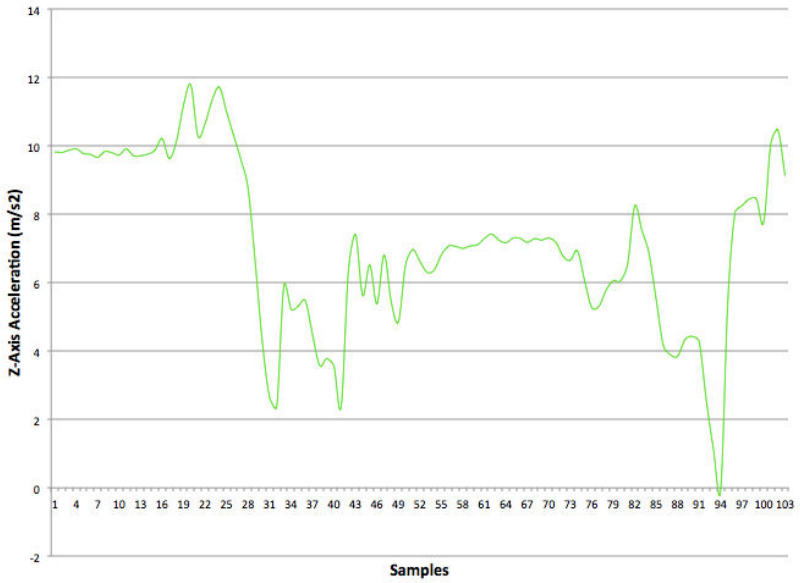


Fig. 7. Efecto sobre el eje Z provocado por un bache a baja velocidad (40 km/h).

Cuando la velocidad del automóvil es mas baja (40 km/h), el impacto es menor registrando una caída de la aceleración en el eje Z promedio de 5.60483 m/s^2.

Tabla I. Aceleración promedio en función de la velocidad.

		Velocity (km/h)	
		40	**80**
Acceleration (m/s^2)	**min**	5.60483	1.05824
	max	11.79503	12.64139

6 Conclusión

Este análisis previo confirma que las altas velocidades inciden con mayor intensidad sobre los vehículos y en consecuencia la probabilidad de daño sobre el mismo se incrementa.

Se analizaron los efectos provocados por los reductores de velocidad vial y los baches determinando que la curva de la aceleración acompaña a la forma de la irregularidad en la ruta. Este análisis permitirá en futuros trabajos identificar con precisión la presencia de baches mediante el empleo de técnicas de inteligencia artificial tales como aprendizaje automático o el aprendizaje profundo.

Por otro lado, en futuros trabajos se desarrollará un sistema embebido con sensores ubicados en los ejes del vehículo para lograr una mejor lectura de las irregularidades de las carreteras.

El uso de un sistema de esta naturaleza requiere la implementación futura de una red de automóviles conectados que informen el estado de las calles en forma permanente. Esto conlleva a una situación polémica ya que el uso de una aplicación que suministra información geo posicionada de personas a una base de datos en forma permanente puede tener implicancias cuestionables relativas a la privacidad de quienes la utilizan. Es por ello que, si un sistema de este tipo se requiere instalar en vehículos para que trabajen en forma colaborativa entonces los administradores municipales podrían plantear una solución intermedia. Esto significa que una posibilidad es recurrir a aquellos vehículos que tienen un espectro de alcance más amplio y mejor distribuido como ser el caso de los taxis a los cuales podrían aplicárseles exenciones impositivas en caso de colaborar con la instalación y posterior suministro de información.

En relación al uso del acelerómetro y su factibilidad, se pudo constatar que se requiere medir, además de la aceleración, los movimientos de cabeceo y rolido mediante el giroscopio del smartphone. Otra posibilidad, consiste en instalar sensores piezoeléctricos adosados a los amortiguadores con el fin de lograr una precisión superior.

Referencias

1. K. Ashton. That 'Internet of Things' Thing: In the real world, things matter more than ideas. RFiD Journal. RFID. (2009).
2. C. Tsirmpas, A. Anastasiou, P. Bountris and D. Koutsouris. A New Method for Profile Generation in an Internet of Things Environment: An Application in Ambient-Assisted Living. IEEE Internet of Things Journal. Volume: 2, Issue: 6. Page(s): 471 – 478. ISSN: 2327-4662. (2015).
3. K. Nahrstedt, H. Li, P. Nguyen, S. Chang, and L. Vu. Internet of mobile things: Mobility-driven challenges, designs and implementations. In Proceedings - 2016 IEEE 1st International Conference on Internet-of-Things Design and Implementation, IoTDI 2016 (pp. 25-36). [7471348] Institute of Electrical and Electronics Engineers Inc.. DOI: 10.1109/IoTDI.2015.41. (2016).
4. K. Nahrstedt. Internet of mobile things: challenges and opportunities. 10.1145/2628071.2635931. (2014).

5. M. U. Farooq, M. Waseem, S. Mazhar, A. Khairi and T. A. Kamal. Review on Internet of Things (IoT). International Journal of Computer Applications (0975 8887). Volume 113, No. 1. (2015).
6. A. Mednis, G. Strazdins, R. Zviedris, G. Kanonirs, and L. Selavo. Real time pothole detection using android smartphones with accelerometers. In IEEE DCOSS. (2011).
7. G. Alessandroni, L. C. Klopfenstein, S. Delpriori, M. Dromedari, G. Luchetti, B. D. Paolini, A. Seraghiti, E. Lattanzi, V. Freschi, A. Carini and A. Bogliolo. SmartRoadSense: Collaborative Road Surface Condition Monitoring. In Proc. of UBICOMM. (2014).
8. H. A. Jamakhandi and K. G. Srinivasa. Internet of things based real time mapping of road irregularities. Assistive Devices Group M. S. Ramaiah Institute of Technology, Proceedings of International Conference on Circuits, Communication, Control and Computing. I4C. (2014).
9. Zigbee alliance, https://www.zigbee.org/what-is-zigbee/494-2/
10. S. Dimple, V. Monica, A. Anirudh and C. Adarsh, "Monitoring Of Road Irregularities Using IOT", International Journal of Management and Applied Science (IJMAS), pp. 127-130, Special Issue. (2016).
11. U. Bhatt, S. Mani, E. Xi and Z. Kolter. Intelligent Pothole Detection and Road Condition Assessment. Data For Good Exchange. (2017).
12. A. S. El-Wakeel, J. Li, M. T. Rahman, A. Noureldin and H. S. Hassanein, "Monitoring road surface anomalies towards dynamic road mapping for future smart cities", Signal and Information Processing (GlobalSIP) 2017 IEEE Global Conference on, pp. 828-832. (2017).
13. B. Lanjewar et al., "Road bump and intensity detection using smartphone sensors", Int. J. Innov. Res. Comput. Commun. Eng., vol. 4, no. 5, pp. 9185-919. (2016).
14. H. Song, K. Baek and Y. Byun. Pothole Detection using Machine Learning. Advanced Science and Technology Letters Vol.150 (AST 2018), pp.151-155. (2018).
15. H. A. Troemel and M. R. Stelts, “Road Health (Pothole) Detection over Wireless Infrastructure”, U.S. Patent 15/662,753, Jul. 28. (2017).

Análisis del diseño adaptativo en sitios web gubernamentales

Roxana Martinez[1], Rocío Rodríguez[1,2], Pablo Vera[1,2]

[1] Centro de Altos Estudios en Tecnología Informática (CAETI).
Universidad Abierta Interamericana, Ciudad Autónoma de Buenos Aires, Argentina
{Roxana.martinez; rocioandrea.rodriguez: pablomartin.vera} @uai.edu.ar

[2]Grupo de Investigación, Desarrollo y Formación en Innovación de Software (GIDFIS)
Departamento de Ingeniería e Investigaciones Tecnológicas, San Justo, Buenos Aires, Argentina
{rocio.rodriguez; pvera} @unlam.edu.ar

Resumen. Hoy en día, los sitios web gubernamentales son de gran importancia para un mejor acercamiento entre las entidades gubernamentales y las personas. Por otra parte, en el sector público, cada vez son más las entidades estatales que apuestan al desarrollo tecnológico para mejorar la comunicación entre el Estado Nacional Argentino y los ciudadanos. La tecnología puesta al servicio del ciudadano conlleva a la implementación y mejora continua del Gobierno Electrónico, posibilitando servicios virtuales que se encuentren al alcance de las personas mediante los distintos sitios web, y por otra parte, el acceso a una fuente de información pública más precisa y actualizada. Con la alta inserción de los dispositivos móviles, el diseño adaptativo se torna fundamental al construir un sitio web. Este artículo considera casos de estudio en donde se analiza el diseño web adaptativo, detectándose problemas que dificultan al usuario la utilización y correcta visualización, de los sitios web gubernamentales. Consideramos que este estudio puede ser replicado por otros colegas a fin de validar otros sitios web gubernamentales para evaluar su estado general.

