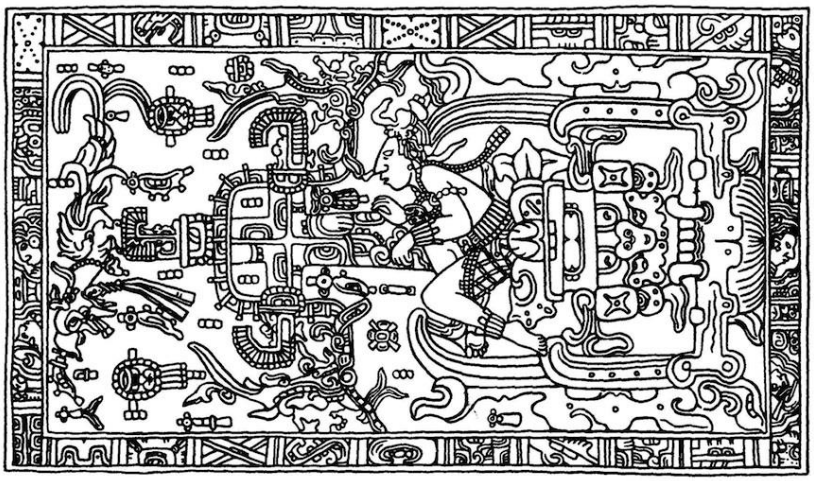
Edgar Serna M. (Ed.)

Avances en Ingeniería



Editorial
Instituto Antioqueño de Investigación

Edgar Serna M. (Ed.)

Avances en Ingeniería

Editorial

Serna, M.E. (Ed.)
Avances en Ingeniería -- 1a ed.
Medellín, Antioquia: Editorial Instituto Antioqueño de Investigación, 2015.
345p. Investigación Científica.
ISBN: 978-958-59127-0-0

Avances en Ingeniería
© Editorial Instituto Antioqueño de Investigación

Edición: octubre 2015
ISBN: 978-958-59127-0-0
Hechos todos los depósitos legales
Publicación electrónica gratuita

Imagen portada: Pakal "el astronauta Maya".
http://alef.mx/pakal-el-astronauta-maya-en-el-centro-del-universo/

Copyright © 2015 Instituto Antioqueño de Investigación (IAI)™. Except where otherwise noted, content in this publication is licensed under the Creative Commons Attribution, NonCommercial, ShareAlike 3.0 Unported License.

Global Publisher: Instituto Antioqueño de Investigación (IAI)
Cover Designer: IAI, Medellín, Antioquia - Colombia.
Editorial Instituto Antioqueño de Investigación is trademarks of *Instituto Antioqueño de Investigación*. All other trademarks are property of their respective owners.

The information, findings, views, and opinions contained in this publication are responsibility of the author and do not necessarily reflect the views of *Instituto Antioqueño de Investigación* (IAI), and does not guarantee the accuracy of any information provided herein.

Diseño, edición y publicación: Editorial Instituto Antioqueño de Investigación (IAI)
Instituto Antioqueño de Investigación
http://fundacioniai.org
contacto(AT)fundacioniai.org

© 2015 Editorial Instituto Antioqueño de Investigación
Medellín, Antioquia – Colombia

Editorial

Prólogo

La sociedad del siglo XXI tiene una relación estrecha con la ingeniería, por eso los temas relacionados con la ciencia, la tecnología y el desarrollo son solicitados diariamente en las publicaciones científicas. En consecuencia, los investigadores en todo el mundo se vuelcan en proyectos cuyos objetivos intentan satisfacer esta demanda. Esta publicación contiene los trabajos de diversos investigadores en ingeniería, a través de una mirada neutral y mediante la participación y aportes en cuestiones clave del desarrollo ingenieril. Es una muestra fehaciente de la colaboración entre pares, con un enfoque en la mejora de los productos de la ingeniería.

Con la idea de entregarle a la Sociedad un conocimiento actualizado y producto de los procesos en investigación en ingeniería, este libro es un sumario de resultados experimentales, teóricos y de reflexión, en el que los autores demuestran su experticia y espíritu de desarrollo para aportar al logro permanente del bienestar social. En las siguientes páginas se resume discusiones, análisis y pruebas, en las que sus autores exploran y superan las barreras al aprendizaje, las lecciones en la transformación de la sociedad y aprovechan el talento técnico de la ingeniería para informar el desarrollo y apoyo necesario a la toma de decisiones, de mecanismos de retroalimentación y a la infraestructura. A lo largo del texto, hacen hincapié en la necesidad de incrementar el trabajo organizado pero mancomunado, para que las nuevas generaciones aprendan y continúen produciendo cada vez mejores resultados para el desarrollo de la ingeniería. Entre las cuestiones más importantes se puede mencionar el aporte en la realineación de los incentivos para obligar a la mejora continua, el fomento de una cultura de liderazgo que refuerza el trabajo en equipo, la mejora de las oportunidades de aprendizaje continuo y la investigación desde diferentes perspectivas, que requieren la perfección en el rendimiento del sistema transdisciplinar, y lo más importante, centrado en procesos de mayor consideración y en la experiencia humana.

Nos gustaría ofrecer nuestros agradecimientos a todos los autores que aportaron su trabajo en estas cuestiones tan importantes para el desarrollo social, por su habilidad y dedicación en la escritura de tan importantes trabajos, y sobre todo por haber confiado en nosotros para la edición de su obra.

Tabla de contenido

	Pag.
Validación de modelos numéricos durante el desarrollo de correlaciones para intercambiadores de calor **Juan Ardila M., Diego Hincapié Z. y Julio Casas M.**	6
Design, implementation and testing of a characterization photovoltaic panels system **Álvaro Jaramillo, Bonie Restrepo, Jorge Arango y Alfonso Berrio**	15
Detección de la fatiga muscular a través de un sensor wearable de bajo costo y análisis de amplitud y frecuencia de la señal EMG superficial **María Montoya V., John Muñoz C. y Oscar Henao G.**	22
Desarrollo de software integrando técnicas de usabilidad y metodologías agiles **Alcides Alvear S. y Graciela Quintero R.**	30
Un método basado en algoritmos genéticos de apoyo al diagnóstico TEA **Juan Cockbaine O.**	44
Reutilización de la investigación científica en Ingeniería Civil usando protocolos experimentales semánticos **Ingrid Torres, Jaime Guzmán L. y Juan Muñoz F.**	61
Desarrollo de una aplicación móvil para la localización en espacios interiores por medio de análisis de radiación de redes WiFi **Carlos Gómez R.**	72
Adaptative particle swarm optimization algorithm with non-iterative electrostatic repulsion and social neighborhood **Daniel Soto y Wilson Soto**	80
Estudio y análisis de la regulación y normativa en el servicio de datos móviles para tecnologías 4G en Colombia **Jhon Gutiérrez L., Germán Góez S., Sergio Serna G.**	88
Capacidad del Ansys Fluent® V 16.0 para determinar la fuerza de arrastre en partículas no esféricas tipo discos cilíndricos. **Jorge Sierra, Mohammed Khalij, Anne Tanière, Diego Hincapié y Sebastián Vélez**	100
Custom processors design using Python-based high level synthesis **Jaime Parra P.**	107
Detección de actividad en señales de voz utilizando la transformada wavelet discreta **Kevin Rodríguez, Andrés Castañeda y Dora Ballesteros**	118
Concentración de redes Wifi en sitios céntricos - Caso de estudio Centro de la ciudad de Medellín Colombia **Roberto Guevara C. y Giovanny Flórez**	127
Implementación de un sistema de conteo de estudiantes utilizando redes de sensores inalámbricos **Auris Parody Z., Jorge Garcia G., Jorge Chinchilla A. y Jorge Ortiz**	136
Cifrado de imágenes en fase usando la transformada fraccionaria de Hartley y operaciones de truncamiento **Juan Vilardy O., Lenin Nuñez R. y Cesar Torres M.**	148
Metodología para el desarrollo de software integrando SOA, BPM y TOGAF **Lourdes De Avila Gutiérrez**	155
Segmentación de contornos ventriculares izquierdos, mediante el uso de la transformada generalizada de Hough **Hernando Velandia V., Eduard García C. y Luis Mendoza**	163
Caracterización del sensor EEG BR8 Plus: Verificación del estado base y los ritmos beta y theta utilizando la prueba Stroop de atención visual **Estefanía Millán D., Carolina Jiménez M., Daniela Ospina T. y Oscar Henao G.**	172
Uso de tecnologías emergentes para el monitoreo de tráfico vehicular **Daniel Silva, Johan Henao, César Pedraza y Félix Vega**	182
Desarrollo de un plataforma para el control inteligente de vehículos basada en identificación por radio frecuencia **Ernesto Neira, John Jairo Pantoja, Natalie Figueroa, Gabriel Mañana, Cesar Pedraza, Félix Vega y Paola Ordoñez**	192
Diseño y fabricación de un sistema monopulso para aplicaciones RFID **Carolina Alcaraz, Juan Balbastre y Félix Vega**	204

Hacia una plataforma de software libre de tutoría inteligente para B-learning: Arquitectura de referencia **Martha Giraldo G. y Armando Ordóñez**	212
Sistema para caracterización de paneles fotovoltaicos **Carlos Robayo L. y Sergio Serna G.**	220
Dinámicas de inventarios en el sector productivo **Federico Liévano M.**	227
Métricas del software - Herramientas de apoyo cuantificables para la toma de decisiones **Piedad Metaute P. y Alexei Serna A.**	237
Descripción de un proceso de autenticación de una firma manuscrita usando procesamiento digital de imágenes **José Espitia, Cesar Torres, Marjorie Jácome y William Lasso**	246
Construcción de una herramienta usando software libre para la gestión de la red nacional de estaciones GNSS permanentes con propósitos geodinámicos de Colombia **Leonardo Cardona, Yuli Corchuelo y Héctor Mora**	255
Desarrollo de una ontología sobre estudios de impacto ambiental orientada a sistemas de recuperación documental **Jaime Guzmán L., Jennifer Cartagena O. y Edna Rodríguez G.**	266
Arquitectura empresarial y metodologías ágiles - Una combinación efectiva para hacer frente a los frecuentes cambios en el negocio **Alejandro Álvarez C. y Albeiro Patiño B.**	277
Caracterización y calibración de un sensor de humedad de bajo costo para optimizar su desempeño dentro de una red WSN en escenarios indoor **Luis Noreña T. y Diego Cortés**	289
Interfaz para navegación autónoma de discapacitados visuales en salas de exposiciones museográficas **Mauricio Arias C., Anderson Arango, Stiven Agudelo, Dany Urrego y Jakeline Serrano**	297
Optimización del enrutamiento de datos en una red de sensores inalámbricos basado en un algoritmo multi-objetivo **Ángela Rodríguez, Armando Ordóñez y Luis Eraso**	308
Análisis y diseño de un prototipo suelo generador de energía eléctrica utilizando materiales piezoeléctricos **Dionel Castro S. y Emilsen Arias C.**	315
Redes inalámbricas de sensores aplicadas al monitoreo de integridad estructural **Gustavo Meneses**	326
Diseño de un sensor de permitividad dieléctrica relativa de un medio empleando una antena de microcinta con estructuras metamateriales **María Domínguez V., Daniel Cataño O. y Erick Reyes V.**	339

Validación de modelos numéricos durante el desarrollo de correlaciones para intercambiadores de calor

Juan Ardila M.
Diego Hincapié Z.
Julio Casas M.
Instituto Tecnológico Metropolitano – Colombia

1 Introducción

Diferentes métodos para mejorar la tasa de transferencia de calor en intercambiadores han sido investigados [1]; esto ha permitido identificar distintas técnicas para el mejoramiento en la eficiencia de transferencia, que se clasifican en dos categorías: 1) técnicas activas y 2) técnicas pasivas. Las primeras requieren alimentación externa (vibración o campos electromagnéticos), mientras que las segundas no la necesitan (geometrías especiales o inserciones dentro del tubo). Un método pasivo usado ampliamente es el de tubos en espiral [2], en el que el flujo secundario causado por la fuerza centrífuga aumenta la transferencia de calor [3].

Computational Fluid Dynamics (CFD) es la rama de la mecánica de fluidos que emplea métodos numéricos, algoritmos y programación para el análisis de problemas y fenómenos en toda tipo de fluidos [4]. Los computadores son empleados para la realización rápida de millones de cálculos relacionados con las ecuaciones y los modelos matemáticos acoplados al fenómeno. Los paquetes CFD comerciales han se utilizan para predecir el desarrollo del flujo de masa y el térmico, obteniendo resultados coherentes con los experimentales en intercambiadores de calor de tubos en espiral concéntricos, en intercambiadores de coraza y en tubos en espiral (rectangular, con flujo bifásico, con flujo monofásico, contorsionado), empleando los programas: FLUENT®, CFX®, PHOENICS® y ANFIS® [1, 5-14].

Cuando se diseñan bancos experimentales se busca simular las condiciones en las que se presenta el fenómeno, sea natural o industrial. Se procede de la misma manera cuando se configura la simulación y se seleccionan los modelos más complejos que utilizan la menor cantidad de simplificaciones. Dichos modelos han mostrado resultados que se acercan a los experimentales, pero requieren alto poder de cómputo. CFD se ha empleado para los siguientes estudios en diversos tipos de intercambiadores de calor: mala distribución del flujo de los fluidos, suciedad o incrustaciones, caída de presión y análisis térmico en la fase de diseño y optimización. Los diferentes modelos de turbulencia disponibles con fines comerciales (estándar, realizable, RNG, RSM y SST), en relación con los métodos numéricos de acoplamiento velocidad-presión (SIMPLE, SIMPLEC, PISO,…), se adoptan para llevar a cabo las simulaciones. La calidad de las soluciones obtenidas a partir de ellas está, en gran parte, dentro del intervalo aceptable, demostrando que CFD es una herramienta eficaz para predecir el comportamiento y el rendimiento de una amplia variedad de intercambiadores de calor [15].

En el presente trabajo se busca validar modelos numéricos de intercambiadores de calor, uno convencional y otros mejorados con técnicas pasivas, desarrollados desde ANSYS®. En este trabajo se contrastan los resultados con los obtenidos en las investigaciones experimentales publicadas en la literatura científica.

2 Metodología

Todos los desarrollos se realizaron utilizando las diferentes herramientas de ANSYS®, las geometrías se desarrollaron en el módulo DesignModeler®, los enmallados se realizaron en el módulo Meshing® y las simulaciones se configuraron con el módulos CFD: CFX® y Fluent® para corroborar los resultados entre ellos. Solamente al final se decide seguir trabajando únicamente con CFX®. El estudio consta de tres proyectos de simulación: en el primero se desarrolló el fluido interno en un intercambiador de calor de tubo liso de 500mm de longitud y 25mm de diámetro. En el segundo se hizo en un tubo torsionado de 560mm de longitud y 18mm de diámetro, con dos regiones lisas en la entrada y la salida del tubo de 30mm cada una y una región corrugada de 500mm, la vena corrugada helicoidal con paso de 15.95mm y profundidad de 1.03mm. En el tercero una espira de tubo flexionado en espiral helicoidal con diámetro del tubo 20mm, diámetro de la hélice 370mm, longitud 1162mm y paso de hélice 40mm. En este último proyecto se compararon los resultados con los obtenidos para dos espiras, debido al interés en saber si basta con una espira para obtener resultados de flujo completamente desarrollado. Para esta segunda geometría se tiene diámetro de tubo 20mm, diámetro de hélice 370mm, longitud 2324mm y paso de hélice 40mm. En la Figura 1 pueden apreciarse las geometrías descritas.

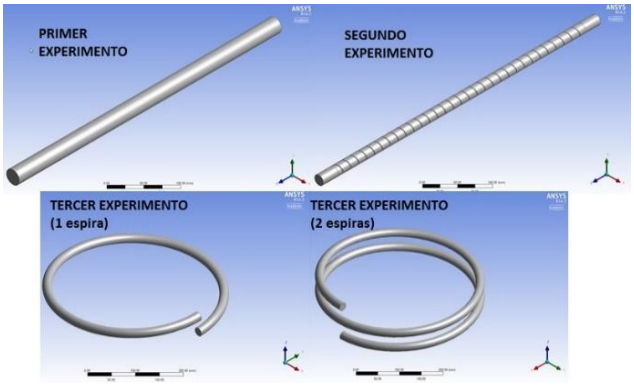


Figura 1: Geometrías aplicadas en los proyectos

En esta investigación se empleó el algoritmo SIMPLEC para la solución de los modelos termo-fluidos que implican el uso de la ecuación de la energía y ecuaciones que caractericen el comportamiento del campo fluido. Allí aparece el modelo de ecuaciones de Navier-Stokes para flujo laminar y los modelos de turbulencia k-ε, k-ε realizable y k-ω. El algoritmo resuelve los modelos en cada nodo de las mallas que describen las geometrías de intercambiadores de calor de tubos concéntricos, y por último se compararon los resultados obtenidos con datos experimentales publicados buscando el acercamiento mediante refinamiento de malla y selección apropiada de modelo matemático.