Palabras Claves: Diseño Adaptativo, Gobernanza, Sitios Web, Normativas.

1. Introducción

La inclusión de las Tecnologías de la Información y la Comunicación (TIC) en la Argentina, se ha convertido en uno de los nexos fundamentales entre los ciudadanos y el Gobierno Nacional. “En todo el mundo las nuevas Tecnologías de la Información y la Comunicación (TIC) están dando lugar a profundas transformaciones socioculturales que afectan tanto a las sociedades y a sus gobiernos, como a sus industrias, sus comunidades y sus individuos” [1]. Esto conduce al concepto de “Gobierno Electrónico que es la transformación de todo el gobierno como un cambio de paradigma en la gestión gubernamental, este concepto de gestión que fusiona la

utilización intensiva de las TIC, con modalidades de gestión, planificación y administración, como una nueva forma de gobierno" [2].

El sistema de modernización a nivel nacional en Argentina, propone una gestión innovadora con una plataforma tecnológica de datos abiertos y un acceso a la información integrada, la cual es aplicada a toda la administración estatal. Para llevar a cabo dicho proyecto, es de vital importancia, la concientización de las personas frente a esta nueva visión de acceso a la información pública, brindando una propuesta en tiempo real, en forma clara y precisa. Es una buena opción, "la posibilidad de que sean los mismos ciudadanos quienes se organicen al margen de los partidos u otras instituciones o que colaboren abiertamente con las organizaciones públicas o bien, que diseñen y dispongan de herramientas para controlar al "poder establecido" y la actividad de sus representantes (vigilancia activa)" [3]. Esto permite tener un "Gobierno Abierto que establece una constante conversación con los ciudadanos con el fin de oír lo que ellos dicen y solicitan, y que toma decisiones basadas en sus necesidades y preferencias, que facilita la colaboración de los ciudadanos y funcionarios en el desarrollo de los servicios que presta y que comunica todo lo que decide y hace de forma abierta y transparente" [4].

Una de las características fundamentales del Gobierno Electrónico es la accesibilidad a cualquier persona, en cualquier momento y en cualquier lugar, es decir, que los servicios públicos e información sobre los mismos sean fácilmente accesibles por los ciudadanos, por otro lado, se encuentra "la capacidad de respuesta (receptividad) desde el Gobierno/administración a las nuevas demandas, ideas y necesidades que presenta la ciudadanía y la oportunidad de participar en el proceso decisorio" [5].

Como se mencionó anteriormente, las características fundamentales para llevar a cabo un Gobierno Abierto son: transparencia, participación y colaboración. Con el fin de llegar a cada una de las personas es importante que el contenido sea transparente y público, por lo que son características necesarias en los sitios web. "En torno al quehacer gubernamental, las herramientas que son utilizadas con mayor frecuencia en el Gobierno Electrónico son los sitios web (incluidos los portales de transparencia, datos abiertos y de compras públicas), a los que se han agregado ahora el uso de aplicaciones" [6].

Junto a la implementación de los sitios web, es necesario incorporar cuestiones técnicas en el diseño, debido a que los usuarios/ciudadanos ingresarán a los mismos desde distintos dispositivos y la solución debería ser adaptable. En la sección 2 se aborda el tema de diseño adaptativo. Luego en la sección 3, se abordará la temática de sitios web gubernamentales, su importancia y el marco normativo nacional en Argentina, en la sección 4, se mostrarán algunos sitios web con diseño adaptativo seleccionados como casos de estudio, y finalmente se presentan las conclusiones alcanzadas.

2. Diseño web adaptativo

El diseño web adaptativo permite utilizar un diseño visual para el usuario, dependiendo del tamaño de la pantalla del dispositivo que se esté utilizando. En la Figura 1, se muestra un ejemplo de diseño adaptativo, visualizando los distintos tamaños de pantalla que se pueden tener. Como se puede observar, se adaptan los elementos que componen la página web, es decir, dependiendo del tamaño de pantalla del dispositivo móvil en la que se visualice, sus componentes se acomodan para una correcta visualización y accesibilidad de la información.



Fig. 1. Diseño Web Adaptativo en distintos dispositivos.

Sin lugar a dudas, la inclusión de los dispositivos móviles es cada vez más notoria en los distintos ámbitos utilizados por los usuarios, desde consultas varias, temas referidos a salud, juegos, trámites gubernamentales, entre otros. La tecnología móvil es un área que ha tenido un importante crecimiento y adopción por parte de las personas en sus vidas cotidianas, siendo los dispositivos más destacados las tablets y los smartphones.

"El diseño web adaptativo o adaptable (en inglés, Responsive Web Design) es una técnica de diseño y desarrollo web que, mediante el uso de estructuras e imágenes fluidas, así como de media-queries en la hoja de estilo CSS, consigue adaptar el sitio web al entorno del usuario" [7], por lo que un sitio se podrá visualizar sin inconvenientes, indistintamente del tamaño del dispositivo en el que se esté ejecutando.

Básicamente, para otorgar un diseño adaptable, se deben llevar a cabo estructuras e imágenes fluidas, valiéndose de HTML5, y las hojas de estilo CSS, que consigue adaptar el sitio web al entorno del usuario. Uno de los conceptos fundamentales es el de Media Queries, que corresponde a las herramientas que se encargan de aplicar diferentes estilos para los dispositivos móviles, proporcionando una experiencia de diseño más amigable para cada tipo de usuario móvil. Básicamente, se crean puntos de ruptura o breakpoints en las hojas de estilos CSS, es decir, detectan el tipo de dispositivo y en base a ello el contenido consigue adaptarse a un determinado tamaño de pantalla. Algunas consideraciones pueden ser el ancho y alto del browser, o bien del ancho y alto del dispositivo móvil. Estos son los puntos donde el diseño cambiará de forma, es decir, se adaptará a los distintos anchos de pantallas y resoluciones. En la Tabla 1, se muestra una propuesta de Ethan Marcotte en su libro, Responsive Web Design, de puntos de quiebre [8].

Una premisa fundamental, es entender que "el diseño adaptativo no es una versión diferente del sitio web, sino que es el mismo sitio que se va adaptando y acomodando sus elementos para una mejor visualización en el dispositivo de acceso. Con lo cual no cuenta con todas las ventajas que tendría la construcción de un sitio móvil" [9].

Tabla 1: Puntos de quiebre propuestos por Ethan Marcotte.

Píxeles	Descripción
320	Para dispositivos de pantalla pequeña, como los celulares en modo vertical
480	Para dispositivos de pantalla pequeña, como los celulares en modo horizontal.
600	Tables pequeñas, 600x800 y 600x1024, en modo vertical.
768	Tables de 10 pulgadas como los iPad (768x1024), en modo vertical.
1024	Tables, iPad (1024x768), en modo horizontal, algunas laptop, netbook y monitores.
1200	Para pantalla anchas.

Para el desarrollo de sitios web móviles el W3C (Consorcio Web a Nivel Internacional) brinda una guía de buenas prácticas [10] y, además, establece el conjunto de características mínimas que un dispositivo móvil debe cumplir para navegar por Internet. Siguiendo esas pautas, es viable diseñar y desarrollar sitios web móviles accesibles desde cualquier dispositivo. A éstas debemos sumar las ventajas en la utilización del diseño adaptativo. Algunas de las ventajas principales del diseño web adaptativo son:

- Mediante una única versión en HTML y las hojas de estilo CSS necesarias, se pueden contemplar la mayoría de las resoluciones de pantalla de los dispositivos móviles estándares, por lo que la experiencia del usuario es totalmente transparente.
- El mantenimiento de los sitios web es mucho menor, ya que no se toma en cuenta las versiones diferentes. Un único sitio es el que se construye y será visualizado correctamente en cada dispositivo.