El número de Nusselt (Nu) es el gradiente de temperatura adimensional en la superficie de transferencia, y caracteriza la transferencia de calor midiendo el aumento de la transferencia de calor por convección comparada con la transferencia si ocurriera solamente por conducción [16]:

$$Nu_D = \left.\frac{\partial T^*}{\partial y^*}\right|_{y^*=0} = \frac{hD}{k_f} = f(Re, Pr) \qquad (1)$$

Donde y^* es una distancia medida perpendicularmente desde la superficie hacia el interior del fluido; h es el coeficiente convectivo de transferencia de calor; D es el diámetro

de la tubería, que resulta ser la longitud característica de la geometría involucrada en este análisis; y k_f es la conductividad térmica del fluido. El número de Nusselt debe ser alguna función universal del número de Reynolds (Re) y del número de Prandtl (Pr). Si se conociera está función podría encontrarse el coeficiente convectivo en el proceso de diseño de intercambiadores de calor. El parámetro adimensional de Prandtl es la razón de difusividades de momento y térmica, y es una característica propia del fluido que mide la efectividad relativa del transporte de momento y energía por difusión en las capas límite hidrodinámica y térmica. El número de Reynolds es una característica propia del flujo y la geometría; es la razón de las fuerzas de inercia a las fuerzas viscosas en la capa límite hidrodinámica y determina la existencia de flujo laminar o turbulento según predominen las fuerzas de velocidad a las de fricción [17]:

$$Re = \frac{\rho V D}{\mu} \qquad (2)$$

$$Pr = \frac{c_p \mu}{k_f} \qquad (3)$$

Donde μ es la viscosidad del fluido; ρ es su densidad; c_p es su calor específico; y V es la velocidad media del flujo. La ecuación (4) es una correlación que se atribuye a Sieder y Tate en 1936 y que aplica al flujo interno turbulento completamente desarrollado en tubos largos circulares con grandes variaciones en las propiedades para 0.7≤Pr≤16700; donde todas las propiedades se evalúan en una temperatura media, excepto μ, y de su uso pueden resultar errores hasta del 25% [16]:

$$Nu_D = 0.027 Re_D^{0.8} Pr^{0.2} \left(\frac{\mu}{\mu_s}\right)^{0.14} \qquad (4)$$

Petukhov desarrolló en 1970 una correlación más compleja y redujo los errores de uso al 10% para 0.5≤Pr≤2000 [16]:

$$Nu_D = \frac{(f/8)\, Re\, Pr}{1.07 + 12.7 (f/8)^{0.5} \left(Pr^{2/3} - 1\right)} \qquad (5)$$

Donde f es el factor de fricción, y donde estas correlaciones para tubo liso son experimentales. Vicente publica en 2004 las medidas experimentales hechas con tubo torsionado en régimen laminar y transicional completamente desarrollados [1]. En la Figura 2 pueden apreciarse dichos resultados.

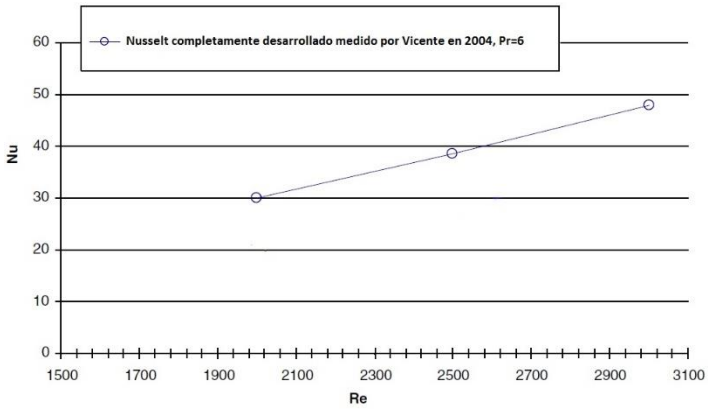


Figura 2: Resultados experimentales para tubo torsionado con flujo en régimen laminar y transicional [1]

Para determinar el Nusselt teórico, a partir de los resultados de las simulaciones, se sigue un procedimiento similar al experimental para inducir un mismo error, partiendo por calcular el coeficiente convectivo a partir de la ley de enfriamiento de Newton:

$$\dot{q} = h(T_s - T_\infty) \qquad (6)$$

Donde $\dot{q}$ es el flux de calor en la superficie; T_s es la temperatura de la superficie; y T_∞ es la temperatura media del fluido. Estos datos son resultados de las simulaciones y llevados al cálculo de coeficientes convectivos y, posteriormente, de números de Nusselt.

3 Resultados

3.1 Primera geometría: Tubo recto liso

Enmallado con método de tetraedros conformes con tamaños mínimo de 0.01mm, máximo de 2mm y máximo de cara 1mm, y con inflación por defecto en pared, reportan para Fluent® 1´183.547 elementos y para CFX® 1´131.383. Se obtuvieron las siguientes condiciones de frontera: salida a presión de 0Pa manométricos, pared temperatura constante 40°C, entrada temperatura constante 20°C y velocidades de 0.5, 0.75, 1, 1.25 y 1.5m/s. Se ejecutaron cuatro experimentos para seleccionar el modelo y la temperatura promedio apropiados. Experimento 1: modelo k-ε estándar con funciones de pared escalables y propiedades constantes a 25°C; experimento 2: modelo k-ω y propiedades constantes a 25°C (Figura 3).

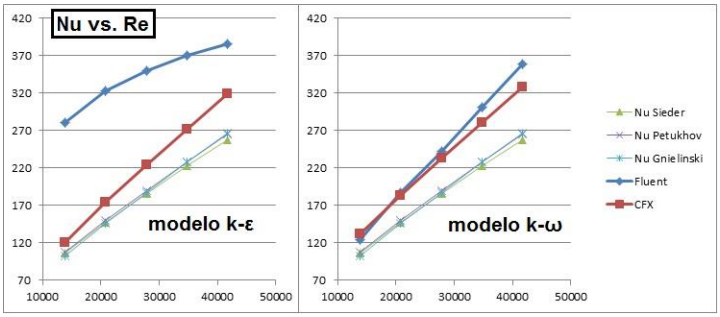


Figura 3: Tubo recto liso comparación de modelos

Se decidió trabajar con k-ω y se empezó la selección de la temperatura iterando con las nuevas temperaturas promedio que resultaban. Experimento 3: modelo k-ω y propiedades constantes a 22°C; experimento 4: modelo k-ω y propiedades constantes a 21.3742°C (Figura 4).

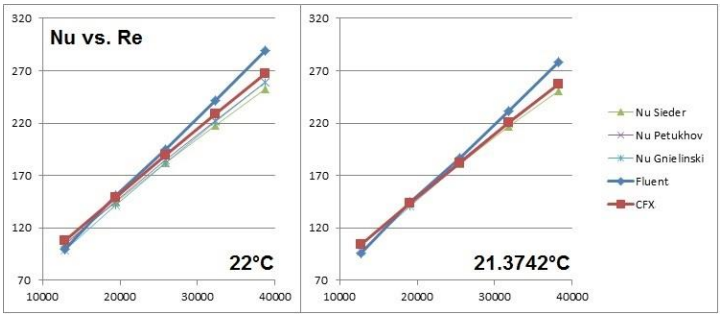


Figura 4: Tubo recto liso comparación de temperaturas medias

3.2 Segunda geometría: Tubo recto torsionado

Enmallado con método de tetraedros conformes con tamaños mínimo 0.01mm, máximo 2mm y máximo de cara 1mm, y con inflación por defecto en pared, que reportan las siguientes métricas: para Fluent® 6´897.372 elementos y para CFX® 6´071.295 elementos. Con las siguientes condiciones de frontera: salida a descarga libre, pared con flux de calor constante 10kW/m2, entrada temperatura constante 20°C y velocidades de

0.097, 0.1075, 0.1211, 0.1344, 0.1452 y 0.1615m/s. Se ejecutaron cuatro experimentos para seleccionar el modelo y la temperatura promedio apropiados, pero además fue necesario evaluar un refinamiento de malla. Experimento 1: modelo laminar y propiedades constantes a 22°C; experimento 2: modelo k-ω y propiedades constantes a 22°C; y experimento 3: modelo k-ε realizable con funciones de pared escalables y propiedades constantes a 22°C (Figura 5).

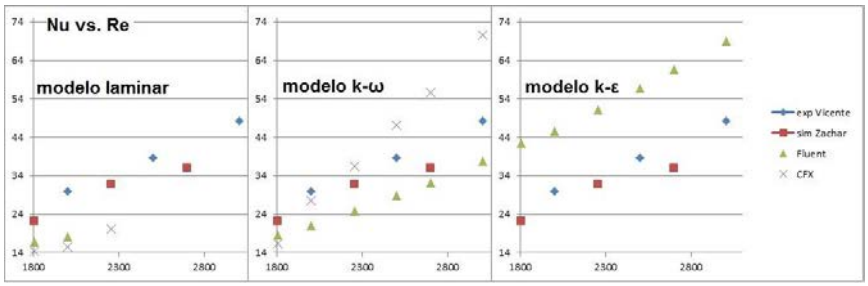


Figura 5: Tubo recto torsionado comparación de modelos

Se decidió trabajar con k-ω y se hizo un refinamiento de malla con tamaños mínimo 0.01mm, máximo 0.5mm y máximo de cara 0.25mm, y con inflación de 10 capas en pared, reportando las siguientes métricas: para Fluent® 29´710.505 elementos y para CFX® 6.071.295 elementos. Experimento 4: modelo k-ω y propiedades constantes a 22.876°C (Figura 6).

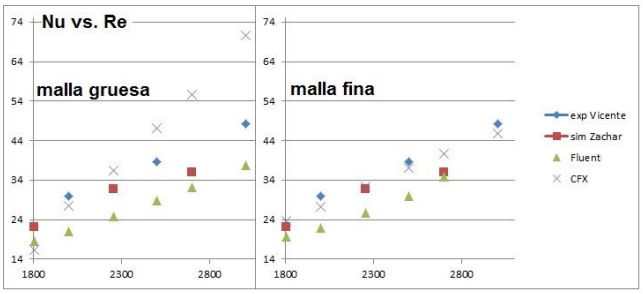


Figura 6: Tubo recto torsionado incidencia del refinamiento de malla

3.3 Tercera geometría: Tubo en espiral helicoidal liso

Enmallado con método de tetraedros conformes con tamaños mínimo 0.01mm, máximo 0.75mm y máximo de cara 0.375mm para CFX®, y mínimo 0.01mm, máximo 1mm y máximo de cara 0. 5mm para Fluent®, y con inflación por defecto en pared, reportando las siguientes métricas: para Fluent® 11.853.843 elementos y para CFX® 19.523.490 elementos. Los resultados se compararon con los obtenidos para un intercambiador de dos espiras (Figura 7), debido al interés en saber si basta con una para obtener resultados de flujo completamente desarrollado.

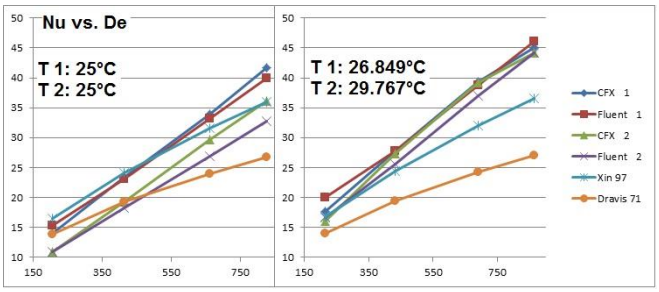


Figura 7: Tubo en espiral helicoidal liso comparación de temperaturas medias

Enmallado con método de tetraedros conformes con tamaños mínimo 0.01mm, máximo 2mm y máximo de cara 1mm para CFX® y para Fluent®, y con inflación por defecto en pared, reportando las siguientes métricas: para Fluent® 4.039.175 elementos y para CFX® 3.581.967 elementos. Las siguientes condiciones de frontera: salida a presión 0Pa manométricos, pared con temperatura constante 60°C, entrada temperatura constante 20°C y flujos másicos de 0.0125, 0.025, 0.04, 0.05kg/s. Se ejecutaron seis experimentos buscando seleccionar el modelo, el refinamiento de malla y la temperatura promedio (para evaluar las propiedades) apropiados. Experimento 1: modelo k-ω y propiedades constantes a 25°C; experimento 2: modelo k-ω y propiedades constantes a 26.849°C (una espira) y 29.767°C (dos espiras).

Los resultados muestran que la diferencia entre una o dos espiras radica en la temperatura promedio de evaluación de propiedades sin importar el refinamiento de la malla, por lo tanto se decidió seguir trabajando con el modelo de una espira por economía de recursos informáticos. También se estudió la incidencia del engrosamiento de malla y se corrió un tercer experimento para una espira con método de tetraedros conformes con tamaños mínimo 0.01mm, máximo 1.5mm y máximo de cara 0.75mm para CFX® y para Fluent®, y con inflación por defecto en pared. Se reportaron las siguientes métricas: para Fluent® 4´321.425 elementos y para CFX® 3.875.839 elementos. Enmallado para dos espiras con método de tetraedros conformes con tamaños mínimo 0.01mm, máximo 1.5mm y máximo de cara 0.75mm para CFX® y para Fluent®, y con inflación por defecto en pared Se reportan las siguientes métricas: para Fluent® 8.135.430 elementos y para CFX® 7.790.153 elementos. Experimento 3 (Figura 8): modelo k-ω y propiedades constantes a 26.849°C (una espira) y 29.767°C (dos espiras).

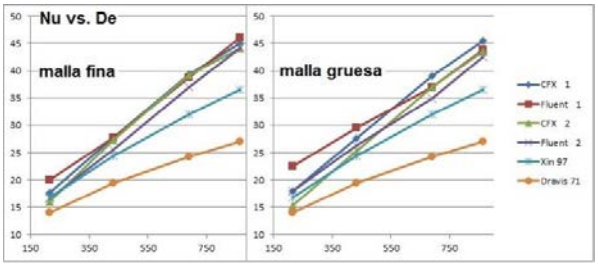


Figura 8: Tubo en espiral helicoidal liso incidencia del refinamiento de malla

Los resultados se diferenciaron en que el refinamiento de malla es el que permite trabajar con una sola espira y obtener resultados muy similares al caso de dos espiras. Se decidió estudiar la incidencia del modelo corriendo algunas simulaciones con modelos k-ε, solamente con una espira, e inicialmente con una malla gruesa con método de tetraedros conformes con tamaños mínimo 0.01mm, máximo 2mm y máximo de cara 1mm para CFX® y para Fluent®, y con inflación por defecto en pared, reportando las siguientes métricas: para Fluent® 2.100.867 elementos y para CFX® 1.874.986 elementos. Experimento 4 (Figura 9): modelo k- ε y propiedades constantes a 26.849°C.

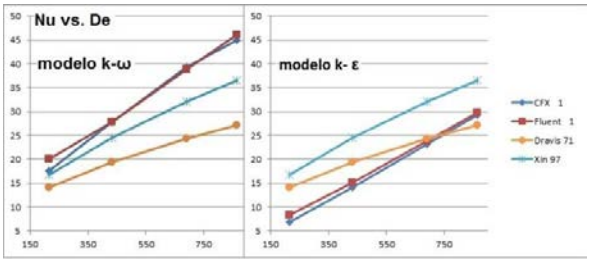


Figura 9: Tubo en espiral helicoidal liso comparación de modelos

A partir de este momento se hizo una modificación en el programa para realizar las comparaciones: las propiedades se venían evaluando en una temperatura promedio para todas las velocidades sabiendo que el incremento de velocidad disminuye la temperatura de salida y por tanto se afecta el valor promedio de temperatura, así que se decidió establecer una temperatura promedio para cada velocidad y evaluar en ella las propiedades, pero configurando de manera separada cada fluido en cada simulación, lo que toma más tiempo pero resulta más realista. Se escogió el modelo k-ω y se refinó la malla con método de tetraedros conformes con tamaños: mínimo 0.01mm, máximo 0.75mm y máximo de cara 0.375mm para CFX® y mínimo 0.01mm, máximo 1mm y máximo de cara 0.5mm para Fluent®, y con inflación por defecto en pared, reportando las siguientes métricas: para Fluent® 11.853.843 elementos y para CFX® 19.523.490 elementos. Experimento 5 (Figura 10): modelo k- ω y propiedades constantes a temperatura media para cada velocidad.

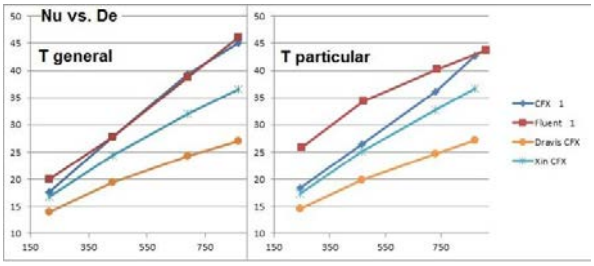


Figura 10: Tubo en espiral helicoidal liso incidencia del modelo de comparación

Aun con el refinamiento de malla y la selección realista de temperatura promedio para cada velocidad, los resultados siguen por encima de Xin con el modelo k-ω, replicando los resultados del primer experimento de esta geometría. Dadas las bajas velocidades se decidió modelar como flujo laminar para ganar algo de tiempo de cómputo y se sigue trabajando con una malla más fina. Debido a la lentitud de procesamiento experimentada con Fluent®, a partir de este punto se trabaja solamente con CFX®. La malla con método de tetraedros conformes con tamaños: mínimo 0.01mm, máximo 0.5mm y máximo de cara 0.25mm para CFX®, y con inflación por defecto en pared reporta 53.720.866 elementos. Experimento 6 (Figura 11): modelo laminar y propiedades constantes a temperatura media para cada velocidad. Siendo estos los mejores resultados alcanzados.

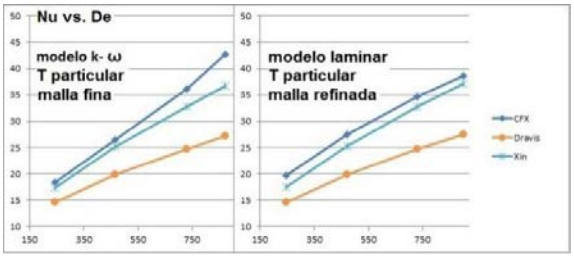


Figura 11: Tubo en espiral helicoidal modelo laminar malla refinada

4 Conclusiones

Este proyecto permitió validar los resultados teóricos que se pueden alcanzar con herramientas CFD comerciales, al mostrar el acercamiento que se pueden obtener comparados con resultados experimentales, por lo tanto válida la herramienta para el desarrollo de correlaciones que facilitan el diseño de diferentes geometrías de intercambiadores de calor.

Los resultados teóricos de la simulación con paquetes comerciales se ven influenciados por el modelo seleccionado y su configuración, la calidad del enmallado, la selección y configuración del algoritmo, y otros parámetros. Todos deben ser evaluados antes de emplearlos en la toma de decisiones de diseño y es necesario usar resultados experimentales para dicha validación.

Las diferencias entre los resultados de la simulación y los resultados experimentales publicados se deben a las simplificaciones que implican los modelos teóricos que idealizan la naturaleza y los errores involucrados en la toma de mediciones de campo.

Hay que tener en cuenta que los modelos semi-empíricos en que se basa la simulación son idealizados y arrojan resultados que por naturaleza deben diferir de los resultados experimentales, de manera que alcanzar los resultados experimentales a través del refinamiento de malla no quiere decir que los resultados de la simulación bajo esas condiciones garanticen igualdad con la realidad, solamente puede confiarse en el resultado de simulación cuando el error relativo se estabiliza.

Las diferencias entre los resultados de CFX® y Fluent® se deben a la utilización de modelos con parámetros predeterminados que son diferentes en cada módulo, y aunque se seleccionaron los mismos modelos los programas resuelven ecuaciones ligeramente diferentes.

En general, los resultados actuales, tanto de CFX® como de Fluent®, se aproximan a los reportados en la literatura, pero se selecciona CFX® para trabajos posteriores porque presenta mayor acercamiento sin cambiar los parámetros por defecto del modelo y porque las soluciones convergen más rápido.

Referencias

[1] Zachár, A. (2010). Analysis of coiled-tube heat exchangers to improve heat transfer rate with spirally corrugated wall. International Journal of Heat and Mass Transfer 53(19), pp. 3928-3939.

[2] Shokouhmand, H. et al. (2008). Experimental investigation of shell and coiled tube heat exchangers using wilson plots. International Communications in Heat and Mass Transfer 35(1), pp. 84-92.

[3] Minton, P. (1970). Designing spiral tube heat exchangers. Chemical engineering 77(11), pp. 145-152.

[4] Tu, J., Yeoh, H. & Liu, C. (2013). Computational Fluid Dynamics. Oxford: Butterworth-Heinemann - ElSevier.

[5] Rennie, T. & Raghavan, V. (2006). Numerical studies of a double-pipe helical heat exchanger. Applied Thermal Engineering 26(11), pp. 1266-1273.

[6] Mehrabi, M., Pesteei, S. & Pashaee G. (2011). Modeling of heat transfer and fluid flow characteristics of helicoidal double-pipe heat exchangers using Adaptive Neuro-Fuzzy Inference System (ANFIS). International Communications in Heat and Mass Transfer 38(4), pp. 525-532.

[7] Huminic, G. & Huminic, A. (2011). Heat transfer characteristics in double tube helical heat exchangers using nanofluids. International Journal of Heat and Mass Transfer 54(19), pp. 4280-4287.

[8] Kumar, V. et al. (2008). Numerical studies of a tube-in-tube helically coiled heat exchanger. Chemical Engineering and Processing 47(12), pp. 2287-2295.

[9] Jayakumar, J. et al. (2010). CFD analysis of single-phase flows inside helically coiled tubes. Computers and Chemical Engineering 34(4), pp. 430-446.

[10] Naphon, P. & Wongwises, S. (2002). An experimental study on the in-tube convective heat transfer coefficients in a spiral coil heat exchanger. International Communications of Heat and Mass Transfer 29(6), pp. 797-809.

[11] Zhao, Z. et al. (2011). Numerical studies on flow and heat transfer in membrane helical-coil heat exchanger and membrane serpentine-tube heat exchanger. International Communications in Heat and Mass Transfer 38(9), pp. 1189-1194.

[12] Conte, I. & Peng, X. F. (2009). Numerical and experimental investigations of heat transfer performance of rectangular coil heat exchangers. Applied Thermal Engineering 29(8), pp. 1799-1808.

[13] Di Piazza, I. & Ciofalo, M. (2010). Numerical prediction of turbulent flow and heat transfer in helically coiled pipes. International Journal of Thermal Sciences 49(4), pp. 653-663.

[14] Vashisth, S. & Nigam, K. (2009). Prediction of flow profiles and interfacial phenomena for two-phase flow in coiled tubes. Chemical Engineering and Processing: Process Intensification 48(1), pp. 452-463.
[15] Aslam, M. et al. (2012). CFD applications in various heat exchangers design: A review. Applied Thermal Engineering 32, pp. 1-12.
[16] Bergman, T. et al. (2007). Fundamentals of heat and mass Transfer. New Jersey: John Wiley and Sons.
[17] Ardila, J. & Hincapié, D. (2012). Intercambiadores de Calor de Tubo en Espiral. Revista UIS Ingenierías 11(2), pp. 203-213.

Design, implementation and testing of a characterization photovoltaic panels system

Álvaro Jaramillo
Bonie Restrepo
Jorge Arango
Alfonso Berrio
Instituto Tecnológico Metropolitano – Colombia

1 Introduction

To understand the operation of an array of photovoltaic (PV) panels, it is necessary to consider the individual behavior of a panel and each of the cells that compose it [1, 2]. To characterize a PV panel, it is only considered direct mode of operation (as a generator). However reverse mode operation of a PV panel (as a load), must be consider, although it is a less common analysis [3, 4]. In a panel, the reverse mode operation can occur for a damaged cell, or because the electrical characteristics of some cells are not the same, or if some of the modules are exposed to partial or total shading. As a result, these affected cells do not generate electricity and otherwise behave as a load, consuming energy coming from rest of the array. Consequently these cells dissipate energy as heat, resulting in some cases in permanent damage to the panel [5-7].

In order to evaluate the performance of the photovoltaic module, in direct and reverse mode of operation, it is necessary to make an electrical and thermal characterization thereof. For electrical characterization, experimental measurements must be made of Current-Voltage (I/V) and Power-Voltage (P/V) curves. As a first step, it was proposed to develop a system for characterization of photovoltaic panels for direct mode operation (as a generator). The second stage of this project, which is not reported in the article seeks to characterize and obtain the curves of a PV panel in the reverse mode of operation (as a load).

To obtain the characteristic curve in direct mode operation of a PV panel, it is possible to use different methods described in the literature [8]. However, the cost of the necessary equipment, the time required to make the test, the accuracy and reliability of the results, are factors that must be considered to choose the most appropriate alternative. The procedure to obtain the characterization curves, consist in connect an element, active or passive, to the panel terminals and measure the voltage and current signals. Ideally, this type of testing have to be made when the panel is exposed to constant radiation and temperature. Below are define some of the methods described in the literature:

- Variable Resistance: The resistance value is changed in incremental steps, starting at zero ohm and until the current is approximately zero amps. This method is only applicable to low-capacity PV panels due to the need of a high power resistors to dissipate energy coming from the panel [9].
- Capacitive Load: When the capacitor is connected to the panel, its charge increases, current drops and voltage rise. When the capacitor is fully charged, the current supplied by the photovoltaic panel becomes zero. To obtain reliable curves, a good quality capacitor is required, this means that it has a low equivalent series resistance [8].

- Electronic Load: Uses a transistor (usually a MOSFET) as a load. When this method is used to trace I/V curve of the PV panel, the MOSFET must operate in three modes: cut-off, active, and in the Ohmic region. During the process, most of the power delivered by the PV panel must be dissipated by the MOSFET, limiting its application to medium power panels [10].
- Bipolar Amplifier Power: It is based on a class B amplifier, which use two transistors BJT as a load. Bipolar transistors must operate in three modes: cut, active and saturation region. Therefore, as in the method of electronic load, most of the power supplied by the PV panel must be dissipated by the transistors. This restrict its application to medium power panels [8].
- Four Quadrants Power Supply: A power supply whose output can be varied by a reference input signal or programmed to perform sweeps on a specific range of values. These types of power supplies are relatively expensive but offer the most reliable results [11].

To characterize different PV panels, it has been proposed to solve the problem in several steps. The first stage, presented in this paper, shows the development and implementation of a test system for the reconstruction of electrical characteristics curves of PV panels in the first quadrant (direct mode), using the method of the electronic load. The electronic load method was chosen over the other mentioned methods above, as it is an available element in the institute labs. This device can be programmed and operated remotely. Thus, it possible to perform the characterization automatically, over different conditions. The capacity of the device available, fits the maximum power of the PV panel available.

2 Methodology

The development environment included in Matlab® was used to implement the prototype and its graphical user interface. This environment was chosen because of its versatility and ability to interact with hardware, in this case, the electronic load. The configuration, operation and reading data of electronic load was done through serial communication using the protocol of electronic load [12]. The program, created in Matlab® development environment, communicates with the electronic load, adjusting its parameters and sending a series of instructions for performing the test. Additionally, the electronic load takes record of voltage, current and power and sends it to the program. The program is designed to store the readings made by the load and once the test is complete, the program displays the results graphically. The operation of the prototype was tested first with a simulator of photovoltaic panels and finally with some panels exposed to the midday sun, in the city of Medellín, Colombia. In both cases, the panel and the simulator, were connected directly to the electronic charge, which was controlled by software. The prototype development was divided into three phases as described below.

2.1 Remote configuration and control of the electronic load

The remote configuration and control of the electronic load were achieved through serial communications using the libraries available in Matlab and according to the communication protocol of the electronic load [12]. All code was written from scratch. The configuration and operation of the load was included in the main function of the program, which is represented in the block diagram shown in Fig. 1. Each instruction of the main function, corresponding to sub-functions that perform a simple task, as the adjustment of a certain configuration parameter, reading of the variable or the operation mode of the

load. The objective was to develop a modular program to facilitate debugging and code reuse in other sub-functions.

As can be seen in Figure 1, the initial part of the program defines the parameters for serial communication with the electronic load. Then, setting for maximum values of voltage, current and power are made, these values are associated with the nominal operating of the PV panel to be characterized.

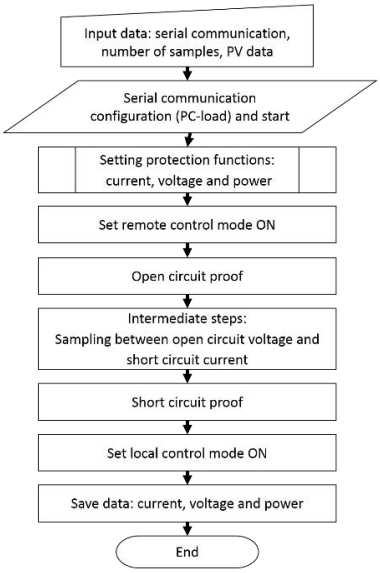


Figure 1: Block diagram of the main function

For the characterization of any panel, was proposed a reconstruction point to point of the IV and PV curves. The test starts measuring the open circuit voltage, then a sweep, point to point, between the open circuit voltages to the nearest value to zero volts is performed by the program. The number of steps to rebuild the graphs is chosen by the user in the GUI (Graphical User Interface). In each stage the operation mode of the load, including current, voltage or constant impedance is changed as is need. For the open-circuit voltage, for example, the constant current mode is set to a value of zero amps. For the intermediate steps, the load mode is set to constant voltage and varying the voltage going in decreasing values. Finally the load is adjusted in constant impedance mode, with a value of zero for short circuit testing.

2.2 Design and development of the prototype

For GUI prototype input parameters, according to their origin were grouped as below; Serial communication: speed of data transmission and communication port; Test characteristics: number of samples per curve; Characteristics of PV panel: Nominal operating voltage and maximum power output. For data input, we designed and developed a graphical interface that offers an experience simple and intuitive user, additionally prevents the entry of invalid data, as can be seen in Figure 2. The application includes an option allows exporting data, IV and PV curves, to an Excel and Matlab (.mat) format; additionally export an image of the curves in png format.

2.3 Prototype testing

Two tests were developed to test the prototype: the first using a programmable DC source, which allows the simulation of a photovoltaic panel; and the second, with a

photovoltaic panel outdoors. In both cases feeding the electronic load controlled by the software prototype developed.

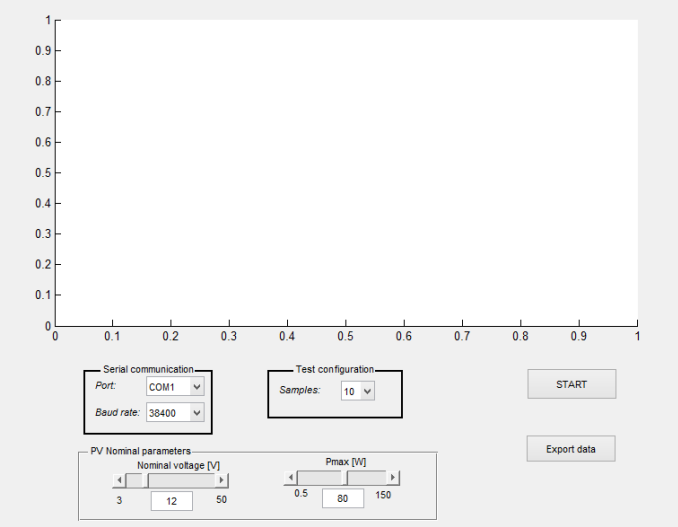


Figure 2: Graphical interface user

3 Results

PV panels simulator Chroma 62000H was used in the first stage of tests [12]. The

Figure 3 Figure 3 shows a schematic of the communication between the PC running the program and the electronic load; further the electrical connection between the load and the simulator panels. Some tests were performed using the characteristics of PV panels HYBRITEC-QM5-85/12, HYBRITEC-M5 30/12 and HYBRITEC 10/12 with standard test conditions (Irradiance: 1000W/m2, cell temperature: 25°C and a spectral distribution of the incident radiation of 1.5). Initially linearly spaced measures, with respect to the voltage, were taken, but to obtain a more reliable behavior of the panel, we chose a denser spacing in the nonlinear region of the curve, which is one of the most important characteristic of the panel, called Maximum Power Point (MPP).

Photovoltaic panel HYBRITEC-QM5-85/12, exposed to irradiation noon on a sunny day was used for the second stage of testing. Irradiation and temperature measurements were taken manually using the MacSolar device [14].

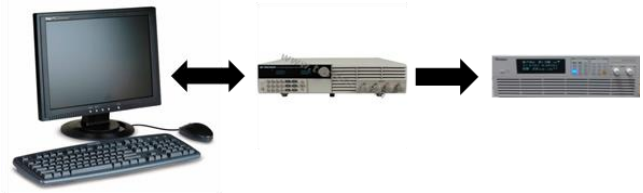


Figure 3: Communication between the computer and the electronic load, and connection between the electronic load and panel simulator

3.1 Test based on a PV panels simulator device

In Figure 4, the results of the simulation of Hybrytec 10/12 PV panel are shown, using the Chroma 62000H PV device. To simulate the behaviour of Hybrytec 10/12 PV panel, Chroma 62000H [11] device requires as input the maximum power: 10 W, maximum peak current: 0.58 A, maximum peak voltage: 17.40 V, open circuit voltage: 21.80 V and short circuit current: 0.63 A.

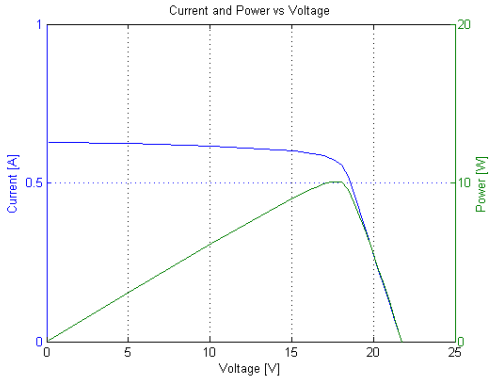


Figure 4: Characteristic curves of the simulated panel Hybrytec 10/12

3.2 Test using the PV panel Hybrytec QM5-85/12

For performed outdoors testing, we used the PV panel Hybrytec QM5-85/12, with the following characteristics: maximum power: 85W, maximum peak current: 4.83 A, maximum peak voltage: 18.02 V, open circuit voltage: 21.85 V and short circuit current: 5.22 A. Curves I/V and P/V presented in Figure 5, were obtained at 13.00 with a temperature of 47°C and an irradiation of 1037 W/m2.

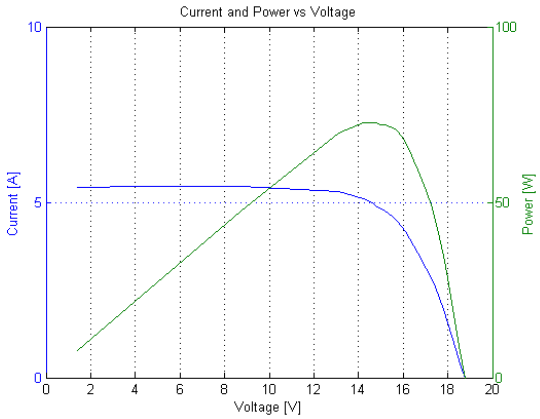


Figure 5: Characteristic curves of the simulated panel Hybrytec-QM5-85/12

Figure 6 shows the variations in the characteristic curves irradiation due to changes caused by the passage of a cloud.

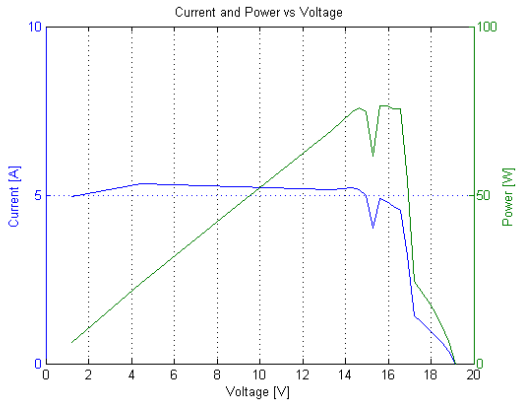


Figure 6: Characteristic curves of the PV panel Hybrytec-QM5-85/12

4 Conclusions

A program with graphical user interface was developed using the Matlab® development environment, which allows the reconstruction of the characteristic curves of a photovoltaic panel automatically. The method of programmable electronic load was used in the test system, demonstrating its good performance and reliability in the results

The performance of the program, considering every features in the internal functions and the GUI, was tested with the panel simulator Chroma 62000H. At this stage of testing, it was determined that it was not necessary to make a uniform sample measurement base on voltage. There are two main reasons to do a dynamic sampling during the test, reduce the time of the measures and obtain a more reliable curves in non-linear regions, where a denser sampling was made.