Cabe destacar que un mal uso del diseño adaptativo incorporará algunos errores a tener en cuenta como ser: (1) Peso de la página, (2) Uso y abuso de imágenes, (3) Imágenes no preparadas para dispositivos móviles, (4) Contenido no priorizado, y (5) Páginas demasiado grandes [11]. Para realizar un correcto diseño de un sitio web utilizando diseño adaptativo es preciso considerar algunos principios que resultan fundamentales, propuestos por [11]. A continuación, se muestran en la Figura 2, las pautas para el Diseño Adaptativo con el Enfoque Mobile First.



Fig. 2. Pautas para el Diseño Adaptativo con el Enfoque Mobile First.

3. Sitio web gubernamentales

"La tecnología puesta al servicio ciudadano, conlleva a la implementación y mejora continua del Gobierno Electrónico, posibilitando servicios virtuales que se encuentren al alcance de los ciudadanos y por otra parte el acceso a una fuente de información pública más precisa y actualizada. Mediante el uso de las TIC, se favorece la inclusión y se fomenta la participación activa de los ciudadanos" [12].

Para los sitios web gubernamentales existen algunas políticas a tener en cuenta, como ser la Resolución 69/2011 de la Secretaría de Gabinete, que hace mención a la "Guía de Accesibilidad para Sitios Web del Sector Público Nacional" [13]. Esta resolución apunta a que el acceso a los sitios web pueda garantizarse adoptando estándares internacionales que aumenten la posibilidad de que la información de una página web estatal pueda ser comprendida y consultada por personas con discapacidad y por personas que posean diversas configuraciones en su equipamiento o en sus programas. Además, en el ámbito internacional, se condice a la utilización de las recomendaciones del W3C [14] sobre WAI (Web Accesibility Initiative) que constituyen la referencia en cuanto a criterios y estrategias de accesibilidad a Internet. Estas recomendaciones y documentos técnicos del Consorcio W3C, posee una consolidación con las pautas tratadas por la Resolución 69/2011.

El artículo 1º de la Ley de Accesibilidad de la información en las Páginas Web establece que "el Estado nacional, entiéndase los tres poderes que lo constituyen, sus organismos descentralizados o autárquicos, los entes públicos no estatales, las empresas del Estado y las empresas privadas concesionarias de servicios públicos, empresas prestadoras o contratistas de bienes y servicios, deberán respetar en los diseños de sus páginas web las normas y requisitos sobre accesibilidad de la información que faciliten el acceso a sus contenidos, a todas las personas con discapacidad con el objeto de garantizarles la igualdad real de oportunidades y trato, evitando así todo tipo de discriminación" [15].

En Argentina, se consideran accesibles los sitios web gubernamentales que cumplan con las prioridades 1 y 2 de las WCAG 1.0 (Pautas de Accesibilidad al Contenido en la Web 1.0) [16]. En base a esto, se han elaborado una serie de documentos técnicos donde se describen y relacionan entre sí las técnicas apropiadas para cumplir con cada una de las pautas referidas y sus puntos de verificación. Según la Resolución 69/2011[13], se presentan 14 pautas técnicas a tener en cuenta para el diseño de un sitio web. Estas pautas se agrupan en:

- Técnicas para las Pautas de Accesibilidad al Contenido en la Web 1.0, que expone los temas de accesibilidad y otras técnicas generales como por ejemplo, validación, test, etc. Además, se analizan los temas referidos a estructuras y presentación de sitios, textos equivalentes, páginas alternativas, acceso a teclados, navegación, etc.
- Técnicas HTML para las Pautas de Accesibilidad al Contenido en la Web 1.0. Acá se analizan las estructuras de documentos y metadatos, listas, tablas, imágenes, enlaces, video y audio, frames, formularios, scripts, etc.
- Técnicas CSS para las Pautas de Accesibilidad al Contenido en la Web 1.0, como ser: crear hojas de estilo, fuentes, estilo del texto, formato, colores, reglas, bordes, etc.

A su vez, en las WCAG 1.0, se describe un método para la validación de documentos web [17].

Esta resolución es una aproximación al análisis de los sitios web gubernamentales, pero no incorpora características propias para poder asegurar un buen diseño adaptativo. Sin embargo, cada vez son más los sitios web que se construyen con diseño adaptativo. Un relevamiento realizado en el 2016 sobre los municipios de la provincia de Buenos Aires indica que el 30% de ellos tienen diseño adaptativo [11]. Es por ello que este artículo propone analizar sitios web adaptativos por medio de una herramienta automática (lo cual es presentado a continuación en la sección 4.1) y además analizar en ellos las pautas elaboradas previamente por el equipo de investigación para dichos sitios (ver figura 2) y evidenciar por medio de casos de estudio la situación actual de los mismos (lo cual se presenta en la sección 4.2).

4. Casos de estudio

4.1. Análisis mediante una herramienta automática

A continuación, se muestran algunos ejemplos de sitios web con y sin diseño adaptativo. En la Figura 3, se presenta el sitio web del Ministerio de Economía de la Provincia de Buenos Aires (en la Parte A: la pantalla del navegador web desde una PC y en la Parte B: el mismo sitio web desde un dispositivo móvil). Se puede observar que no se toma en cuenta una adaptación del diseño de dicha página web.



Fig. 3. Sitio del Ministerio de Economía de la Provincia de Buenos Aires. Parte A: Sitio accedido desde un navegador de PC. Parte B: Sitio accedido desde un dispositivo móvil.

En la Figura 4, se muestra el sitio web del Ministerio de Modernización, (en la parte A, la pantalla del navegador web desde una PC y en la parte B: el mismo sitio web desde un dispositivo móvil). Se puede observar que, en este caso, se toma en cuenta una adaptación del diseño de dicha página web.



Fig. 4. Sitio del Ministerio de Modernización. Parte A: Sitio accedido desde un navegador de PC. Parte B: Sitio accedido desde un dispositivo móvil.

Google posee una herramienta en forma gratuita y on-line, para realizar pruebas de optimización para páginas web móviles [18], que permite probar de forma rápida y sencilla si una página de un sitio web está optimizada para móviles o no. Mediante esta herramienta, a continuación, se muestra el resultado para el análisis del sitio del Ministerio de Economía de la provincia de Buenos Aires, el cual arroja un resultado indicando que la página no está optimizada para móviles, esto se muestra en la Figura 5. Esta herramienta da algunas recomendaciones a tomar en cuenta para mejorar el sitio: (1) El texto es demasiado pequeño para leerlo. (2) El contenido es más ancho que la pantalla. (3) Los elementos en los que se puede hacer clic están demasiado cerca uno de otros. (4) La ventana gráfica no se ha definido.

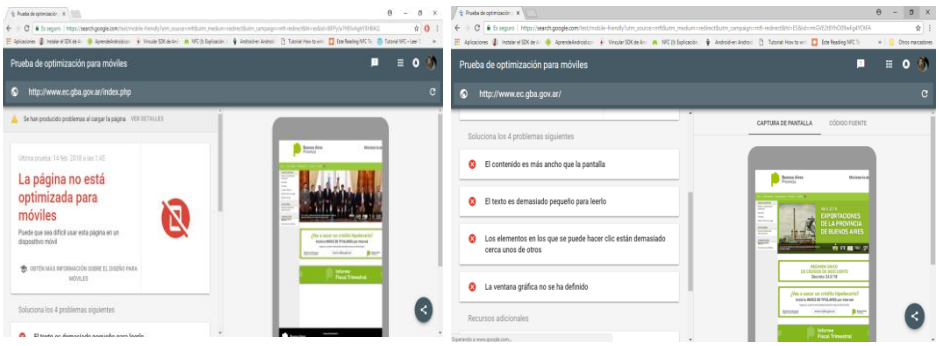


Fig. 5. Utilizando la herramienta de prueba de optimización para móviles para el Sitio del Ministerio de Economía de la Provincia de Buenos Aires.

Por otra parte, en la Figura 6, se muestra el resultado de la herramienta de prueba de optimización de Google, para el sitio web del Ministerio de Modernización.

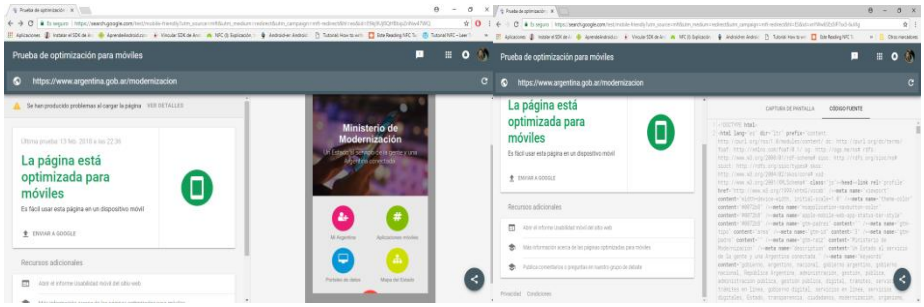


Fig. 6. Utilizando la herramienta de prueba de optimización para móviles para el Sitio del Ministerio de Modernización.