A real tests was made with PV module Hybrytec-QM5-85/12 where the correct performance of the test system was proved, software developed and electronic load.

5 Further work

For the presentation of the data in the graph, it is possible to use curve fitting techniques, such as least squares or kernel regression, in cases where there are unexpected changes of radiation and temperature by the temporal passage of clouds, ensuring a curve that represents the real behavior of the panel.

For outdoor testing, the data must be taken in the shortest time to avoid unexpected changes in radiation and temperature. A new version of the system will be implemented with a separate step of acquiring data, to obtain the values of voltage, current, plus temperature and irradiation in a faster procedure. With this additional subsystem is expected to obtain more detailed and reliable information from the PV panels.

For the second stage of the system, is intended to obtain the characteristic curves in the reverse mode of operation of a PV panel. In order to perform the complete characterization of the electrical behavior of the PV panels. These data will be used for the design of electronic devices that integrate the panels in the most appropriate manner, for isolated and plugged systems. Additionally, these data will be used for energy evaluation of PV solar generation projects.

Acknowledgements

This article was supported by Instituto Tecnológico Metropolitano under the project PM14105. Equipment used belong to the Group of Automation and Electronics Research - GIAE, of Instituto Tecnológico Metropolitano.

Referencias

[1] Salmi, T. et al. (2012). MATLAB/Simulink Based Modeling of Photovoltaic Cell. International Journal of Renewable Energy Research (IJRER) 2(2), pp. 213–218.

[2] Liu, J. (2014). Solar Cell Simulation Model for Photovoltaic Power Generation System. International Journal of Renewable Energy Research (IJRER) 4(1), pp. 49–53.

[3] Sidawi, J. et al. (2011). The effect of reverse current on the dark properties of photovoltaic solar modules. Energy Procedia 6, pp. 743–749.

[4] Zegaoui, A. et al. (2011). Photovoltaic Cell/Panel/Array Characterizations and Modeling Considering Both Reverse and Direct Modes. Energy Procedia 6, pp. 695–703.

[5] Bishop, J. (1988). Computer simulation of the effects of electrical mismatches in photovoltaic cell interconnection circuits. Solar Cells 25, pp. 73–89.

[6] Wang, Y. & Lin, S. (2011). Analysis of a partially shaded PV array considering different module connection schemes and effects of bypass diodes. Proceedings International Conference & Utility Exhibition on Power and Energy Systems: Issues and Prospects for Asia (pp. 1–7). September 28-30, Pattaya City, Thailand.

[7] Restrepo, B., Trejos, A. & Ramos, C. (2013). Modeling of PV Systems Based on Inflection Points Technique Considering Reverse Mode. TecnoLogicas edición especial octubre, pp. 237-248.

[8] Durán, E. et al. (2008). Different methods to obtain the I–V curve of PV modules: A review. Proceedings 33rd IEEE Photovolatic Specialists Conference (pp. 1-6). My 11-16, San Diego, USA.

[9] Andújar, J. et al. (2004). Sıstema para la Generación Automátıca de Curvas I-V , P-V Y Monıtorızacıón de Módulos Fotovoltaıcos. Memorias XXV Jornadas de Automática (pp. 1-6). September 8-10, Ciudad Real, Spain.

[10] Durán, E. et al. (2012). Determination of PV Generator I-V/P-V Characteristic Curves Using a DC-DC Converter Controlled by a Virtual Instrument. International Journal of Photoenergy 2012, pp. 1–13.

[11] Adamo, F. et al. (2011). Characterization and Testing of a Tool for Photovoltaic Panel Modeling. IEEE Transactions on Instrumentation and Measurement 60(5), pp. 1613–1622.

[12] BK PRECISION. 300W Programmable DC Electronic Load. Online [July 2014].

[13] © Chroma ATE Inc. (2013). Model 62000H series Programmable DC Power Supply. Online [July 2014].

[14] SOLARC. Dispositivo de medición portátil de radiación solar y temperatura (con kit de fijación/orientación). Online [July 2014].

Detección de la fatiga muscular a través de un sensor wearable de bajo costo y análisis de amplitud y frecuencia de la señal EMG superficial

María Montoya V.
John Muñoz C.
Oscar Henao G.
Universidad Tecnológica de Pereira – Colombia

1 Introducción

La fatiga es un estado de estabilidad del cuerpo y la conciencia, también conocido como cansancio, letargo, languidez, lasitud y apatía, que se asocia a debilidad física o mental [1]. Sin embargo, el concepto es complejo debido a los diferentes fenómenos fisiológicos y psicológicos que la contribuyen y caracterizan. Además, como un proceso dependiente del tiempo es empleado exitosamente por ingenieros y físicos en procesos de medición [2].

Para medir la fatiga existen varios puntos de vista subjetivos: el test de Yoshitake, que busca el sentimiento de cansancio mental y físico; el cuestionario SOFI-SM, utilizado para establecer la carga de trabajo percibida; y la escala de Krupp o Escala de Fatiga Severa [3], probablemente la más usada hasta ahora. Por otro lado, existen diferentes técnicas para la medición cuantitativa de la misma, por ejemplo, los test químicos, que evalúan gases respiratorios, orina, sangre, saliva, sudor y hormonas en busca de índices de fatiga; los test de funciones vegetativas, que evalúan temperatura, peso, apetito y energía gastada; entre otros. Una de las técnicas más utilizadas y preferidas por su carácter no-invasivo es la electromiografía de superficie (sEMG), la cual permite estudiar la actividad del músculo esquelético en tareas tanto dinámicas como estáticas [4]. El componente principal del análisis de fatiga a través de la medición de la señal eléctrica muscular es la identificación de características prominentes de los datos sEMG.

En la literatura se establecen dos componentes principales y que se presentan como bio-marcadores de fatiga en tiempo y en frecuencia: el valor RMS (Root Mean Square), que es la raíz cuadrada de la energía promedio de la señal, y los valores de frecuencia media (MNF) y la frecuencia mediana (MDF) del espectro de densidad de potencia de la curva de frecuencia. Cobb y Forbes [5] observaron un incremento consistente en la amplitud de la señal EMG recolectada con electrodos de superficie. Edwards y Lippold [6], Eason [7], Vredenbregt y Rau [8] y Maton [9] han atribuido ese incremento al reclutamiento de unidades motoras adicionales. Ellos postulan que mientras una contracción progresa, unidades motoras adicionales deben ser requeridas para mantener el nivel de fuerza constante. Bajo condiciones de fatiga, factores como el reclutamiento de las fibras musculares de contracción rápida, sincronización de unidades motoras en el musculo y el patrón de reclutamiento no-lineal, causan desplazamiento espectral hacia las regiones de baja frecuencia y una mayor magnitud en las señales [10]. Cuando una contracción es sostenida, varía el comportamiento de la forma del potencial de acción de la unidad motora (MUAP) [11]. El tiempo de contracción del MUAP es proporcional a la duración de la contracción, y para contracciones sostenidas la forma del MUAP se afecta principalmente por los cambios intracelulares de pH causados por los productos ácidos de la fatiga, como consecuencia la velocidad de conducción de las fibras musculares decae,

lo que causa la disminución de MNF y MDF y una comprensión en el espectro de frecuencias de la señal sEMG [12].

Se ha explorado una amplia cantidad de técnicas para extraer las características de la señal sEMG. Cuando es estacionaria existe una relación lineal entre la amplitud y la fuerza que se presenta en contracciones isométricas; y si no es estacionaria, es decir, no existe linealidad entre amplitud de la señal y fuerza, se presenta en contracciones isotónicas o dinámicas. También es conocido que durante las contracciones a una fuerza constante los factores que afectan las características de la señal se pueden reducir a indicadores de fatiga, como el cambio en amplitud y el espectro de frecuencias [13]. Por ejemplo, Sarmiento et al. [14] utilizan la técnica de ventana deslizante en contracciones isométricas (el músculo no se contrae ni se alarga) e isotónicas (tensión cuasi-estática), y extraen valores característicos de fatiga tales como RMS, MNF y MDF, en los rangos esperados en la literatura. En las últimas dos décadas los investigadores han empezado a estudiar la sEMG no-estacionaria para extraer indicadores de fatiga. Venugopal et al. [10] hacen la comparación del uso de diferentes técnicas de extracción de características de múltiples ventanas para extraer características de la señal sEMG, y logran extraer los indicadores de fatiga de la señal sin realizar transformaciones en el dominio de la frecuencia. Bonato et al. [15] realizaron un estudio de contracciones dinámicas cíclicas en busca de fatiga, para ellos utilizan transformaciones de clase de Cohen logrando un acercamiento ideal para el comportamiento no-estacionario de la señal.

Ahora bien, para evaluar fatiga muscular existen diferentes protocolos que dependen del tipo de contracción que se desea medir. Hernández et al. [16] evalúan sEMG con un protocolo de fatiga durante contracciones isotónicas durante entrenamiento de Taekwondo. Ellos establecen dos sesiones: la primera para determinar la repetición máxima (1RM) de cada sujeto y en la segunda midieron las contracciones cuando el atleta ejecutaba una serie de patadas con su máximo esfuerzo hasta el cansancio. Por otro lado, para contracciones isométricas, Sarmiento et al. [14] midieron fatiga en bíceps braquial, deltoides anterior y trapecio. Esta tarea isométrica se ejecutó usando cargas constantes de 30%, 60% y 80% de la contracción máxima voluntaria (MVC) del sujeto. También usaron un goniómetro (instrumento para medir ángulos [17]) para medir el decremento de 10° con respecto a la posición inicial del brazo en 90°, lo que indicaba cansancio y por lo tanto el fin de la tarea.

A pesar de que los bio-marcadores para la fatiga muscular en tiempo y frecuencia de la señal sEMG han sido ampliamente estudiados, existen aún limitaciones técnicas que impiden su uso extensivo fuera de ambientes controlados. Por ejemplo, el uso excesivo de cables y conexiones para la medición de la señal sEMG. Sin embargo, novedosos sensores para la medición inalámbrica de la señal eléctrica muscular han sido propuestos en campos como la interacción humano computador (HCI), donde se les conoce como interfaces musculares [18]. Quizá el más reconocido sea el brazalete Myo [19], que es ampliamente usado para el control gestual de dispositivos y periféricos. El uso de estos sensores ha sido limitado a aplicaciones interactivas, y hasta el momento no se encuentra evidencia del uso de sensores *wearables* para monitorear y detectar fatiga muscular en extremidades superiores o inferiores.

El presente estudio propone el uso del sensor Myo para la detección de fatiga muscular en extremidades superiores, específicamente en bíceps braquial de ambos brazos, bajo un protocolo de fatiga durante contracciones isotónicas al 80% del valor 1RM. Se describen las técnicas de procesamiento de señal en tiempo y en frecuencia usadas y se presentan resultados previos.

2 Metodología

2.1 Sujeto

En este estudio participó un atleta amateur masculino de 24 años y 76 Kg. de peso, que fue seleccionado debido a su buena condición física, lo cuál le permite ejercitarse con altas cargas de trabajo.

Figura 1: Asignación de electrodos según Thalmic Labs

El participante es diestro, su participación fue voluntaria y conoció con anterioridad el procedimiento y los posibles riesgos descritos en un consentimiento informado.

2.2 Instrumentación

Las señales sEMG fueron tomadas usando el sensor Myo Armband, el cual tiene ocho electrodos bipolares secos enumerados secuencialmente (ver Figura 1), con una frecuencia de muestreo de 200 Hz cada uno. La recolección de los datos se realizó con el SDK del Myo y el código de Thalmic Labs para la señal EMG [19]. La señal digital fue analizada en Matlab[R]. Durante la prueba se utilizó un goniómetro analógico de 1° de resolución para asegurar el rango de movimiento durante el protocolo. La ejecución se llevó a cabo en un banco predicador de 60° de inclinación.

2.3 Procedimiento experimental

Se llevaron a cabo dos sesiones, en un gimnasio local en ambiente controlado. Durante por lo menos dos días antes de cada sesión el sujeto no realizó ejercicio con cargas mayores al 40% de bíceps braquial. Debido a que es un protocolo de fatiga solo se llevo a cabo una toma de datos en cada sesión.

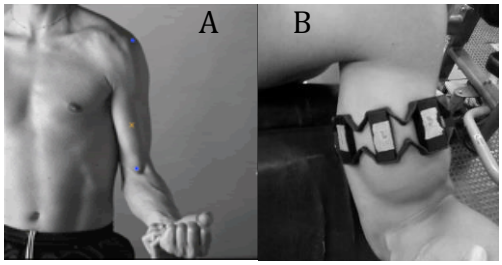


Figura 2: A) Posición de los electrodos según SENIAM. B) Posición del Myo en montaje experimental según SENIAM

En la primera sesión se aplica al sujeto el test 1RM en cada brazo para el bíceps braquial en el banco predicador usando mancuernas, se decidió usar este banco ya que reduce el uso de músculos del hombro y tronco durante la flexión del codo, dejando casi todo el trabajo al bíceps. Se escogió un rango de movimiento de 70° en la flexión del codo, teniendo en cuenta el ángulo en el que debía hacer mayor fuerza sin ayuda de la

inclinación del banco. Al finalizar esta sesión se simula la toma de datos para que el sujeto se familiarice con el protocolo de la siguiente sesión.

En la segunda sesión el protocolo empieza con una serie de ejercicios de calentamiento general de 5 minutos y calentamiento especifico de bíceps braquial de 5 minutos, luego se posiciona el brazalete Myo con el electrodo principal paralelo al eje longitudinal del húmero apuntando a la punta del acromion, de esta manera los electrodos 5 y 6 para el brazo derecho y 2 y 3 para el brazo izquierdo (Figura 3A) quedaron posicionados cada uno sobre una de las ramificaciones del bíceps, como se recomienda en el manual del SENIAM [20] (ver Figura 2).

El goniómetro se adhirió al brazo del sujeto teniendo en cuenta no influir en el buen flujo de la sangre, se ubicó según el estándar: el eje del goniómetro sobre el epicóndilo lateral del húmero, el brazo estático paralelo al eje longitudinal del húmero apuntando al acromion y el brazo móvil paralelo al eje longitudinal del radio (Figura 3B). Al tener todos los dispositivos ubicados correctamente, el sujeto se sentó en el banco predicador a una altura confortable, de esta manera se realizaron flexiones y extensiones del codo repetidas veces con una carga del 80% del valor 1RM (Figura 3C).

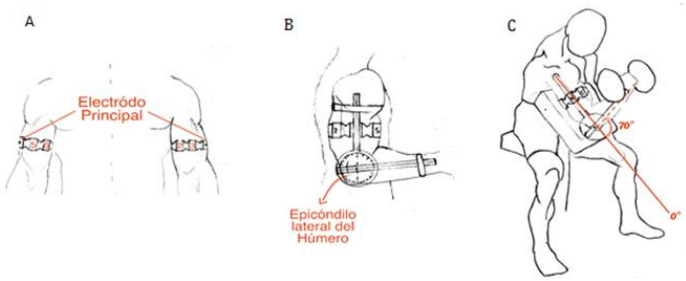


Figura 3: Montaje experimental: A) Posición del sensor Myo en cada brazo; B) Posición del goniómetro sobre el sensor; C) Posición del sujeto en el banco predicador usando los dispositivos de medición

La repetición inició con el codo totalmente flexionado sobre el banco sosteniendo la mancuerna, seguidamente el sujeto hizo una completa extensión del codo e inmediatamente lo flexionó hasta encontrar los 70°. En esta posición el sujeto debe sostener la contracción muscular durante 30 segundos, luego debe realizar una extensión total del codo y nuevamente flexionarlo hasta la misma posición. Las repeticiones terminan cuando no es capaz de repetir la flexión del codo o cuando no es capaz de sostener la carga sin que haya un decremento menor a 10° en el goniómetro. El tiempo fue tomado por el investigador y notificando al sujeto cuando alcanzaba el rango de movimiento de 70°.

3 Procesamiento de la señal

Para el análisis de la señal sEMG se tomaron los valores RMS y MDF. Primero se realizó un pre-procesamiento de la señal removiendo la tendencia restándole a cada dato el valor medio de la serie. Posteriormente la señal fue rectificada. Luego, tanto el valor RMS como el MDF fueron extraídos con ventanas deslizantes. Para el valor RMS se utilizó una ventana con 1s de duración y 15 ms de *overlap*. Para calcular la frecuencia media, primero se obtiene el espectro de densidad de potencia de la curva de frecuencias con la transformada de Fourier, y luego se le aplica una ventana de 15ms de duración y *overlap* de 14 ms. Se decide usar la frecuencia media como característica principal del espectro, porque es menos sensible al ruido, al *aliasing* y en la mayoría de los casos es más sensible a los factores biomecánicos y fisiológicos que ocurren en el músculo [11].

Posteriormente, se calculó una curva de tendencia del desenvolvimiento temporal de cada bio-marcador, esto se realizó mediante el cálculo de la curva de regresión que mejor se ajustara a los datos. La regresión lineal de esta curva presenta dos parámetros: la pendiente *a* y el punto de corte con el eje vertical, el intercepto *b*. Se decide usar esta técnica ya que permite limitar el análisis a épocas, que son intervalos de tiempos determinados. Estos intervalos se escogen de manera que coincidan con los instantes cuasi-isométricos de la contracción muscular, es decir cuando hay contracción sostenida la mayor parte del tiempo, de esta manera se puede extrapolar la interpretación del análisis con base en los resultados de cada época, aunque la contracción sea isotónica por ser repetitiva [11].

4 Resultados

Los dos bio-marcadores, el temporal y el frecuencial, mostraron comportamientos en las direcciones esperadas según la teoría de la medición de la fatiga muscular. En primera instancia, los electrodos ubicados en el bíceps braquial (5 y 6 brazo derecho, 2 y 3 brazo izquierdo) describieron curvas con tendencias positivas (*Canal 5 a= 0.78; Canal 6 = 0.83*) para los valores RMS durante todo el registro de la actividad muscular. También se encontró un comportamiento similar en los otros canales. Las Figuras 4 muestran el comportamiento temporal de los valores RMS en el brazo derecho en los electrodos 5 y 6.