En estos dos casos la herramienta resultó ser una efectiva prueba para evidenciar problemas de diseño de forma rápida, sin embargo, en la mayoría de los sitios web testeados hemos podido notar que en diversas pruebas la herramienta da error de carga del sitio o bien el mensaje que no sea ha podido acceder a la página.

4.2. Análisis por medio de pautas para diseño adaptativo

A continuación, se analizan los sitios web gubernamentales relacionados al Poder Legislativo: (1) Congreso de la Nación; (2) Senado y (3) Cámara de Diputados. Estos sitios se muestran en la Figura 7; en la parte A, se puede observar que, al achicarse el tamaño de las ventanas, se enciman los objetos de la página (de hecho, en vez de achicar su tamaño se agrandan cuando hay menos espacio). En la Figura 7 parte B, con un tamaño menor de pantalla, se puede observar fácilmente superposición de contenidos. A continuación, se muestra la Tabla 2, el análisis de cada una de las pautas presentadas para el diseño adaptativo con el Enfoque Mobile First, mediante la opción de Desarrollador Web del navegador Firefox.

PARTE A **PARTE B**



Fig. 7. Se muestran los 3 sitios del Poder Ejecutivo, siendo: Congreso de la Nación, Senado y Cámara de Diputados correspondientemente.

Cabe destacar que, para estos sitios no funcionó la herramienta automática presentada en la sección previa de forma correcta, existiendo algunos recursos que no pudieron ser cargados en su totalidad. Por otra parte, estos sitios no cumplen con la mayor parte de las pautas consignadas (ver tabla 2). Llama la atención que usando Boostrap no hayan aprovechado los puntos de quiebre para ofrecer distintas vistas al usuario y tengan tan sólo 2 vistas: la completa y otra en tamaño pequeño.

Tabla 2: Análisis de las 10 Pautas para el Diseño Adaptativo con el Enfoque Mobile First.

Pautas para el Diseño Adaptativo	Congreso de la Nación	Senado de la Nación	Cámara de Diputados
P1- Prioridades de Contenido	NO	NO	NO
P2 – Limitar Contenido	NO	NO	NO
P3 – Control de Scroll	SI	SI	SI
P4 – Páginas Livianas	NO	NO	NO
P5 – Barra de Navegación	SI	SI	SI
P6 – Gráficos Vectoriales	NO	NO	NO
P7 – Evitar sobre carga de imágenes	SI	NO	SI
P8 – Diseñar distintas vistas	NO	NO	NO
P9 – Usar unidades relativas	NO	NO	NO
P10 – Hojas de estilo	NO	NO	NO

5. Conclusiones

Este trabajo plantea la importancia del diseño adaptativo en los sitios web actuales, para permitir que los usuarios puedan tener vistas distintas dependiendo de sus dispositivos de acceso. Esto cobra una importancia mayor aún, cuando se trata de un sitio web gubernamental ya que representa la única opción como medio de comunicación en la web para poder ofrecer contenidos y servicios a los ciudadanos. Se tomaron como casos de estudio sitios web gubernamentales, pudiéndose observar que aquellos que cuentan con diseño adaptativo presentan dificultades para ser visualizados, estas dificultades pueden advertirse al analizarse mediante un conjunto de pautas presentadas previamente por el grupo de investigación, esto expresa que aún falta contemplar en el marco normativo nacional, cuestiones técnicas que puedan mejorar la accesibilidad y usabilidad de las soluciones presentadas a los usuarios.

Como trabajo futuro se analizarán otros sitios web gubernamentales con diseño adaptativo para poder extraer estadísticas generales de los errores más comunes.

Referencias

[1] Martín, E., & Marchesi, A. (2006). La integración de las tecnologías de la información y la comunicación en los sistemas educativos. *Argentina, IIPE–UNESCO Sede Regional Buenos Aires. Ministerio de Educación Ciencia y Tecnología/PROMSE*, 12-19.

[2] Naser, A., & Concha, G. (2011). El gobierno electrónico en la gestión pública. CEPAL. *Santiago de Chile. Instituto Latinoamericano y del Caribe de Planificación Económica y Social (ILPES)*.
[3] Ramírez-Alujas, Á. (2011). Open Government y Gobernanza Colaborativa: El (inevitable) camino hacia un nuevo paradigma de Gobierno y Gestión Pública. Estado del arte, desafíos y propuestas. In *Ponencia presentada en el X Congreso de AECPA:"La política en la red", Murcia, del* (Vol. 7).
[4] Lorenzo, S., & César, C. (2010). *Open government: gobierno abierto*. Jaén, España: Algón Editrores MMX, 2010.
[5] Ramírez-Alujas, Á. (2012). Gobierno Abierto y Modernización de la Gestión Pública. Tendencias actuales y el (Inevitable) Camino que Viene-Reflexiones Seminales (Open Government and Modernization of Public Management: Current Trends and the (Inevitable) Way Forward-Seminal Reflections).
[6] Meléndez, C. C., & Vázquez, A. Z. (2017). Municipios y gobierno abierto, más allá del gobierno electrónico. *Revista Opera*, (21), 55-77.
[7] González-Bañales, D. L., & Monárrez Armendáriz, C. (2015). Aplicación de principios de diseño adaptativo para el acceso a la plataforma Moodle en dispositivos móviles.
[8] Marcotte, E. (2017). *Responsive web design: A book apart n° 4*. Editions Eyrolles.
[9] Rodríguez, R. A., Marko, I., Vera, P. M., Valles, G. Y., & Acevedo Zain, G. (2017, August). Evaluación de la usabilidad en sitios web móviles adaptativos. In *XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires)*.
[10] W3C. Mobile Web Best Practices 1.0. 2008. Fuente: https://www.w3.org/TR/mobile-bp/
[11] Vera, Pablo M.; Rodríguez, Rocío A.; Martínez, M. Roxana (2016). Guía metodológica para el desarrollo de sitios web móviles adaptativos con el enfoque Mobile First. *Congreso Nacional de Ingeniería en Informática / Sistemas de información (4°: 2016 nov. 17-18: Salta)*. (2016). [10 p.]
[12] Martínez, M. R., Rodríguez, R. A., & Vera, P. M. (2017). Incrementando la participación ciudadana mediante herramientas de crowdsourcing. In *I Workshop de Ciencia Abierta y Ciudadana-Argentina (La Plata, 2017)*.
[13] InfoLeg. Resolución 69/2011. SECRETARIA DE GABINETE. Apruébase la "Guía de Accesibilidad para Sitios Web del Sector Público Nacional".
[14] W3C. ALL STANDARDS AND DRAFTS. Fuente: http://www.w3.org/TR/
[15] Gobierno Argentino - Educar. Accesibilidad Web.
https://www.educ.ar/institucionales/10/accesibilidad-web
[16] Accesibilidad Web - WCAG 1.0 - Pautas y puntos de verificación de WCAG 1.0 - Universidad de Alicante. Fuente: http://accesibilidadweb.dlsi.ua.es/?menu=pautas-puntos-1.0
[17] Accesibilidad Web - WCAG 1.0 - Método - Universidad de Alicante.
http://accesibilidadweb.dlsi.ua.es/?menu=metodo-1.0
[18] Prueba de optimización para móviles. Google, Ayuda de Search Console.
https://support.google.com/webmasters/answer/6352293?hl=es

Comparison of services for the recognition of flora images. Uses in augmented reality and tourism.