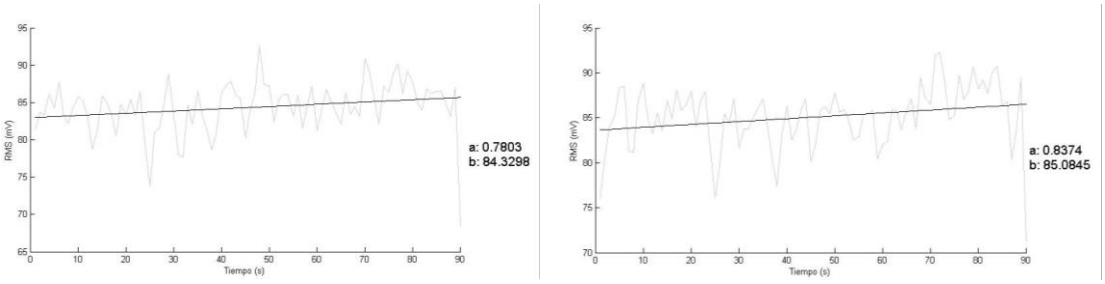


Figura 4: Brazo derecho. A) La curva gris es el valor RMS de cada ventana, la negra es la pendiente para el canal 5; B) la curva gris es el valor RMS de cada ventana, la negra es la pendiente para el canal 6

Este mismo comportamiento incremental fue observado en la Figura 5 para los electrodos del brazo izquierdo, con valores de *a=2.37* en el canal 2 y de *a=1.69* en el canal 3. Los valores de pendiente del brazo izquierdo son mayores al los del brazo derecho, probablemente causados por un alcance de la fatiga con mayor rapidez en el brazo no dominante. Se puede observar que la duración del esfuerzo del brazo izquierdo solamente alcanzo los 50s, mientras que en el derecho fue de 90s. Estos resultados son consistentes con la literatura, pudiendo atribuir este comportamiento a un mayor número de MUAPs activos, es decir un aumento en el reclutamiento de las fibras musculares a lo largo del esfuerzo [2, 8, 9, 21].

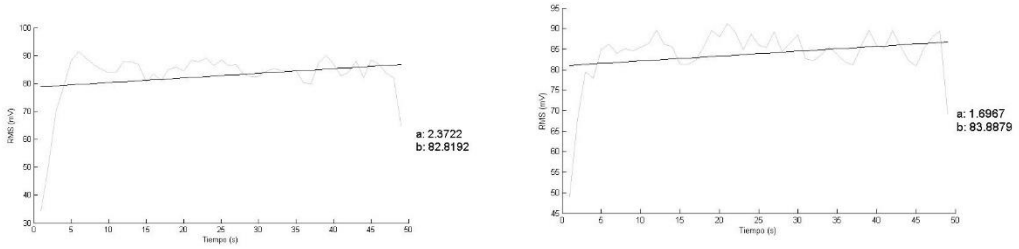


Figura 5: Brazo izquierdo. A) La curva gris representa el valor RMS de cada ventana, la negra es la pendiente para el canal 2. B) La curva gris representa el valor RMS de cada ventana, la negra es la pendiente para el canal 3

Debido al protocolo llevado a cabo, con contracciones sostenidas de larga duración y por el procesamiento de la señal cuasi-isométrica, también se esperaba una modificación del espectro de frecuencia en los mismos electrodos. Las Figuras 6 y 7 muestran las gráficas del valor MDF y su desarrollo en el tiempo, donde para cada brazo y en los canales respectivos, la curva de tendencia muestra un decremento de la amplitud del valor MDF en cada ventana, dando como resultado una pendiente *a* negativa.

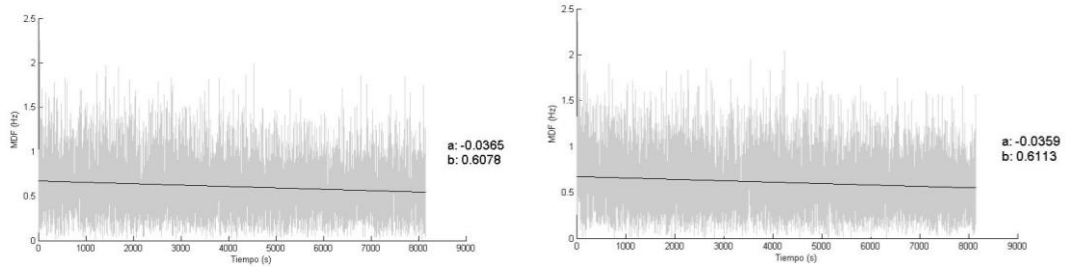


Figura 6: Brazo derecho. A) La curva gris representa el valor MDF de cada ventana, la negra es la pendiente para el canal 5. B) La curva gris representa el valor MDF de cada ventana, la negra es la pendiente para el canal 6

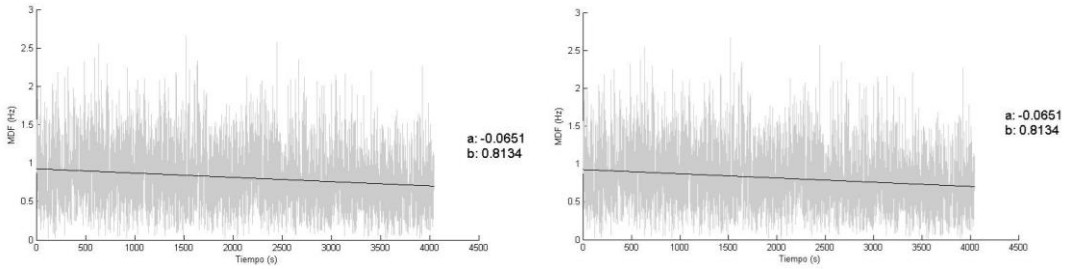


Figura 7: Brazo izquierdo. A) La curva gris representa el valor MDF de cada ventana, la negra es la pendiente para el canal 2. B) La curva gris representa el valor MDF de cada ventana, la negra es la pendiente para el canal 3

Para el brazo derecho la pendiente en el canal 5 tuvo un valor de *a=-0.036* y para el canal 6 un valor de *a=-0.035*, muy cercanos entre sí y comparables (en escala de dimensión) con hallazgos anteriormente publicados [14]. Para el brazo izquierdo los valores de pendiente son más negativos que en el derecho, con *a=-0.065* en el canal 2 y 3. De nuevo, esto puede ser causado porque el esfuerzo hecho por el brazo izquierdo tuvo menor duración que el derecho, es decir, llegó al cansancio más rápido.

5 Conclusiones

Se conoce que la sEMG es una técnica ampliamente usada para la detección de fatiga muscular, teniendo en cuenta que en conjunto depende del protocolo de medición, el procesamiento de la señal y el sensor usado. El uso popularizado de nuevas interfaces musculares como tecnologías interactivas, plantea nuevas posibilidades de aprovechar bit a bit la información que puede ser recolectada de sensores *wearables*. En este caso y según los resultados obtenidos, el sensor sEMG de bajo costo Myo Armband, a pesar de su limitación en frecuencia de muestreo y de ser adecuado solamente para extremidades superiores, prevé como una buena alternativa para la cuantificación de los niveles de fatiga muscular durante una tarea particular. Más allá de esto, el procesamiento de la señal muscular a través de sEMG plantea retos importantes para su uso en aplicaciones en tiempo real, lo que presupone el desarrollo de algoritmos que optimicen el tiempo de respuesta sin sacrificar precisión en la detección.

La detección de la fatiga muscular a través de sistemas de sEMG *wearables* de bajo costo puede ser una pieza clave para el mejoramiento de la prescripción del ejercicio, tanto en personas saludables como en pacientes con múltiples trastornos musculo-esqueléticos [22, 23]. La información del comportamiento eléctrico de los músculos no había estado en manos de sensores no-invasivos, inalámbricos y de fácil acceso.

Los trabajos futuros encierran la validación de esta herramienta con grupos estadísticamente significativos y la inclusión de nuevos sensores de bajo costo, como el sensor Kinect, para la medición de los ángulos de movimiento, en reemplazo del goniómetro.

Referencias

[1] Sanei, S. (2013). Adaptive processing of brain signals. USA: John Wiley & Sons.

[2] De Luca, C. (1979). Physiology and mathematics of myoelectric signals. IEEE Transactions on Biomedical Engineering 26(6), pp. 313-325.

[3] Krupp, L. et al. (1989). The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus erythematosus. Archives of neurology 46(10), pp. 1121-1123.

[4] Criswell, E. (2010). Cram's introduction to surface electromyography. London: Jones & Bartlett Publishers.

[5] Cobb, S. & Forbes, A. (1923). Electromyographic studies of muscular fatigue in man. American Journal of Physiology--Legacy Content 65(2), pp. 234-251.

[6] Edwards, R. & Lippold, O. (1956). The relation between force and integrated electrical activity in fatigued muscle. The Journal of physiology 132(3), pp. 677-681.

[7] Eason, R. (1960). Electromyographic study of local and generalized muscular impairment. Journal of applied physiology 15(3), pp. 479-482.

[8] Vredenbregt, J. & Rau, G. (1973). Surface electromyography in relation to force, muscle length and endurance. In Desmedt, J. (Ed.), New Developments in EMG and Clinical Neurophysiology (pp. 607–622). Berlin: Karger Basel.

[9] Maton, B. (1981). Human motor unit activity during the onset of muscle fatigue in submaximal isometric isotonic contraction. European journal of applied physiology and occupational physiology 46(3), pp. 271-281.

[10] Venugopal, G., Navaneethakrishna, M. & Ramakrishnan, S. (2014). Extraction and analysis of multiple time window features associated with muscle fatigue conditions using sEMG signals. Expert Systems with Applications 41(6), pp. 2652-2659.

[11] De Luca, C. (1997). The use of surface electromyography in biomechanics. Journal of applied biomechanics 13(2), pp. 135-163.

[12] Brody, L. et al. (1991). pH-induced effects on median frequency and conduction velocity of the myoelectric signal. Journal of Applied Physiology 71(5), pp. 1878-1885.

[13] Merletti, R. & Parker, P. (2004) Electromyography: physiology, engineering, and non-invasive applications. USA: John Wiley & Sons.

[14] Sarmiento, J. et al. (2011). Comparative muscle study fatigue with sEMG signals during the isotonic and isometric tasks for diagnostics purposes. Proceedings Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 7163-7166). August 30 – September 3, Boston, USA.

[15] Bonato, P. et al. (2001). Time-frequency parameters of the surface myoelectric signal for assessing muscle fatigue during cyclic dynamic contractions. IEEE Transactions on Biomedical Engineering 48(7), pp. 745-753.

[16] Hernandez, C. et al. (2010). Traditional sEMG fatigue indicators applied to a real-world sport functional activity: Roundhouse kick. Proceedings 20th International Conference on Electronics, Communications and Computer (pp. 154-158). February 22-24, Cholula, México.

[17] Taboadela, C. (2007). Goniometría: una herramienta para la evaluación de las incapacidades laborales. Buenos Aires: Asociart ART.

[18] Allanson, J. & Fairclough, S. (2004). A research agenda for physiological computing. Interacting with computers 16(5), pp. 857-878.

[19] Thalmic Labs. (2014). Myo SDK.

[20] Hermens, H. et al. (1999). European recommendations for surface electromyography. Roessingh Research and Development 8(2), pp. 13-54.

[21] Cifrek, M. et al. (2009). Surface EMG based muscle fatigue evaluation in biomechanics. Clinical Biomechanics 24(4), pp. 327-340.

[22] Rojas, M. et al. (2011). Evaluación de la Función Neuromuscular mediante Electromiografía Multicanal en Procesos de Rehabilitación motora. Revista Iberoamericana de Automática e Informática Industrial 8(2), pp. 35-44.
[23] Lindeman, E. et al. (1999). Surface EMG of proximal leg muscles in neuromuscular patients and in healthy controls - Relations to force and fatigue. Journal of Electromyography and Kinesiology 9(5), pp. 299-307.

Desarrollo de software integrando técnicas de usabilidad y metodologías agiles

Alcides Alvear S.
Graciela Quintero R.
Universidad del Turabo – Puerto Rico

1 Introducción

En la sociedad actual, la ciencia y la tecnología constituyen un poderoso pilar del desarrollo cultural, social y económico. Su influencia en la vida cotidiana es tal que la creciente avalancha de productos procedentes de ambas esferas inundan los ambientes de trabajo, los hogares y la cotidianidad, al punto que hoy en día el manejo sistemático y apropiado de estos productos constituye un importante indicativo de modernidad. Desde esta perspectiva, la sociedad moderna demanda por aplicaciones ágiles y novedosas, en las que la integración de la usabilidad sea un tema relevante para investigadores, creadores y diseñadores de software, como para las personas encargadas de utilizar los sistemas, es decir, los usuarios.

En la era del Internet y la telefonía celular la creciente demanda de pequeños sistemas ha propiciado que los desarrolladores de software utilicen metodologías simples, con requisitos de implementación rápidos, de bajo costo, alternativos y flexibles, que permitan analizar, diseñar, evaluar y administrar software de calidad. En estos proyectos, los requisitos pueden ser introducidos, modificados o eliminados en iteraciones sucesivas, transformando el concepto iterativo en el paradigma principal de todos los procesos ágiles. Como es de esperase, el entorno cambiante en el negocio del software también afecta los procesos de desarrollo, lo que hace que se necesiten herramientas y procedimientos especiales para enfrentar los cambios que se presentarán en el ciclo de vida de un producto. A manera de reflexión este trabajo pretende responder los siguientes interrogantes: ¿Es posible disminuir el tiempo y el costo en el desarrollo de un producto software? ¿Es posible invertir menos tiempo y menos esfuerzo en los procesos de mantenimiento de los programas existentes? ¿Es posible integrar al desarrollo de software las metodologías ágiles y la usabilidad?

2 Ingeniería del Software

La Ingeniería del Software (IS) es una disciplina que se ocupa de los aspectos de la producción de software, desde las primeras etapas de especificación del sistema hasta el mantenimiento después que se haya puesto en uso. Se ocupa de las teorías, métodos y herramientas para el desarrollo profesional de software, teniendo en cuenta el problema a solucionar, las restricciones de desarrollo y los recursos disponibles. A través de IS se busca ofrecer métodos y técnicas para desarrollar y mantener un software con criterios de calidad [1]. Desde la perspectiva de la IEEE [2], consiste en la aplicación de un enfoque sistemático, disciplinado y cuantificable para el desarrollo, operación y mantenimiento del software; metodología con frecuencia denominada modelo de proceso o desarrollo de software, en la que se define quién está haciendo qué, cuándo lo hace y cómo alcanzar los objetivos [3]. IS es una tecnología con varias capas:

1. La primera se basa en un compromiso organizacional continuo con la *calidad*, que permite desarrollar productos cada vez más eficaces. Según la norma ISO-14598 [4] la calidad se refiere a la totalidad de aspectos y características de un producto o

servicio que tienen que ver con su habilidad para satisfacer las necesidades declaradas o implícitas de un usuario. Es decir, no es el objetivo último del producto sino el resultado de la valoración objetiva de cada etapa que satisface las necesidades del cliente [5].

2. La segunda es de *proceso* y define la estructura base para la administración de proyectos software y establece el contexto en el que se aplican los métodos técnicos, se generan los productos del trabajo (modelos, documentos, datos, reportes, formatos, etc.), se establecen los puntos de referencia, se asegura la calidad y se administra el cambio de manera apropiada. En esta capa se reconocen cinco actividades genéricas conocidas como estructurales: a) *comunicación*: consiste en habilitar un canal permanente entre el desarrollador y los usuarios, para conocer las características y funciones del software, saber sus objetivos, sus requerimientos y lo que se espera obtener; b) *planeación*: consiste en una programación de las actividades donde se describe las tareas técnicas, los riesgos probables, los recursos que se requieren y los productos del trabajo que se obtendrán; c) *bosquejo o prototipado*: se usa para ver el panorama general del proyecto, entender mejor los requisitos del software y establecer un diseño tentativo que los satisfaga; d) *construcción:* esta actividad combina la generación de código y las pruebas que se requieren para descubrir posibles errores; y e) *despliegue:* cuando el software parcialmente terminado se entrega al usuario para que lo evalúe y retroalimente. Para muchos proyectos de software las actividades estructurales se aplican en forma iterativa a medida que este avanza. Es decir, se ejecutan a través de cierto número de repeticiones del proyecto, y cada una es una iteración que produce un incremento en la funcionalidad del software.
3. La tercera capa corresponde a los *métodos* y en ella se determina la experiencia técnica para elaborar software, es decir, las herramientas computacionales que indican cómo construir un buen producto. Abarcan un amplio espectro de tareas que incluyen: análisis de requisitos, diseño, construcción de programas y pruebas de mantenimiento.
4. La cuarta es la de *herramientas* de IS, que proporcionan apoyo automatizado o semi-automatizado al proceso y los métodos. Cuando se integran las herramientas de modo que la información creada por una pueda ser utilizada por otra, se establece un sistema conocido como IS asistido por computador, que apoya el desarrollo de gran cantidad de aplicaciones software de carácter moderno.

Como se evidencia, aunque existe una estructura establecida dentro de IS no se puede olvidar que cada producto software necesita un proceso diferente, y estas etapas genéricas pueden organizarse de diferente manera y en diferentes niveles según las necesidades del desarrollador e incluso del propio usuario. Un uso inapropiado del proceso software puede reducir la calidad o la usabilidad del producto e incluso puede incrementar los costos de desarrollo. De ahí la necesidad de idear una serie de modelos, que tengan como finalidad orientar al desarrollador sobre las distintas tareas que se deben cumplir para fabricar un producto de calidad. En la literatura los modelos de proceso también se conocen con el nombre de ciclo de vida del software y comprenden las etapas por las que pasa un proyecto software desde que se concibe hasta que está listo para usarse [6].

Durante varios años uno de los modelos más utilizados para el desarrollo de software fue el modelo tradicional en cascada. Pero contemporáneas a él surgieron otras propuestas como el modelo en V, el incremental, el iterativo, el de prototipos y el modelo

en espiral. Todos con aciertos y desaciertos, sobre todo en lo que respecta a las necesidades cambiantes de los negocios y a la versatilidad que se necesita hoy en día de todas las aplicaciones. Desde este punto de vista, los modelos modernos son evolutivos e iterativos, de manera que permiten desarrollar versiones cada vez más completas de un mismo software. A este grupo de *modelos* pertenece el desarrollo de software basado en componentes, el modelo de métodos formales y el desarrollo de software orientado a objetos. En el primero se construyen aplicaciones a partir de fragmentos de software prefabricados, fomentando el reciclaje cibernético. Esta reutilización reduce el tiempo de desarrollo y disminuye el costo del proyecto. Los modelos formales llevan a la especificación matemática del software, y permiten especificar, desarrollar y verificar un sistema por medio de una notación matemática rigurosa. El modelo orientado a aspectos es un paradigma que proporciona un enfoque metodológico para definir, especificar, diseñar y construir interfaces de usuario, bajo plataformas de trabajo colaborativas.

En este sentido, no se encuentra un modelo único ideal y es claro que en las ultimas décadas lo que se ha observado es un intento por obtener una propuesta integradora que tenga en cuenta las mejores prácticas de los modelos tradicionales, sin dejar de lado los principios del desarrollo ágil de software. El modelo de proceso unificado (RUP) se clasifica dentro de estas propuestas, porque permite realizar trabajos a gran escala mediante un proceso continuo de pruebas y retroalimentación, garantizando la calidad del producto. Con la implementación de (UML), esta propuesta constituye una de las metodologías más utilizadas para el desarrollo de software.

Otro modelo de desarrollo híbrido es el Desarrollo Rápido de Aplicaciones (RAD), que combina el desarrollo iterativo y la construcción de prototipos (en cualquier etapa de desarrollo). Los principios básicos de RAD consisten en realizar entregas rápidas y de calidad a un costo de inversión relativamente bajo; reducir riesgos inherentes al proyecto dividiéndolo en segmentos más pequeños y más fáciles de manejar; y promover la participación intensa de los usuarios y el uso de herramientas de desarrollo computarizadas.

Con lo anterior es posible concluir que no existe una metodología universal para hacer frente con éxito a cualquier proyecto de desarrollo de software, por lo que la IS ha pasado por distintos momentos históricos tratando de dar respuesta a las necesidades propias de cada contexto. Es decir, no ha faltado un modelo que no pretenda ajustarse a las necesidades inherentes e históricas de cada aplicación, de cada desarrollador, e incluso de cada usuario. A cada desarrollador le compete seleccionar el que mejor se ajuste a las necesidades de cada aplicación y de cada usuario en particular.

3 Metodologías ágiles

En la era tecnológica, el internet, la telefonía celular y las redes sociales han cambiado los conceptos de producto y de demanda, revolucionando la forma de hacer negocios. En este contexto surge la Ingeniería de Software ágil como una filosofía orientada al desarrollo de software colaborativo y caracterizado por cuatro aspectos básicos: 1) conformación de grupos de trabajo con privilegios en áreas que anteriormente solamente eran de dominio del desarrollador; 2) reducción de la documentación requerida al mínimo; 3) integración activa de los clientes al grupo de trabajo; y 4) aceptación de la incertidumbre como parte inherente de todo desarrollo de software.

Con esta estrategia de trabajo, tanto el cliente como el desarrollador optan por una perspectiva en común. Por lo tanto, un producto de este trabajo es el incremento en la cantidad de software operativo y funcional que se entrega al cliente a tiempo. En el

contexto de las metodologías ágiles, los requisitos cambian continuamente por lo que se requiere una retroalimentación sobre resultados obtenidos para determinar si es lo requerido o no, es decir, se va planeando a medida que se avanza en el proyecto, y por tanto la planeación es gradual y constante. A lo largo de todo el proceso de desarrollo permanecen las actividades estructurales fundamentales de comunicación, planeación, modelado, construcción y despliegue, como etapas recicladas de los modelos iniciales [7].

Históricamente, desde la publicación del manifiesto ágil, ha pasado más de una década en la que la propuesta ha pasado por distintas etapas: los primeros años sirvieron de base para la adopción de los métodos ágiles [8, 9] y para entender las características del desarrollo de software desde esta perspectiva [10-12]. En años más recientes, el enfoque ha sido sobretodo administrativo y ha incursionado en la dinámica del trabajo en equipo, el juego de roles, la auto-organización y las estrategias de comunicación grupal para la eficiencia [13]. La adopción y aceptación de las metodologías ágiles ha sido clave para entender las nuevas perspectivas [14, 15] y los retos de su implementación en diversos entornos [16].

Las primeras propuestas de desarrollo ágil se caracterizaron por afrontar con distintas ópticas los principios básicos del manifiesto ágil y con esa idea surgieron propuesta tales como XP, Scrum, Crystal, LSD y FDD. A continuación se enfatizará en cuatro metodologías agiles cuyos principios y estrategias serán utilizadas para construir el marco de integración que hace parte de la fase experimental de este trabajo.

3.1 Programación eXtrema (XP)

Este método, más conocido como XP [17, 18], ha sido señalado ampliamente como el punto de partida de los distintos enfoques de desarrollo de software ágil. XP se enfoca en potenciar las relaciones interpersonales entre el equipo de trabajo y el cliente a través de historias de usuario dinámicas y flexibles. Mediante esta estrategia y con la ayuda de fichas, el cliente describe brevemente las características que el sistema debe poseer. Cada historia de usuario se descompone en tareas de programación de forma suficientemente comprensible y delimitada para que un desarrollador pueda implementarlas en poco tiempo, es decir, durante una iteración [19].

En todas las iteraciones del ciclo el grupo de trabajo aprende y se asegura de que el sistema tenga la mayor calidad posible. El principal reto de XP es disminuir la curva exponencial del costo a lo largo del proyecto, como para que el diseño evolutivo funcione. Esto se consigue mediante la utilización de tecnologías disponibles que, aunque no son novedosas, se integran de forma efectiva y complementada con otros fundamentos teóricos del campo de los negocios, los valores humanos y el trabajo en equipo, dando vida a una metodología adaptada para la modernidad [19]. Las prácticas de XP incluyen *planning game*, pequeñas entregas, diseño simple, programación en pareja, pruebas, *refactoring*, integración continua, propiedad común del código, paso sostenible, cliente en sitio, metáforas y estándares de código.

3.2 SCRUM

El nombre de esta metodología proviene de un concepto deportivo, propio del rugby, relacionado con la formación utilizada para la recuperación rápida del juego ante una infracción menor [20]. El modelo fue identificado y definido por Nonaka y Takeuchi a principios de los 80, al analizar cómo desarrollaban los nuevos productos las principales empresas de manufactura tecnológica: Fuji-Xerox, Canon, Honda, 3M y Hewlett-Packard [21, 22]. Por lo tanto, inicialmente Scrum surgió como un modelo para el desarrollo de

productos tecnológicos. No obstante, hoy en día se emplea en entornos que trabajan con requisitos inestables y que requieren rapidez y flexibilidad. Para el desarrollo de software la metodología propone un estructura básica para su funcionamiento, partiendo de equipos en proyectos pequeños con diferentes roles que inician un *juego* estructurado mediante un conjunto de reglas y artefactos. Scrum utiliza un enfoque incremental y empírico que garantiza la visibilidad de las cosas que pueden afectar el resultado del proceso; una inspección que ayuda a detectar variaciones indeseables en el proceso; y una adaptación que realiza los ajustes pertinentes para minimizar su impacto [23].

Los equipos Scrum son auto-gestionados, multifuncionales y trabajan en iteraciones. Cada iteración crea nuevas funcionalidades o modifica las que el dueño del producto requiera. Scrum define tres roles: 1) ScrumMaster, que procura facilitar su aplicación y gestionar los cambios, 2) ProductOwner, que representa a las partes interesadas, y 3) Team, que ejecuta el desarrollo y los demás elementos relacionados. Cada periodo de trabajo es conocido como *sprint*, y básicamente es un periodo que toma entre una y cuatro semanas para que el equipo cree un incremento de software potencialmente entregable (utilizable). Las características que forman cada *sprint* provienen del Product Backlog, un conjunto de requisitos de alto nivel priorizados que definen el trabajo a realizar.

Los elementos del Product Backlog que forman parte del *sprint* se determinan durante la reunión de Sprint Planning. Durante esta reunión se planea qué se va a entregar y cómo se hará, es decir, el diseño del sistema y la estimación de cantidad de trabajo. Esta actividad dura ocho horas para un *sprint* de un mes. Si tiene una duración menor se asigna el tiempo de manera proporcional. El Daily Scrum es un evento que e toma al equipo de desarrollo alrededor de 15 minutos, y se realiza cada día con el fin de explicar lo que se ha alcanzado desde la última reunión, lo que se hará antes de la siguiente y los obstáculos que se han presentado. El Monitoreo de Progreso consiste en la suma del trabajo que falta por realizar en el *sprint*. Tiene como característica que se puede dar en cualquier momento, lo que le permite al dueño del producto evaluar el progreso del desarrollo.

Para que esto sea posible, los integrantes del equipo actualizan constantemente el estado de los requisitos que tienen asignados, indicando cuánto consideran que les falta por terminar. El ciclo de vida de este marco de trabajo se compone de cuatro fases: planeación, puesta en escena, desarrollo y entrega. En la planeación se establece la visión, se fijan las expectativas y se asegura el financiamiento. En la puesta en escena se identifican más requisitos y se priorizan para la primera iteración. En la implementación se desarrolla el sistema y en la entrega se hace el despliegue operativo.

3.3 Agil Unified Process (AUP)

AUP es una metodología utilizada comúnmente para el desarrollo de software en el área de los negocios que utiliza técnicas y conceptos afines a la metodología RUP, lo que la convierte en una versión simplificada de esta [24]. Se considera como un marco de trabajo extensible que puede ser adaptado a organizaciones o proyectos específicos, y se caracteriza por la utilización de los casos de uso, porque se centra en la arquitectura y porque es iterativa e incremental. AUP, es una metodología que adopta muchas técnicas ágiles utilizadas en XP y las formalidades de RUP, y su objetivo es adaptarse a las necesidades del proyecto, contrario a la filosofía de las metodologías tradicionales.

Uno de los principales planteamientos de AUP es su composición en dos partes: fases y disciplinas. Las fases son cuatro: 1) *concepción*, cuyo objetivo es obtener una comprensión común cliente-equipo de desarrollo del alcance del nuevo sistema y definir una o varias arquitecturas candidatas para el mismo; 2) *elaboración*, cuyo objetivo es

lograr que el equipo de desarrollo profundice en la comprensión de los requisitos del sistema y en validar la arquitectura; 3) *construcción*, en la que el sistema es desarrollado y probado completamente dentro del ambiente de desarrollo; 4*) transición*, donde el sistema es llevado a entornos de preproducción y se somete a pruebas de validación y aceptación, y finalmente se despliega en los sistemas de producción. En AUP se considera el modelado, la implementación, las pruebas, el despliegue, la gestión de configuración, la administración del proyecto y el entorno, como etapas suficientes para hacer de esta metodología de fácil y rápido entendimiento por todo el grupo de desarrollo. Además, utiliza técnicas ágiles tales como *test driven development* (TDD], *agile model driven development* (AMDD), *agile change management* y *database refactoring*.

3.4 Lean Software Development (LSD)

LSD es una adaptación al desarrollo ágil del Lean Manufacturing de Toyota. Es una metodología que tiene como objetivo eliminar desperdicios seleccionando las características que realmente aportan valor, velocidad y eficiencia a un producto software [25]. La metodología incluye siete principios importantes:

1. *Eliminación de desperdicios*. Consiste en evitar todo aquello que no agregue valor al proyecto y sobre todo al cliente. Un desperdicio es cualquier actividad que puede ser excluida o dar el mismo resultado si no se llega a hacer; son procesos que finalmente no son usados por el cliente pero que pueden incidir en la calidad del producto.
2. *Ampliar aprendizaje*. En este caso consiste en aprender de las experiencias, es decir, es mejor probar pronto un producto antes que acumular defectos que luego hagan más difícil mejorar las características. En resumen, se trata de llevar a cabo muchos experimentos rápidos, crear documentación concisa y aplicar la mejor alternativa.
3. *Decidir lo más tarde posible*. Todo desarrollo tiene asociada una incertidumbre y los mejores resultados se alcanzan con un enfoque basado en opciones. Por lo que se pueden retrasar las decisiones tanto como sea posible hasta que se basen en hechos y no en suposiciones y pronósticos inciertos. Las iteraciones promueven este principio adaptándose a los cambios y corrigiendo los errores, porque corregir un error puede resultar costoso si se detecta demasiado tarde.
4. *Reaccionar tan rápido como sea posible*. Consiste en reducir el tiempo de las iteraciones, porque cuanto más cortas sean mejor es el aprendizaje y la comunicación del equipo. Cuanto antes se entrega un producto final más pronto se reciben propuestas para incorporarlas en la siguiente iteración.
5. *Potenciar el equipo*. Se basa en el compromiso del grupo, que no debe ser numeroso porque dificulta la comunicación; se recomienda seleccionar un líder para la gestión; el grupo debe tener una misión clara y una meta a corto plazo; y los participantes deben reunir capacidades necesarias para desarrollar.
6. *Crear integridad*. Esto significa que los conceptos del sistema trabajan como una totalidad armónica de arquitectura coherente.
7. *Ver el todo como un conjunto*. Porque pensar en el todo es más importante que las mediciones de desempeño individuales. Un criterio globalizado ayuda a perfeccionar y corregir un área especifica percatándose del conjunto total del producto.

4 Usabilidad e integración persona-computador

De acuerdo con ACM, esta es una disciplina relacionada con el diseño, evaluación e implementación de sistemas informáticos interactivos para el uso de seres humanos, y su

importancia radica en que mejora el uso del computador como herramienta de trabajo, ocio y aprendizaje. A través de ella se busca incrementar la satisfacción de los usuarios y reducir su esfuerzo para construir tareas para computador, sin que esto disminuya su capacidad para desarrollar actividades y procesos interactivos con el sistema.

Según este principio, la interfaz es el punto en el que seres humanos y computadores se ponen en contacto, y se transfieren mutuamente información, órdenes y datos, tales como sensaciones, intuiciones y nuevas formas de ver las cosas. Para el caso, si la interfaz presenta alguna limitación, esto afectará directamente al usuario porque lo que no es posible expresar a través de ella permanecerá por fuera del espacio relacional. Desde esta perspectiva, aunque se tenga en cuenta las cualidades del usuario promedio, en muchos casos la interfaz se convierte en una barrera debido a un deficiente proceso de elaboración del software y a una escasa atención a los detalles de la tarea a realizar. Como resultado, aunque el diseño no estará completamente libre de barreras, esta integración trata de minimizar la situación colocando sus procesos y herramientas al servicio del usuario [26].

A finales del siglo XX Diaper indicaba que la usabilidad constituía una de las principales preocupaciones de estas herramientas, porque a través de ellas se integraban no solamente los factores relacionados con la IS, si no los aspectos sociológicos, psicológicos, ergonómicos, culturales y sociales que debían interactuar para producir un software de calidad. En la misma línea de pensamiento, Bevan afirma que estas técnicas tienen como objetivo incrementar el nivel de usabilidad a través de todo el proceso de creación del software, porque de esta manera se logra que dicho producto sea comprendido, aprendido, usado y que a la vez sea atractivo para el usuario. Esta definición hace énfasis en los atributos internos y externos del producto, los cuales contribuyen a su funcionalidad y eficiencia [27].

No obstante, aunque los autores coinciden en la importancia de la usabilidad dentro de la IS, todavía existe un amplio desconocimiento entre los desarrolladores de software sobre estas técnicas, y muchos no tienen una percepción clara de cómo y cuándo aplicarlas. Además, no les resulta cómodo hablar de usabilidad en el campo de IS. De hecho, para otros la usabilidad solamente tiene sentido en la fase funcional, cuando la parte interna del sistema ya ha sido diseñada [28]. Para algunos autores los procesos de estas técnicas se centran en tecnologías blandas, es decir, se basan en temas humanos. Por ejemplo, se enfocan en el diseño, evaluación e implementación de sistemas computacionales interactivos y en la forma en que un producto se comporta y es usado en el mundo real. En muchos casos se considera blando a los conocimientos tecnológicos de tipo organizacional, administrativo y de comercialización, pues se trata de información no necesariamente tangible [29].

Esta dicotomía entre pensamiento y ejecución es lo que ha mantenido separados por tanto tiempo procesos que desde el sentido común debían permanecer unidos. Por lo tanto, buscar puntos de encuentro entre la IS y estas técnicas ha sido el principal objetivo de muchas organizaciones de desarrollo de software, las cuales quieren incrementar el nivel de usabilidad de sus productos, pero que no están dispuestas a cambiar completamente su proceso de desarrollo [28]. Por otra parte, las nuevas tecnologías también demandan diseños de sistemas interactivos, donde el usuario se transforma en parte esencial del proceso y en ciertos momentos del equipo de trabajo. Con estas tecnologías también emergen nuevas necesidades, obligando a la IS a darle importancia a la usabilidad, pues de alguna manera la masificación de los productos software implica trabajar en pro de un grupo cada vez más grande de usuarios, con un rango cada vez más amplio de necesidades.

5 Marco de integración

En este trabajo se propone integrar las metodologías ágiles y la usabilidad al desarrollo de Software, para averiguar si es posible disminuir el tiempo y el costo de desarrollo de un producto. La hipótesis central se basa en que al integrar las técnicas de usabilidad al modelo estándar de IS y al aplicar metodologías ágiles en las diferentes capas del proceso, se puede realizar una aplicación en menor tiempo, utilizando un sin numero de herramientas automatizadas de libre costo que simplifican la documentación y aceleran los procesos, sin alterar los requisitos de calidad del producto. El marco de integración propuesto se muestra en la Figura 1.