Guillermo Feierherd[1], Federico González[1], Leonel Viera[1], Rosina Soler[2], Lucas Romano[1], Lisandro Delía[3], Beatriz Depetris[1]
[1] Instituto de Desarrollo Económico e Innovación, Universidad Nacional de Tierra del Fuego Antártida e Islas del Atlántico Sur
Hipólito Irigoyen 880, 9410, Ushuaia, Tierra del Fuego, Argentina
{gfeierherd, fgonzalez, lviera, lromano, bdepetris}@untdf.edu.ar
[2] Centro Austral de Investigaciones Científicas, CONICET
Houssay 200, 9410, Ushuaia, Tierra del Fuego, Argentina
rosinas@cadic-conicet.gob.ar
[3] III-LIDI - Universidad Nacional de La Plata
50 y 120, 1900, La Plata, Buenos Aires, Argentina
ldelia@lidi.info.unlp.edu.ar

Abstract. Tourism information services are evolving rapidly. With Internet, tourists organize their trips by managing information before arriving at their destination. Nature is the main tourist attraction in Argentina. However, the information tools as field guides, have had few improvements in their digital version compared to printed ones. This work compares machine learning, deep learning, artificial intelligence and image recognition services, to evaluate the app development for mobile phones that offers the recognition in real time of flora species in natural areas with low or no internet connectivity. Recognition of three *Nothofagus* tree species were evaluated in the Tierra del Fuego National Park, using IBM Watson and Microsoft Azure, with good results in general. A next iteration of this work expects to use assisted learning to improve the efficiency of the neural network obtained to know the adaptation capacities for each evaluated service.

Keywords: tourism, smart tourist destinations, smart cities, foliar morphology, *Nothofagus*, augmented reality, artificial intelligence, image recognition, computer vision, IBM Watson, Microsoft Azure, Google Cloud, Amazon Web Services.

1 Introduction

This work is part of the project "Virtual and Augmented Reality, Big Data and Mobile Devices: Applications in Tourism" that since 2017 has been developed at the National University of Tierra del Fuego. The project seeks to reveal the uses that the tourism industry is making of these emerging technologies individually or in combination, to propose alternatives for application in the area of Tierra del Fuego.

Considering that the main tourist attraction of Tierra del Fuego is wildlife (flora and fauna), tourists often enrich their trip with field guides, species guides or

information brochures about the local flora and fauna. However, it would be much more interesting if they could recognize a species at the same moment of taking a photograph and in this context, the potential of augmented reality and the recognition of images associated with artificial intelligence, is promising.

Biological organisms are not always easy to photograph. Animals, whether they are walkers, fliers or swimmers, besides being mobile (and some of them very fast) are elusive to human presence or have cryptic behavior. Photographing plants, on the other hand, is much easier, accessible and attractive for anyone without technical knowledge of photography, even when they have only a smartphone. On the other hand, throughout the world, the proportion of identified plant species is greater than any other biological group [1].

Despite this greater knowledge about plant species, in some cases the interspecific morphological differences are minimal and difficult to recognize with the naked eye. Even in herbaria that contain a large number of specimens, their identification is being carried out with artificial vision and machine learning approaches applied to scans of leaves or images of plants in the field [2]. The trees of the genus *Nothofagus* are the most representative of the forest of Tierra del Fuego, but the specific differences of their leaves are difficult to recognize when comparing them without expert human eye, which represents an ideal difficulty to test the augmented reality together with the recognition of images.

Considering the complexity and constant evolution of the technologies related to machine learning, deep learning, artificial intelligence and image recognition, we decided to use the services provided in the cloud to train the neural network from a database of images to identify tree species typical of the forest of Tierra del Fuego. The ultimate goal is to use that knowledge and to develop an offline app in the future to recognize flora and fauna without connectivity, which is a peculiarity of the protected areas.

2 Information in smart tourist destinations

Tourist information systems have traditionally been organized in three chronological stages: promotion, planning and stay. In this way, the person who deals with the management of a tourist destination, first makes marketing, then provides information tools to plan the itinerary and, finally, accompanies the tourists with useful information to help them know the destination during their stay. In recent years, the concept of stay was reoriented to that of experience and a fourth component was added to the stages: sharing what was lived.

This work focuses on the stage of the experience, particularly when it takes place in Protected Areas. In those cases, the traditional methods to provide information were based on printed brochures (which could be obtained at a visitor center), on strategically located posters, signage of the trails, etc.

This scheme, with its advantages and disadvantages, but the only one possible until a few years ago, must be complemented and enriched with the resources that technology offers today. This also allows adapting them to the characteristics of

younger visitors, digital natives, accustomed to using new technological resources to obtain the information they need.

Doug Lansky [3] questions the role of visitor centers. Their questions take into account the investment that many of them represent and the increasingly scarce utility of their services. In his opinion its strategic location makes many tourists visit them with the idea of collecting some free maps and brochures, or maybe get an air-conditioned space, shelter from rain or cold, or the need for toilets. But, if they were not there, that does not mean tourists could not get the same information services using their phones and other mobile devices.

Obviously there will be some people who do not carry a smartphone with them, but they are becoming a limited minority. And it is likely that the "analogue" visitors will already arrive prepared with a printed guide, newspaper clippings or the advice they have requested from the hotel receptionist.

It is known that smartphones have been replacing several useful old devices for tourists, such as photo cameras, video recorders, GPS among others. It is also true that many visitors consider the Protected Areas as a symbol of nature and feel that it is worthwhile to remain "disconnected" while they go through them.

People over 50 only use the telephone (Fig. 1) to perpetuate moments and understand the place - take pictures, look at a map - while younger generations incorporate communication tools to interact with social networks and for other recreational uses such as listening music while they visiting the protected area [4].

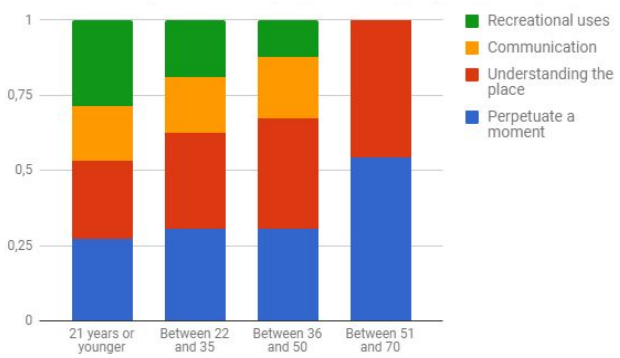


Fig. 1. Uses of the smartphones according to generations grouped by activity [4].

The biggest difference between young and old generations is linked to recreational use and communication (with the outside world of the protected area). This mean that older generations are prone to disconnect themselves within protected areas, while digital natives prefer to remain connected and give intensive use to all the analyzed functions of the telephone when visiting protected areas.

From the above, it is clear that interactive identification of species using a smartphone can be a motivation against traditional brochures or guides with descriptions and figures.

3 Methodology

3.1 Morphological aspects of the compared species

For this work 3 tree species of the genus *Nothofagu*s were selected, which predominate in the Fuegian forest: lenga (*N. pumilio*), ñire (*N. antarctica*) and guindo (*N. betuloides*). The first two are deciduous and phylogenetically closest species [5], so hybridization between them is possible [6]. While the third species is an evergreen tree with morphological characteristics that differentiate it from its congeners. The main foliar characteristics of these trees (Table 1) allow to recognize with an ordinal eye, with a minimum of botanical knowledge, what kind of species they are [7] [8].

Table 1. Main foliar characteristics of the *Nothofagus* species considered in the study

Foliar characteristic	Lenga	Ñire	Guindo
Habit	deciduous	deciduous	evergreen
Shape	oblong-elliptical	oblong	oval
Length	2-3 cm	2-3 cm	1-2.5 cm
Width	1.5 cm	1.5 cm	1 cm
Apex	blunt	rounded	sharp
Edge	crenate	irregular lobed	regular serrated
Base	slightly asymmetric	asymmetric	symmetric
Petiole length	0.5 cm	0.5 cm	0.3 cm

3.2 Visual recognition services

Neural networks have proven effective at solving difficult problems but designing their architectures can be challenging, even for image classification problems alone. Different provider services aim to minimize human participation, employing evolutionary algorithms to discover such networks automatically [9]. Despite significant computational requirements, it is now possible to develop models with high accuracies.

Different complex artificial intelligence and machine learning services are being offered in the cloud by companies such as IBM, Microsoft, Google or Amazon. With different levels of complexity regarding its implementation, this paper compares the services of IBM Watson (Visual Recognition) and Microsoft Azure (Custom Vision), since they are the only ones that incorporate integral products based on visual platforms that facilitate its implementation in great measure. Google and Amazon

recently announced their own visual platforms, but they are still in alpha phase and their access is restricted to a limited number of users.

3.3 Collection of photographs

For each species of *Nothofagus* trees, 45 photographs of leaves or branches with different level of clearly differentiable leaves were provided, to be used as machine learning models. The images were taken at different seasons of the year (except winter) by people without specific knowledge of photography, but with knowledge of plant species.

In addition to the datasets created and categorized, a stock of 30 images (10 for each species) that did not belong to the database and were used to compare the certainty of the results offered by Watson and Azure was established.