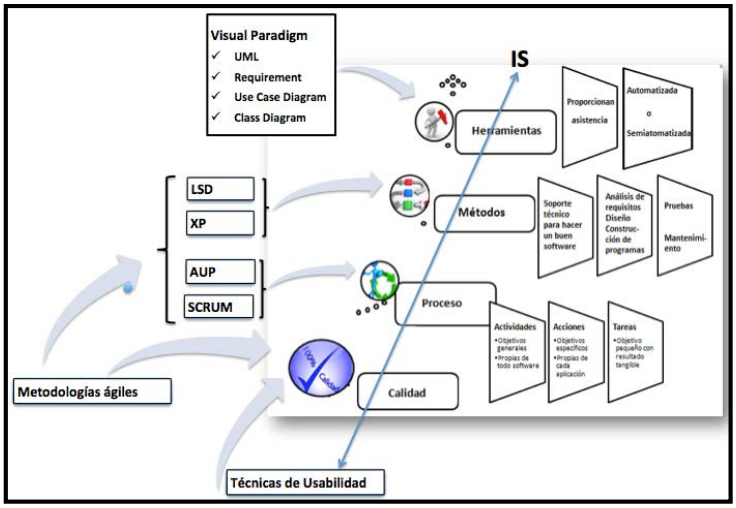


Figura 1: Marco de integración de las metodologías ágiles y la usabilidad a la IS

Cada una de las capas de IS esta cruzada por un eje transversal que propone integrar las técnicas de usabilidad para aumentar significativamente la calidad del producto. Por su parte, las metodologías ágiles (SCRUM, AUP, XP y LSD), se integran en distintos momentos del proceso, con el fin de potenciar los principios de usabilidad. Por ejemplo, teniendo en cuenta a SCRUM se adopta una estrategia de desarrollo incremental, se tiene en cuenta el conocimiento tácito de las personas y se permite el solapamiento de las diferentes fases del desarrollo, siendo una buena alternativa para administrar cualquier proceso software. Por su parte, la formalidad del proceso unificado de AUP puede ser útil para gestionar los riesgos inherentes a todo proceso de desarrollo software. AUP propone que aquellos elementos con alto riesgo tengan prioridad en el proceso de desarrollo y que sean abordados en etapas tempranas del mismo. Mediante el desarrollo de prototipos ejecutables se demuestra la validez de la arquitectura para los requisitos clave del producto, aspecto que minimiza el tiempo de desarrollo y asegura la calidad del producto final.

Desde esta propuesta se propone integrar XP, porque esta metodología se centra en potenciar las relaciones interpersonales como clave para el éxito en el desarrollo de software, y promueve el trabajo en equipo y se preocupa por el aprendizaje de los desarrolladores. XP se basa en la realimentación continua entre el cliente y el equipo de desarrollo, la comunicación fluida entre todos los participantes, la simplicidad en las soluciones implementadas y el coraje para enfrentar los cambios. También se define especialmente adecuada para proyectos con requisitos imprecisos y muy cambiantes, donde existe un alto riesgo técnico. El marco de integración no quedaría completo si no se tienen en cuenta las bondades de la metodología experimental LSD. Esta metodología

facilita el control de calidad del producto a partir de la eliminación del desperdicio, es decir, lo que no añade valor para el cliente puede obstaculizar el avance de un proceso, y se debe considerar como un desperdicio.

Finalmente, y para completar el marco de integración, se propone el uso de herramientas automatizadas o semi-automatizadas tales como Visual Paradigm, que se puede utilizar a lo largo de todo el ciclo de vida del producto. A través de un enfoque orientado a objetos esta herramienta permite utilizar diferentes tipos de diagramas, entre los que se encuentran UML, SysML y ERD, facilitando el proceso de documentación, y en algunos casos ayudando incluso a generar código fuente en distintos lenguajes de programación. Esto hace más fácil el proceso de análisis y diseño de un producto software.

Hoy en día, tanto las comunidades que trabajan en usabilidad e integración persona-computador como la de IS juegan un rol importante en el proceso del desarrollo de sistemas. Las primeras tienen el conocimiento sobre cuáles son las características que se le deben proporcionar al sistema para hacerlo usable, y las otras sobre el desarrollo en sí del mismo. La interacción humano-computador ha desarrollado en las últimas décadas una variedad de técnicas para mejorar la usabilidad de los sistemas, pero todavía no se han extendida lo suficiente en el campo de la IS [30]. A su vez, los desarrolladores de software solamente reciben capacitación básica en usabilidad [31], por lo que normalmente no tienen el conocimiento necesario para construir software usable.

Aunque se tiene conocimiento de más de 50 de estas técnicas, únicamente 35 les permiten a los ingenieros de software con poca experiencia en el tratamiento de la usabilidad, utilizar sus bondades para realizar aplicaciones más eficientes. Estas técnicas fueron organizadas por Ferré [51], constituyéndose en el primer acercamiento e interconexión entre IS y usabilidad. Estos investigadores recopilaron diversas técnicas de usabilidad y crearon una nomenclatura en la que se contemplan tres niveles de uso (alto, medio y bajo) y tres aspectos denotados por una letra mayúscula: (F) Formación, (A) Aplicabilidad y (AP) Aportación, de cada técnica al desarrollo de software. De este criterio surge la taxonomía que se muestra en la Figura 2.

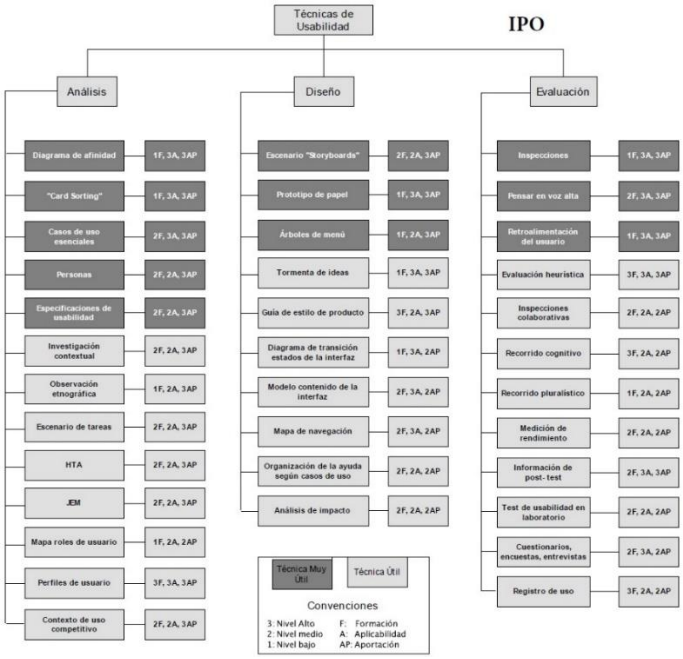


Figura 2: Taxonomía de técnicas de usabilidad de acuerdo con el criterio de uso [51]

A pesar del creciente interés de las organizaciones por la aplicación de métodos y técnicas de usabilidad en el proceso de desarrollo de software, seleccionar y poner en práctica las que más se adecuen a las características de un determinado proyecto es una tarea compleja que tiene poco soporte. Para facilitar el proceso de inclusión, Ferre y Baver [32] diseñaron Usability Planner, una herramienta web para la planificación de técnicas de usabilidad en el ciclo de vida del software. Su objetivo es dar soporte a la selección de técnicas y métodos de usabilidad en el proceso de desarrollo minimizando riesgos y maximizando beneficios. Por otra parte, para proponer una integración de la usabilidad a la IS, no es suficiente con asignar técnicas de usabilidad a actividades de desarrollo, puesto que no todas son aplicables en cualquier momento de un desarrollo iterativo. La idea principal detrás de los procesos ágiles es el mejoramiento iterativo e incremental, que le permita al desarrollador sacar ventaja de lo que ha aprendido en las primeras entregas y le ayude a avanzar en el proceso, para entregar versiones cada vez más ajustadas a las necesidades del cliente. A partir de esta idea, en la Figura 3 se propone una reestructuración del ciclo de vida de un proyecto software, para que de forma deliberada y sistémica se puedan incluir las cuatro metodologías ágiles propuestas en la integración.

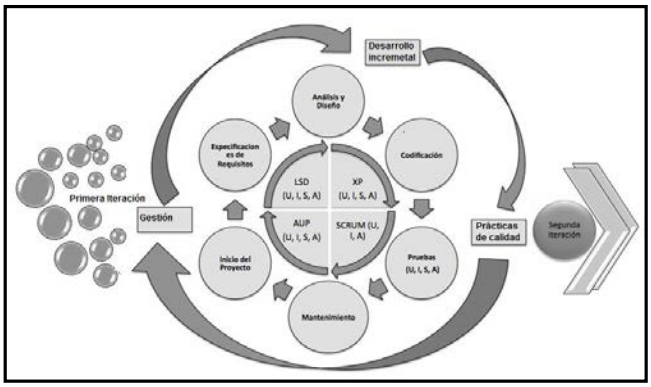


Figura 3: Integración de las metodologías ágiles al ciclo de vida de un producto software

De la Figura 3 se analiza el compromiso constante de la propuesta con la gestión administrativa del proyecto, con el fomento de un desarrollo incremental y con la calidad de cada iteración, lo cual redundará en la calidad final del producto. El diagrama también resalta la importancia de las pruebas (U: unitarias, I: de integración, S: del sistemas y A: de aceptación), pues constituyen la manera formal de asegurarla, y propone un mecanismo de ahorro de tiempo con el uso de las distintas metodologías agiles.

6 Resultados. Implementación del marco ágil

Con el fin de comprobar las bondades del marco ágil propuesto, se diseñó un producto software que buscaba facilitar el conteo de esporas fúngicas presentes en la atmósfera y el monitoreo móvil del mismo a través de *smartphones* [35]. Este tipo de estudio hace parte de una rama del conocimiento científico denominada aerobiología, que se define como la ecología de la atmósfera. Es la disciplina que se ocupa del estudio de los organismos vivos aerotransportados, determina su diversidad, analiza sus modos de vida y describe los procedimientos competentes para la recuperación del entorno [33]. En este contexto es que surge la necesidad de construir una aplicación cuyo primer objetivo era facilitar el conteo de esporas, atendiendo a la clase y a la ocurrencia de las mismas en la atmósfera. Estudios de este tipo son necesarios desde el punto de vista médico, porque se ha comprobado que la densidad y diversidad de este particulado fúngico puede alterar las vías respiratorias de personas con problemas respiratorios.

Para el caso de estudio, el cliente es un grupo de investigadores del Recinto de Ciencias Médicas de la Universidad de Puerto Rico, cuya función consiste en recopilar y evaluar la información aerobiológica para luego enviarla a la Academia Americana de Asma Alergia e Inmunología; actividad que diariamente les consumía un promedio de seis horas de trabajo. Sabiendo que un proceso automatizado podría disminuir ese tiempo y a la vez facilitar la integración de otras variables de interés como las meteorológicas, se propuso la creación de un software de esporas llamado posteriormente Spore Count [34]. La primera versión de este producto fue realizada en 23 semanas, sin el uso del marco ágil propuesto en este trabajo. Otras características relevantes a esta versión son: grupo de desarrollo, dos personas; metodología utilizada: XP; herramientas de ingeniería del software utilizada: UML de forma muy general.

Después de varias reuniones con los clientes para conocer a fondo sus necesidades y con ayuda de la herramienta Visual Paradigm, se crearon las primeras historias de usuario, que sirvieron de insumo para seleccionar las técnicas de usabilidad con inclusión que mejor se ajustaban a los requisitos de Spore Count. Con Usability Planner se construyó la estructura de desarrollo cuya síntesis se muestra en la Figura 4.

Para llevar a cabo la etapa de análisis se realizaron dos visitas de aproximadamente dos horas cada una al sitio de trabajo de los usuarios finales. Después de procesar la información obtenida se incluyeron dos grandes actividades*: contexto de uso* y *especificaciones de usabilidad*, identificadas en la parte inferior de la figura con los números 1 y 2 respectivamente. A su vez, para el *contexto de uso* se tuvo en cuenta tres aspectos: 1) *usuarios*, incluyendo sus conocimientos, experiencia, actitudes, aptitudes y necesidades; 2) *tareas*, con el fin de conocer qué hace el usuario y qué dificultades tiene; y 3) *entorno*, desde las perspectivas social-laboral, físico-emocional y locativa.

Para las *especificaciones de usabilidad* se propuso que cada usuario evaluara cuantitativamente el grado de satisfacción alcanzado durante cada iteración y a su vez la calidad y aplicabilidad de cada entrega. Estas actividades podían a su vez ser fácilmente realizadas con la ayuda de las cinco técnicas de usabilidad señaladas en la parte superior de la figura: *Casos de uso esenciales, personas, especificaciones de usabilidad, escenario de tareas y perfiles de usuario.*

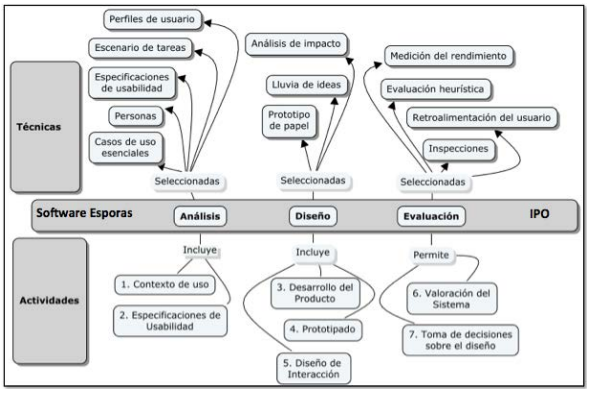


Figura 4: Esquema de integración de las técnicas y actividades de usabilidad para Spore Count

En el área de *diseño* se incluyeron tres grandes actividades: *desarrollo del producto, prototipado y diseño de interacción*, identificadas en la figura con los números 3, 4 y 5 respectivamente. Para el *desarrollo del producto* se tuvo en cuenta tres aspectos: 1) modelos mentales del usuario, para especular sobre cómo funciona el sistema; 2) modelos mentales del diseñador, para tener una imagen lógica del funcionamiento del sistema; y 3) el concepto del producto, que permite la definición clara del mismo.

El *Prototipado* permitió analizar la relación entre diseñador/usuario, usuario/sistema, recursos/tiempo (costo) y todo/parte, y para ver el comportamiento del sistema como un conjunto de partes (sub-módulos) al mismo tiempo cada una de forma independiente. El *diseño de interacción* permitió conocer los entornos de interacción usuario/sistema (parte del prototipado) y la interfaz gráfica.

Para materializar esta etapa de diseño se hicieron cuatro visitas de hora y media cada una al entorno de trabajo de los usuarios, y se mantuvo comunicación con los involucrados según fuera requerido. Se aplicaron tres técnicas de usabilidad: *prototipo de papel, lluvia de ideas y análisis de impacto,* fundamentadas en heurísticas que incluyen el entorno y el contexto de desarrollo y teniendo en cuenta los lineamientos de las cuatro metodologías seleccionadas en el marco ágil de trabajo. Finalmente, en el área de *evaluación* se incluyeron dos grandes actividades: *valoración del sistema y toma de decisiones sobre el producto*, identificadas en la Figura 3 con los números 6 y 7 respectivamente:

La *valoración del sistema* se utiliza para saber si satisface las necesidades del usuario y si encaja en el contexto de uso. La *toma de decisiones sobre el diseño* ayudó a retroalimentar el proceso, a evaluar los objetivos alcanzados por el usuario y a monitorear el uso del sistema a largo plazo. Para realizar esta etapa se hicieron dos visitas, la primera de aproximadamente tres horas, donde se realizaron las instalaciones, tanto de la aplicación como del software necesario para su funcionamiento. Además, se realizaron las primeras pruebas del sistema. La segunda visita fue de cuatro horas aproximadamente y en ella se realizaron las pruebas faltantes, Se utilizaron cuatro técnicas de usabilidad: *inspecciones, retroalimentación del usuario, evaluación heurística y medición del rendimiento;* todo esto fundamentado en heurísticas que incluyen al entorno y el contexto de desarrollo, y teniendo en cuenta los lineamientos de las metodologías XP, AUP y LSD.

Para responder el interrogante sobre el consumo de tiempo empleado en el diseño de la aplicación Spore count se presenta la Figura 5, donde se visualiza el diagrama de actividades con el tiempo empleado en cada iteración.

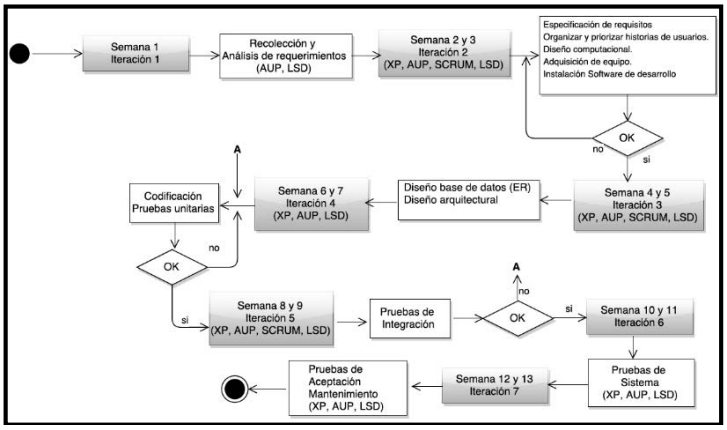


Figura 5: Diagrama de proceso e iteraciones durante el ciclo de desarrollo del software de esporas

La Figura 5 muestra las siete iteraciones realizadas durante el ciclo de desarrollo completo para la aplicación de esporas. La primera iteración fue de una semana y las restantes de dos semanas; en cada iteración se muestra el uso de las metodologías ágiles empleadas en cada una de ellas, hasta la entrega final del producto software. El marco ágil propuesto también fue utilizado en dos aplicaciones más, obteniendo resultados similares a los descritos anteriormente: 1) aplicación web para el proceso de admisiones a la Escuela Graduada de la Universidad del Turabo y 2) aplicación de Realidad virtual para el

tratamiento de fobias a las alturas [36] utilizada por el Departamento de Psicología de la Universidad del Turabo.