The quality of this 30 photographs (Table 2) was variable to represent different situations into the database, all of them in jpg format.

Table 2. Main characteristics of used photographs for each species for this study. Nb: guindo (*Nothofagus betuloides)*, Np: lenga (*N. pumilio)*, Na: ñire (*N. antarctica)*

Photo	Leaves/Branches	Noise	Human recognition	Photo quality	Colour
Nb1	35/10		easy	good	green
Nb2	70/10		easy	good	green
Nb3	50/5	fruits	easy	medium	green
Nb4	45/2		easy	good	green
Nb5	30/10	gall	easy	good	green
Nb6	60/5		easy	good	green
Nb7	80/10		easy	good	green
Nb8	50/15	fence	easy	good	green
Nb9	55/10		medium	good	green
Nb10	20/5	stems	easy	medium	green
Np1	85/2		easy	good	green
Np2	50/10		easy	good	yellow
Np3	20/5		easy	bad	green
Np4	60/15		easy	good	green
Np5	80/2		medium	medium	red
Np6	40/5		tricky	good	green
Np7	10/2		tricky	good	green
Np8	75/5		easy	good	green
Np9	60/2		easy	good	green
Np10	55/20		medium	good	green
Na1	60/10		tricky	bad	yellow
Na2	35/10		easy	good	green/yellow
Na3	80/0		easy	medium	green
Na4	75/2		easy	bad	green
Na5	55/5		easy	good	green
Na6	50/20	fruits	easy	good	green
Na7	75/2		easy	good	green

Na8	45/15		easy	good	green
Na9	20/15	flowers	easy	good	green
Na10	15/20	hand	easy	good	green

4 Results

The recognition results are shown in Table 3. The case with greater precision was guindo, since Watson (9 of 10 tests) and Azure (8 of 10 tests) showed ≥ 80% accuracy. Similar results were obtained for ñire, since Watson recognized 8 of 10 tests while Azure recognized 7 of 10 tests. However, the probability of recognition was low for this species (≥65% accuracy for both services) and cases with greater confusion were observed. For example, the first and second probability results for Na2, 6, 7, 8 and 10 photos were similar with less than 10% difference. Finally, both services were less successful to recognize lenga since only 4 of 10 tests were right with ≥75% accuracy. Also, the first and second probability results for Np2, 4 and 9 photos were similar with less than 10% difference.

Table 3. Comparison of recognition probability of tree species between Watson and Azure services. Match cases are highlighted in green and failed cases in red. Cases with the first and second probabilities ≤10% of difference are shown. Nb: guindo, Np: lenga, Na: ñire

	Watson			**Azure**		
	Nb	Np	Na	Nb	Np	Na
Nb1			0.86	0.99		
Nb2	0.90			0.93		
Nb3	0.79			0.94		
Nb4	0.83			0.87		
Nb5	0.90			0.98		
Nb6	0.80					0.97
Nb7	0.87			0.96		
Nb8	0.91					0.76
Nb9	0.90			0.99		
Nb10	0.90			0.92		
Np1		0.91			0.99	
Np2	0.57	0.46		0.94		
Np3		0.79				0.55
Np4	0.59		0.67	0.94		
Np5			0.85		0.99	
Np6			0.91			0.98
Np7		0.75			0.98	
Np8		0.78			0.89	
Np9	0.77			0.69		0.75

Np10		0.91			0.83
Na1		0.88			0.81
Na2	0.68	0.57			0.97
Na3		0.90			0.98
Na4		0.89			0.84
Na5		0.85			0.64
Na6	0.60	0.67			0.90
Na7	0.68	0.60			0.98
Na8		0.59	0.73		0.72
Na9		0.89	0.54		
Na10		0.91	0.42	0.40	

5 Conclusions

Watson and Azure services had similar performance, but Watson was more successful in recognizing two of the three analysed tree species. Guindo (Nb) and ñire (Na) seem to have more defined morphological characteristics, while lenga (Np) had intermediate and similar morphological characteristics to other species and this generated high confusion for the systems. However, a future guided training for the recognition system should be improved. On the other hand, the accuracy of the recognition was not related to the quality of the photographs, nor the proportion of leaves they contained.

This work focused on the interpretation of images in the foreground (leaves, branches) since for tourist purposes it is of great value for the visitors to be able to know what they have at their fingertips. Beyond this objective, it would also be interesting and possible in future studies to broaden the point of view towards the interpretation of specimens at medium distances to answer questions like "what kind of trees are those?" or to count "how many individuals of each species are there?" in a determined group.

In future iterations it is expected to include Amazon Web Services and Google Cloud, as they release their visual tools for image classification.

Furthermore, it is necessary to incorporate new species in the training process in the near future to generate a database with knowledge about local flora of high interest for tourists. The three species studied in this work are the most representative trees of the Tierra del Fuego National Park, and also the best known by people. The recognition of other plants as forbs, grasses, or ferns is a major challenge from a technical point of view, since a greater variety of species very similar among them could potentially generate more confusion in the recognition or less accuracy of results.

Regarding this last point, it is important to refine the recognition capabilities of the machine learning for the current database and for new species, manually correcting the errors of the first iteration and testing a new group of photographs.

Finally, and as final outcome of our project "Virtual and Augmented Reality, Big Data and Mobile Devices: Applications in Tourism" we expect to develop a public offline app that incorporates image recognition and offers results from the compared platforms, while generating at the same time new databases to refine knowledge with assisted machine learning. Offline mode is a key feature for the usefulness of apps in countries such as Argentina, where the main destination of international tourism are the National Parks (Perito Moreno Glacier, Iguazú Falls). Although Argentina is one of the countries with higher access to the internet in Latin America, such connectivity occurs in cities or urban areas where population is concentrated [10]. The protected areas do not have any type of internet connection and in most cases, no access to mobile phone network [11]. To achieve this, the application must be able to identify the species offline without external services. With the previous training of a neural network based on machine learning services, we hope to have enough potential to achieve it.

References

1. Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., Worm, B.: How many species are there on earth and in the ocean? PLoS Biol. 9(8): e1001127 (2011)
2. Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., Joly, A.: Going deeper in the automated identification of herbarium specimens. BMC Evolutionary Biology 17:181 (2017)
3. Lansky, D.: Has The Smartphone Killed The Tourist Office? Digital Tourism Think Tank, (2016) http://thinkdigital.travel/opinion/has-the-smartphone-killed-the-tourist-office/
4. González F.: Digital tourist information tools in Protected Areas for Millennials and Gen Z. Master Thesis, University of Girona, Spain (2017)
5. Jordan, G., Hill, R.S.: The phylogenetic affinities of *Nothofagus* (Nothogaceae) leaf fossils based on combined molecular and morphological data. Int. J Plant Sci. 160: 1177-1188 (1999)
6. Burns, S.L., Cellini, J.M., Lencinas, M.V., Martínez Pastur, G.J., Rivera, S.M.: Descripción de posibles híbridos naturales entre *Nothofagus pumilio* y *N. antarctica* en Patagonia Sur (Argentina). Bosque: 31: 9-16 (2010)
7. Moore, D.M.: Flora of Tierra del Fuego. Anthony Nelson, England (1983)
8. Donoso, C..: Las especies arbóreas de los bosques templados de Chile y Argentina. Autoecología. Marisa Cuneo Ediciones, Chile (2006)
9. Real, E., Moore, S., Selle, A., Saxena, S., Suematsu, Y., Tan J., Le, Q., Kurakin, A.: Large-Scale Evolution of Image Classifiers. Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, PMLR 70, (2017)
10. Open Signal: Global cell coverage maps, Argentina (2018) https://opensignal.com/networks?z=4&minLat=-56.9&maxLat=-16.6&minLng=-109.7&maxLng=-18.3&s=&t=4
11. APN: Mapa oficial con polígonos de las Áreas Protegidas Nacionales de Argentina. Administración de Parques Nacionales (2018) http://mapas.parquesnacionales.gob.ar/layers/geonode%3Aapn_areasprotegidas_01

Value enhancement of the Artistic Heritage of Tierra del Fuego using Augmented Reality

Guillermo Feierherd[1], Leonel Viera[1], Federico González[1], Francisco Huertas[1], Lucas Romano[1], Lisandro Delía[2], Beatriz Depetris[1]

[1] Instituto de Desarrollo Económico e Innovación, Universidad Nacional de Tierra del Fuego, Antártida e Islas del Atlántico Sur
Hipólito Irigoyen 880, 9410, Ushuaia, Tierra del Fuego, Argentina
{gfeierherd, lviera, fgonzalez, fhuertas, lromano, bdepetris}@untdf.edu.ar
[2] III - LIDI - Facultad de Informática - Universidad Nacional de La Plata
ldelia@lidi.info.unlp.edu.ar

Abstract. The province of Tierra del Fuego developed tourism as one of its main industries since its beginnings, managing to establish well-established tourist circuits. With the increase of tourists received every year, the typical circuits are saturated. Through the use of new technologies, new tourist circuits can be promoted, contributing to the development of new sectors of the city and reducing the saturation of typical places. This work presents a mobile application with augmented reality capabilities that encourages the visit of the Artistic Heritage of Tierra del Fuego, trying to establish new tourist circuits. In turn, it compares the different techniques for the recognition of points of interest, applied on Artistic Heritage of Tierra del Fuego.

Keywords: tourism, smart tourist destinations, smart cities, augmented reality, artistic heritage, mobile apps, software engineering.

1 Introduction

This work is part of the project "Virtual and Augmented Reality, Big Data and Mobile Devices: Applications in Tourism" that since 2017 has been developed at the National University of Tierra del Fuego. The project seeks to reveal the uses that the tourism industry is making of these technologies individually or in combination, in order to propose alternatives for application in the field of Tierra del Fuego (TDF).

During the 2016-2017 season, Ushuaia – the main tourist city of TDF – received one third of its visitors through tourist cruises. Due to the way these boats operate, their passengers have a few hours to enjoy the city. As a result, a large number of people rush to make quick trips to the Tierra del Fuego National Park, a City Tour or simply buy and eat in the center of the city, generating saturation in those spaces.

Through the use of new technologies that complement the real world with the virtual one, new tourist circuits can be promoted. When generating other circuits, it contributes to the development of new sectors of the city, decreasing the saturation of typical places.

Combining the work done by the project "Survey and value enhancement of Artistic Heritage (busts, compositions, monoliths, plates, statues and masts) in public spaces of the city of Ushuaia, province of Tierra del Fuego, Antarctica and South Atlantic Islands, Argentine Republic", it is possible to generate an application that encourages tourists to know the Artistic Heritage. Complementing the application with augmented reality (AR) it will offer a more attractive and immersive experience [1], while promoting history and culture.

However, one of the challenges of augmented reality is to link and align reality with virtual information. There are many approaches: using physical markers, detecting geographical position, detecting the user's environment, among others. The following questions then arise: What approaches are currently feasible? What is the appropriate approach for this type of application? The present work seeks to explore the alternatives for the recognition of Artistic Heritage through the use of different techniques and tools, exposing the advantages and disadvantages of each one.

2 Augmented reality

Augmented reality employs computer vision, image processing and graphic techniques to fuse digital content in the real world. It also allows real-time interaction between the user, real objects and virtual objects.

To achieve this task, the user explores the environment through a screen that shows the real world, together with digital information superimposed. At the same time, a software makes constant monitoring to determine, in real time, the location and orientation of the camera and to react to the user's environment depending on the purpose of the application [2]. Specifically, for this work, it is interesting to be able to determine if the user is observing any of the locations where the Artistic Heritage is located.

There are different solutions to face this challenge. In the field of mobile devices, the most commonly used solutions can fall into two categories: AR based on markers and AR without markers. The border between both categories is diffuse, since the AR term itself is constantly evolving. Next, the chosen interpretation for each category will be defined, to then correctly frame the chosen technologies.

2.1 Augmented Reality based on markers

This category groups RA experiences that require the presence of a physical object to recognize the space over which virtual objects overlap.

A typical solution found in this category is to add a predefined pattern, easily detectable in the environment, and use image recognition techniques to detect it. This pattern is known as a marker.

Most markers are formed by black squares on a white background, forming two-dimensional bar codes. It is also possible to use colors or elaborate images, as

long as their characteristics can recognize them quickly. An example of a widely used marker is the QR code.

Beacons are also considered within this category, although in clear decline. These are physical devices that emit a signal (Bluetooth, NFC, WiFi), which can be perceived by a smartphone. Through the use of an app that monitors and interprets the reception of the indicated signal, the device can react proactively and provide the user with information pertinent to the place where it is located.

AR based on markers has as its main advantage its simple implementation and the low use of device resources. Once the marker is recognized, it is easy to superimpose virtual objects in the desired position.

In contrast, the main disadvantage is the need to intervene in the real physical space to include markers. The materials and form in which it intervenes must be analyzed in each case in order to take the best advantage of the environment, considering the expected time of life and the conditions they must endure. Climatic erosion? Can people or animals damage them? How long should they last? Should they be easily modified?

2.2 Augmented Reality without markers

As its name implies, this category refers to the AR experiences that do not require physical markers to identify the real space over which the virtual object should be represented. There are several methods to achieve this task, being the most used the geographical positioning of the points of interest and the recognition of the environment through computer vision [3].

Typically, these methods use various sensors to determine the position and alignment with which virtual objects should be represented over the real world. The most used sensors are: GPS, accelerometer and gyroscope; The magnetometer, the light sensor and the proximity sensor are also used [4].

In parallel, points of interest (POI) must be identified. Under ideal conditions, they must be observable from different points of view and under different lighting conditions. This is called repeatability [5].

Of course, this continuous processing of information added to the constant collection of data from the sensors penalizes the efficiency and resources used [6]. Added to this, the monitoring of environmental conditions presents several challenges: dealing with large scenarios, variable lighting conditions, materials with low texture, reflective and transparent properties.

The introduction of ARCore[1] y ARKit[2] (official AR SDK of Android and iOS platforms respectively), contributed to improve the performance of AR experiences on mobile devices. They facilitate the development of applications with six levels of freedom; term used to describe the experiences of AR that respond to the rotation and

[1] https://developers.google.com/ar/discover/

[2] https://developer.apple.com/arkit/

position of the user as he travels. Additionally, they provide facility for the detection of different types of surfaces and mechanisms to react to different lighting situations.

However, due to their immature state, few devices on the market support them, limiting the set of users with access to these technologies. Currently both platforms publish compatible devices, being a very small group of devices in both of them.

3 Tools for the identification of points of interest

Understanding the available methods for the implementation of AR and, having as an objective the recognition of POI on the Artistic Heritage of TDF, tests were carried out with different tools.

The team decided to narrow the set of tools to those that address the problem using physical markers, computer vision and geographical position.

3.1 Considerations

When selecting and evaluating the tools, those that were open source and / or that provide some form of free use were prioritized. Additionally, they must allow multiplatform development. In this sense, due to the team's previous experience, solutions aimed at hybrid and interpreted development for mobile devices were sought, specifically Cordova / PhoneGap and React Native.

In spite of the above, it was considered of vital importance to evaluate the new SDK ARCore and ARKit, since these will be the bases for the future frameworks of development in AR. In addition, there is a library for React Native called ViroReact that allows to use both solutions under a common API, allowing to test the potential of the native SDK, without sacrificing the multiplatform objective.

On the other hand, despite its high cost license, it was also decided to include the Wikitude SDK as it is the only tool that allows detecting the points of interest with the three chosen methodologies. For its test the free version that overlaps watermarks was used.

3.2 Based on computer vision

About this method. It seeks to recognize a POI by evaluating frames to see a particular object, based on previous samples of the same - images.

Tools. For this approach, tests were conducted mainly with online services, highlighting IBM Watson Discovery[3] and Microsoft Azure Custom Vision[4].

[3] https://www.ibm.com/watson/services/discovery/

[4] https://www.customvision.ai/

Among the online services, the Azure platform was chosen because it has an interesting addition: the ability to export the learning of the AI to the native formats TensorFlow[5] (Android) and Core ML[6] (iOS), allowing offline recognition. Although this feature has not been tested, it is an important plus since the potential users of the application might not have access to the Internet.

Additionally, both Wikitude and ViroReact – and by transition ARCore and ARKit – allow this type of recognition.

Conclusions. This methodology is seriously affected by real-world variability. The artistic patrimonies are outdoors, severely affected by natural light and weather conditions. To achieve a reasonable result you need to use a large number of sample images, trying to contemplate different times of the year and points of view, affecting the user's storage space. Finally, the distance with which these photographs are taken will influence the minimum and maximum distance that the user needs to maintain with the artistic heritage to see the virtual information.

On the other hand, the use of online services is conditioned by internet access and connection speed. You must select the frames, upload them and wait for the answer. This could be solved by requiring the user to manually activate the recognition, but it impairs the feeling of immersion. In this sense, offline solutions are the best prepared, although they are equally affected by the variability of the environment.

Therefore, it is considered that this approach is not useful for this project, or for anyone who has similar conditions in the environment.

3.3 Based on physical markers

About this method. It uses physical markers with ideal characteristics for its recognition, facilitating its rapid recognition and reducing the use of resources.

Tools. The number of tools that support this type of AR is quite broad, making it difficult to make a selection. Without forgetting to highlight argon.js, the one that gave the best performance and flexibility was AR.js, in its version combined with A-frame. These combined javascript libraries achieve an incredible speed when detecting bookmarks, while also offering the possibility of taking this experience to the web. On the other hand, both ViroReact and Wikitude solve this approach.

Conclusions. Due to the presence of the physical marker, the variability of the natural environment presented by the previous problem is solved. Although climate and luminosity affect these markers, their characteristics make it easier to recognize the desired pattern.

[5] https://www.tensorflow.org/mobile/android_build

[6] https://developer.apple.com/machine-learning/

Of course, this solution requires intervening all the spaces where you want to have the experience and the maintenance of them. The size of the markers directly influences the distance required to recognize the marker.

Regarding the user experience, it can be criticized that it reduces immersion by having to focus on a particular marker, which may not be on the artistic heritage. Also, if there were a large number of people trying to scan the marker, it would generate an unpleasant situation for the tourist and the place.

3.4 Based on geographical position (GPS)

About this method. Knowing the user's position, the direction where he is pointing and the position of the POI, the distance between them can be calculated. If there is a safe distance when focusing, the virtual information is displayed.

Tools. From the survey conducted by the team, the only tool that offers this type of experience explicitly is Wikitude. However, there are libraries that allow conversions between the GPS coordinate system and the ECEF - Cartesian coordinate system, which allows to represent the field of vision observed by the user through his camera. Because of these libraries, this approach could be carried out using ViroReact.

Wikitude provides the best and simplest development experience, while the solution with ViroReact, has some flaws in the refreshment of the marker when the user is in motion.

Conclusions. Although the lack of tools is not very encouraging, this approach seems to be appropriate for situations where the point of interest is outdoors, as in this case. Knowing the distance between the user and the place, it can be manipulated with precision the correct moment to show the virtual object. In addition, the position is constant and is not affected by weather conditions or brightness, including being able to operate at night.

As a disadvantage, it is difficult to show the virtual element in the desired position. The GPS usually have a margin of error in meters, distorting (in the area of vision of the camera) the desired location of the virtual object.

4 Recognition of the Artistic Heritage of TDF

Based on the data provided by the project "Survey and value of the Artistic Heritage (busts, compositions, monoliths, plates, statues and masts) in public spaces of the city of Ushuaia, province of TDF, Antarctica and South Atlantic Islands, República Argentina", a mobile application was designed that combines this information with cutting edge technological aspects, to value the artistic heritage of TDF; seeking to encourage visits to the different POIs and disclose what they represent.

4.1 Mobile application design

Although the focus of this work is to compare the different AR tools to recognize the Artistic Heritage, a complete mobile experience was designed taking into account the previously mentioned objectives.

The mobile application was raised in a simplistic way, inviting users to discover the historical heritage by topic, searching directly on a map or accessing a filterable list of them.

Each point of interest can be marked as favorite, allowing to build a tour based on personal preferences. Each artistic heritage has an extended file where all the information about it can be seen.

Also, users will be able to make comments and rate each place, achieving a better interaction between them, which favors disclosure. In the future, the scores could be analyzed to detect patrimonies that need some kind of intervention, for example a settlement of the place.

Finally, the app offers a last alternative to know the POI, through augmented reality. The user will be able to access a mode where the camera will identify an artistic heritage. The initial objective is to show a virtual indicator that encourages the user to touch it to finally access the heritage file.

4.2 POI detection tools applied to Artistic Heritage

Of the totality of functions enunciated in the previous section, this work focuses on the last one, performing a technical test that allows to evaluate the advantages and disadvantages of the tools mentioned in section 3 for the recognition of the artistic heritage of TDF.

For this purpose, a mobile application was developed as a technical demonstration, focusing identification on three artistic heritage: the ARA Guaraní plaza, the monument to the Fallen in Malvinas and the monolith of the ARA General Belgrano.

For the test, only solutions based on physical markers and geographic positioning were used for the reasons explained in section 3.



Fig. 1. From left to right: the ARA Guaraní plaza, the monument to the Fallen in Malvinas and the monolith of the ARA General Belgrano.

Physical markers. To check the performance of the tools, a 12x12 cm marker was printed representing each Artistic Heritage. Then, the recognition was implemented with each technology, to finally test them in the same device and environment, being able to establish the following comparisons:

Table 1. Comparison of the experience with each tool.

Tool	Max. Distance	Detection in slight movement	Perceived speed
AR.js	130 cm	Yes. Very good	Excellent
ViroReact	80 cm	No	Regular
Wikitude	60 cm	Yes. Good	Very good

As an extra data, ViroReact performs a scan of the environment allowing to keep the object in the position where it was located, avoiding to recognize it again. This was clearly seen when making a 360 turn and visualizing the virtual object in its place as if it had always been there.

GPS. With each tool, a screen was implemented using the user's position and the coordinates of the POI, plus a fourth point a few meters from the ARA Gral. Belgrano monolith, exemplifying the visualization of nearby POIs.

Using Wikitude it was easy to develop this solution, since it has integrated this functionality. On the other hand, as mentioned above, a customized solution was developed using ViroReact and libraries to convert GPS coordinates to Cartesian coordinates and to manage user positioning.



Fig. 2. From left to right: ViroReact, AR.js and Wikitude (Markers); ViroReact and Wikitude (geographical positioning).

Both tools allowed to position the virtual object with relative precision, being affected by the margin of error of the GPS of the device. Virtual objects moved slightly when the user moved but a comfortable experience was achieved. With both tools it was easy to adjust the altitude of the virtual object and the size relative to user's position.

As with physical markers, ViroReact benefits from scanning the environment, allowing to maintain the position of virtual objects when they leave the focus of the camera.

5 Conclusions

Taking into account the strengths and weaknesses of each approach and the tools available, it can be concluded that the method based on geographical position offers the greatest benefits for the detection of the Artistic Heritage of TDF. This is mainly due to the fact that it is not affected by weather or light conditions. However, the high license cost of Wikitude and the limited support of ViroReact devices can be a problem.

Considering that the designed application uses RA as a means of highlighting, allowing various mechanisms to make contact and knowledge of the Artistic Heritage, it is proposed to use a mixed approach, where physical markers can be combined with geographical positioning. A base experience can be offered using AR.js, for its performance and support of various devices, along with physical markers located in the most relevant POIs. In this way, all users will be able to enjoy the AR without being restricted by the device they have, while also probing the interest for the Artistic Heritage in the visualization of it through AR.

Then, in the same application, if it is detected that the device has support for ARCore or ARKit, it will be possible to expand the experience through geographic positioning. Users with this ability will be able to use AR to visualize all the points of interest, not only those that have a physical marker.

The closing of the experience, with the complete development of the application, its implementation and the analysis of the benefit achieved remain pending.

As an addition, with the metadata generated by the application, a system could be designed that learns from the user's behavior, punctuation and opinion, for identification of carelessness of the artistic heritage, launching campaigns to raise awareness about any of them in particular or take advantage of the popularity of some of them to enhance tourist circuits.

References

1. Scholz, J., & Smith, A. N. Augmented reality: Designing immersive experiences that maximize consumer engagement. Business Horizons (2016)
2. Anuroop Katiyar, Karan Kalra and Chetan Garg, Marker Based Augmented Reality, ACSIT. (2015)
3. Behzadan, A. H., Timm, B. W., & Kamat, V. R. General-purpose modular hardware and software framework for mobile outdoor augmented reality applications in engineering. Advanced Engineering Informatics (2008)
3. Fang, W., Zheng, L., Deng, H. and Zhang, H. Real-time motion tracking for mobile augmented/virtual reality using adaptive visual-inertial fusion. Sensors (2017)
5. C.R. Huang, C.S. Chen and P.C. Chung, Contrast context this to graman efficient discriminating local descriptor for object recognition and image matching, Pattern Recog. (2008)
6. Ziegler, E. (2010). Real-time markerless tracking of objects on mobile devices. Bachelor Thesis, University of Koblenz and Landau