7 Conclusiones

La solución propuesta está dirigida a una gran variedad de proyectos y organizaciones dedicadas a la creación de software que cumplan con dos características fundamentales: 1) ciclo de vida basado en los principios de la IS y 2) poco conocimiento del equipo de desarrollo de conceptos y técnicas de usabilidad con inclusión. De esta manera, no solamente se involucran pequeñas y grandes compañías de software o desarrolladores independientes, si no también el desarrollo que se pueda crear desde la perspectiva de lo académico.

El marco ágil de trabajo propuesto permite una considerable disminución del tiempo de desarrollo. Por ejemplo, la primera versión del software de esporas (sin el uso del marco ágil propuesto), necesitó de 23 semanas, mientras que la nueva versión con el uso del marco ágil propuesto necesitó 13 semanas, lo que significa también una disminución en los costos del producto.

Tanto las técnicas y actividades de usabilidad, junto con el uso adecuado de las metodologías ágiles seleccionadas en el marco de trabajo propuesto, no solamente permitieron la disminución del tiempo, sino también producir un software de mejor calidad.

El marco propuesto es uno atractivo para cualquier desarrollador, porque no es necesario desechar el proceso de desarrollo actual o existente, sino que basta con incluir las técnicas de usabilidad, según el marco de integración, permitiendo además la inclusión de las bondades de las metodologías ágiles seleccionadas durante el ciclo de vida, según sea el proyecto que se esté desarrollando.

Las metodologías ágiles seleccionadas tienen varios años de uso en el desarrollo de productos software, se encuentran activas y cumplen con prácticas y principios de usabilidad y respecto a la calidad, tales como: fiabilidad, simplicidad, trazabilidad, usabilidad, claridad, precisión, mantenibilidad, modularidad y portabilidad, Haciendo de estos principios el fundamento que se persigue con el marco ágil propuesto, y cuya finalidad es obtener productos software de calidad, bajo costo y corto tiempo de producción.

Referencias

[1] Boehm, B. (2006). A view of 20th and 21st century Software Engineering. Proceedings 28th international conference on Software engineering (pp. 12-29). May 20-28, Shanghai, China.

[2] IEEE. (1990). IEEE standard glossary of software engineering terminology. Std 610.12-1990.

[3] Jacobson, F., Booch, G. & Rumbaugh, J. (1999). The Unified Software Development Process. USA: Wesley.

[4] ISO. (2014). Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Planning and management. ISO/IEC 25001:2014.

[5] Dávila, A., Melendez, K. & Flores, L. (2006). Determinación de los requerimientos de calidad del producto software basados en normas internacionales. IEEE Latin America Transactions 4(2), pp. 100-106.

[6] Oktaba, H. et al (2007). Software process improvement: The competisoft project. IEEE Computer 40(10), pp. 21-28.

[7] Pressman, R. (2010). Ingeniería del Software: Un enfoque práctico. México: McGraw Hill.

[8] Boehm, B. (2002). Get ready for agile methods, with care. Computer 35(1), pp. 64-69.

[9] Nerur, S., Mahapatra, R. & Mangalaraj, G. (2005). Challenges of migrating to agile methodologies. Communications of the ACM 48(5), pp. 72-78.

[10] Nawrocki, J., Walter, B., Wojciechowski, A. (2001). Toward maturity model for extreme programming. Proceedings 27th Euromicro Conference (pp. 233-239). September 4-6, Warsaw, Poland.

[11] Nawrocki, J. & Wojciechowski, A. (2001). Experimental evaluation of pair programming. In: Maxwell, K., Oligny, S., Kusters, R. & van Veenedaal, E. (Eds.), Project Control: Satisfying the Customer (269–276). London: Shaker Publishing.
[12] Williams, L. (2010). Agile software development methodologies and practices. Advances in Computers 80, pp. 1-44.
[13] Moe, N., Dingsøyr, T. & Røyrvik, E. (2009). Putting agile teamwork to the test – An preliminary instrument for empirically assessing and improving agile software development. Lecture Notes in Business Information Processing 31, pp 114-123.
[14] Cao, L., Mohan, K., Xu, P. & Ramesh, B. (2009). A framework for adapting agile development methodologies. European Journal of Information Systems 18, pp. 332–343.
[15] Mangalaraj, G., Mahapatra, R. & Nerur, S. (2009). Acceptance of software process innovations – the case of extreme programming. European Journal of Information Systems 18, 344–354.
[16] Ramesh, B., Cao, L., Mohan, K. & Xu, P. (2006). Can distributed software development be agile? Communications of ACM 49 (10), pp. 41-46.
[17] Beck, K. (1999). Extreme programming explained: Embrace change. USA: Addison-Wesley.
[18] Beck, K. (1999). Extreme programming: A discipline of software development. Lecture Notes in Computer Science 1687, pp. 1.
[19] Wake, W. (2002). Extreme programming explored. New York: Addison-Wesley Professional.
[20] Mushtaq, Z. & Qureshi, M. (2012). Novel hybrid model: Integrating Scrum and XP. International Journal of Information Technology and Computer Science 6, pp. 39-44.
[21] Takeuchi, H. & Nonaka, I. (1986). The new product development game. Harvard Business Review January-February, pp. 137-146.
[22] Schwaber, K. (1997). SCRUM development process. Business Object Design and Implementation 1997, pp. 117-134.
[23] Alliance, S. (2012). Scrum basics. Online.
[24] Nuñez, J. (2010). Usabilidad en metodologías ágiles, en informática. Tesis de master en ingenieria del software. Universidad Politecnica de Madrid.
[25] Poppendieck, M. & Poppendieck, T. (2003). Lean software development: An agile toolkit paperback. USA: Pearson.
[26] Chi, E. (2013). The false dichotomy between accessibility and usability. Proceedings 10th International Cross-Disciplinary Conference (Article 19). May 13-15, Rio de Janeiro, Brazil.
[27] Bevan, N. (2009). UsabilityNet methods for user centered design. Human-Computer Interaction: Theory and practice. In Jacko, J & Stephanides, D. (Eds.) Human-Computer Interaction: Theory and Practice (pp. 434-438). New Jersey: Lawrence Erlbaum Associates.
[28] Ferré, G. (2005). Marco de integración de la usabilidad en el proceso de desarrollo software. Tesis Doctoral. Univesidad Politécnica de Madrid.
[29] Jin, Q. (2002). Design of a virtual community based interactive learning environment. Information Sciences 140(1–2), pp. 171–191.
[30] Seffah, A. & Metzker, E. (2004). The obstacles and myths of usability and software engineering. Communications of the ACM 47(12), pp. 71-76.
[31] Holzinger, A. (2005). Usability engineering methods for software developers. Communications of the ACM 48(1), pp. 71-74.
[32] Ferré, X. & Baver, N. (2011). Usability planner: Development of a tool to support the process of selecting usability methods. Lecture Notes in Computer Science 6949, pp. 652-655.
[33] Quintero, E., Rivera, F. & Bolaños, B. (2010). Analysis of environmental factors and their effects on fungal spores in the atmosphere of a tropical urban area (San Juan, Puerto Rico). Aerobiologia 26(2), PP. 113–124.
[34] Alvear, A. & Quintero, G. (2013). Aeroallergens database design. Proceedings IWBBIO (pp. 559-567). March 18-20, Granada, Spain.
[35] Alvear, A. & Rivera, J. (2015). Web & mobile application for particulate fungi monitoring. Proceedings LACCEI Conference (In press). July 29-31, Santo Domingo, Dominican Republic.
[36] Alvear, A., Santamaría, M. & Claudio, E. (2013). Virtual Reality: A tool for treating phobias of heights. Proceedings LACCEI (1-10). August 14-16, Cancun, Mexico.

Un método basado en algoritmos genéticos de apoyo al diagnóstico TEA

Juan Cockbaine O.
Universidad Tecnológica de Chile – Chile

1 Introducción

Los individuos, sean niños o niñas, diagnosticados tempranamente como autistas o pertenecientes al espectro autista [17, 12] demandan grandes esfuerzos económicos, dedicación, paciencia y mucho amor de sus familiares, con el objeto de lograr paso a paso una lenta rehabilitación que, en la esperanza de sus padres, podría mejorar su calidad de vida a la vez que de la familia. Idealmente, esta rehabilitación debe ser apoyada por fonoaudiólogos, psiquiatras, psicólogos, kinesiólogos, nutricionistas y terapeutas ocupacionales, además de utilizar los servicios de establecimientos de educación especial o diferencial, entre otros aspectos prácticos.

Se indica que el autismo afecta a uno de cada 88 niños [11], y en diversas publicaciones [13] se expone la prevalencia que usualmente hace referencia a la presencia o proporción de sujetos con determinada condición en una población, tratándose de esta forma la epidemiología del problema de salud TEA. Las primeras sospechas de presencia de autismo en el individuo es entre los dos a tres años de edad o antes [3, 10] en que un médico podría observarlo y derivarlo a un especialista para una evaluación basada en la observación experta, reacción frente a estímulos y cuestionarios a los padres, entre otras técnicas.

Los padres, al recibir este diagnóstico médico angustiante e incierto respecto al impacto que producirá en el niño y su entorno familiar, se plantearán inmediatamente o con el pasar del tiempo una serie de interrogantes, con el objeto de entender y finalmente aceptar lo que significa tener un hijo autista. Por ejemplo, ¿Mi hijo será *normal* en algún momento de su vida? ¿Qué tan grave es esta condición… podrá mi hijo jugar con otros niños o permanecerá inmóvil con su vista perdida? ¿Cuándo se comunicará normalmente, como lo hacen los niños de su edad? ¿Cómo abordar tempranamente esta situación… con dietas especiales, terapias? ¿Qué tan acertado fue el diagnóstico médico? ¿Está el médico actualizado respecto del autismo? ¿Mi hijo presenta o presentará signos de deficiencia mental, en qué grado? ¿Cómo acepto y controlo sus rabietas y reacción a la frustración? ¿Cómo debo educar a sus hermanos para que acepten y entiendan la situación y no impacten fuertemente su propia niñez? ¿Debo vacunar normalmente a mi hijo? ¿Podrá mi hijo trabajar en el futuro? ¿Logrará independencia de sus padres? ¿Qué pasará con él si sus padres no están? ¿Cómo generaré recursos económicos suficientes para lograr un mejor estándar de vida para mi hijo? ¿Alguien lo amará aceptando su condición? Entre muchos otros.

Es así que el diagnóstico podría no tener la exactitud que los padres requieren para tratar de la mejor forma y tempranamente a sus hijos, pues dependerá de la experiencia del médico que diagnostica y los instrumentos de medición que emplea, pudiendo no establecerse con precisión la severidad del autismo que presenta el individuo de temprana edad, situación que podría retrasar tratamientos adecuados. Se indica que las intervenciones tempranas y certeras mejoran el pronóstico en estos individuos [1] y de la misma forma diversidad de situaciones e inquietudes son planteadas actualmente en foros de discusión vía Internet [14, 15].

Consciente del problema diversas instituciones han elaborado cuestionarios estándares para apoyar el diagnóstico del autismo, por ejemplo, el cuestionario Cociente de Espectro Autista Abreviado (AQS) [5-7]. También se han desarrollado videos y juegos educativos para el desarrollo de niños y adultos utilizando pictogramas [22]; aplicaciones personalizables para mejorar la comunicación, planificación de tareas y disfrutar de actividades de ocio [23]; sistema de comunicación alternativa basada en código QR para descifrar acciones, situaciones u objetos [24]; comunicadores digitales para ayudar al niño en su proceso de comunicación [25]; aplicaciones IPad que han permitido mejorar la sociabilización y la memoria visual de niños autistas [27]; software educativo para enseñanza de competencias emocionales y sociales [29], entre otros. Complementariamente, el productor de software SAP proyecta que para el año 2020 un 1% de los 65.000 empleados de la compañía serán autistas, precisando que el autismo y la tecnología congenian especialmente en pruebas y control de calidad de software [26].

Otras iniciativas se han focalizado en inferir computacionalmente, a partir de una conversación de un individuo con autismo, los déficit de comunicación de tal forma que se pueda ofrecer compensaciones utilizando agentes inteligentes [28]; aplicación de robots como herramientas de terapia centrada en comportamientos sociales [30]; simulación de funciones cognitivas utilizando redes neuronales [31]; utilización de Fuzzy Synchronization Likelihood para investigar conectividad funcional en pacientes con autismo [32]; discutir la utilización de *learning machine* y métodos de discriminación para diagnosticar autismo en niños [33]; modelos Bayesianos para explicar la relación entre percepción y sensación y su aplicación al autismo [34]; técnicas automáticas para cuantificar dos tipos de frases repetitivas, lo que el niño auto-repite y lo escuchado por un interlocutor [35].

Este trabajo plantea un método para la generación automática de conocimiento, utilizando técnicas computacionales, específicamente mapas mentales, ontologías, minería de texto y algoritmos genéticos. Se busca cuantificar con tal conocimiento los criterios narrativos definidos en DSM [2] para diagnosticar niños con autismo. El problema radica en la escasa precisión numérica de los criterios DSM para diagnosticar individuos con alta probabilidad de padecer autismo.

2 Materiales

Los materiales utilizados en este trabajo son: 1) Herramienta computacional para la confección de mapas conceptuales, 2) Herramienta computacional para realizar minería de texto, 3) Fuentes de datos que relatan vivencias de autismo en niños, 4) Manual DSM, 5) Cuestionario AQ-Short, 6) Herramientas computacionales para el desarrollo del algoritmo genético, 7) Equipamiento computacional.

2.1 Mapas Conceptuales

Buzan [16] define que los mapas conceptuales representan una expresión del pensamiento, que es una función natural de la mente humana. Los mapas se representan gráficamente como una red de conceptos interconectados que expresan un entendimiento abstracto. En [19] se puede encontrar una lista de herramientas para confeccionar mapas, pero en este trabajo se utiliza XMind [21]. Para el correcto uso de esta herramienta, es recomendable basarse en definiciones o estándares, que son el resultado de la discusión e investigación de comunidades especializadas y que para efectos de este trabajo corresponde al DSM. Por ejemplo, la Figura 1 muestra el concepto resumido TGD definido en el DSM, utilizando XMind.

Figura 1: Mapa Conceptual TGD (Elaboración propia utilizando XMind [21])

2.2 Minería de Texto

El método propuesto requiere obtener información desde fuentes de texto contenedores de relatos (vivencias) de familiares de niños autistas. En la sección 3.4 se puede observar cómo la información generada aporta en la formulación de la función objetivo del algoritmo genético desarrollado. Diversas herramientas se han desarrollado para realizar minería de texto, por ejemplo, SAS Text Analytics, IBM Text Analytics, SAP Text Analytics, Lexalytics Text Analytics, Smartlogic, Attensity, ai-one, Rapid Miner, ATLAS.ti, Semantria etc. Otras herramientas de libre acceso o en modalidad *trial* son KH Coder, Carrot2, GATE, Natural Language Toolkit etc. [36]. Pero en este trabajo se utilizó Semantria, que se instala como plug-in en MS Excel [37]. Se evaluó esta herramienta y otras, tales como AtlasTi, GATE, KHCoder, y se decidió que Semantria es la más adecuada para los fines, por su facilidad de uso, incorporación avanzada de análisis de Sentimiento (tono o énfasis con que se redacta) y facilidades para la definición de taxonomías.

2.3 Fuentes de textos

Como se indicó en la sección anterior, la estrategia planteada en este trabajo requiere relatos sobre los que se pueda realizar minería de texto. Para estos efectos las fuentes se obtuvieron desde Internet, provenientes de 21 sitios web que tratan el autismo en niños, y desde los cuales se han utilizado 1 o 2 relatos para efectos de la minería de texto. Estos materiales no se muestran en esta sección, por la cantidad de texto que contienen, pudiéndose consultar los diversos blogs en que los padres tratan el autismo de sus hijos, por ejemplo, en el índice de blogs y foros de habla hispana sobre autismo [15].

2.4 Manual DSM

En este trabajo se han considerado los criterios definidos en el DSM [2]. El conjunto de criterios DSM constituye la base para definir los mapas de conceptos. Cabe señalar que el Comité Elaborador del DSM ha establecido contactos con American Psychiatric Association y con más de 60 organizaciones y asociaciones interesadas en la confección de esta normativa. Por ejemplo, American Health Information Management Association, American Nurse's Association, American Occupational Therapy Association, American Psychoanalytic Association, American Psychological Association, American Psychological Society, Coalition for the Family, Group for the Advancement of Psychiatry, National Association of Social Workers, National Center for Health, Statistics y World Health Organization. El propósito del DSM es proporcionar descripciones de las categorías diagnósticas, con el fin de que los clínicos y los investigadores puedan diagnosticar, estudiar, intercambiar información y tratar los distintos trastornos mentales. Cabe señalar que los criterios específicos DSM, referidos al autismo, no especifican información cuantitativa para lograr mayor precisión.

2.5 Cuestionario Autism-Spectrum Quotient, AQ-Short

El Cociente de Espectro Autista abreviado es un cuestionario con 50 preguntas orientadas a determinar rasgos asociados a las condiciones de un individuo en el espectro

autista. Se abordan habilidades sociales, de comunicación, imaginación, atención y tolerancia al cambio. La versión para niños AQC se pueden responder por Internet [5]. La puntuación del cuestionario es entre 0 y 150 puntos. Se indica que una puntuación superior a 76 puntos sería indicativa de la presencia de comportamientos que justificarían realizar una valoración diagnóstica por parte de un profesional de la salud. El AQC se utiliza en este trabajo con el objeto de contrastar que los valores (representación numérica de un individuo con alta probabilidad de padecer autismo, Tabla 4) generados por el algoritmo genético, se correspondan con una puntuación alta en la aplicación del test. Otros test se pueden consultar en [38].

2.6 Algoritmos Genéticos

El potencial de los algoritmos genéticos se basa en la teoría de la evolución de Darwin. Tal teoría fue adoptada por Holland [4, 18], que a fines de los 60s desarrolló una técnica que permitió automatizarla computacionalmente. La técnica se basa en los mecanismos de selección de la naturaleza, en que los individuos más aptos de una población son los que sobreviven al adaptarse fácilmente a los cambios que se producen en su entorno. Por ejemplo, aquel individuo de una población más apto que otros respecto una función objetivo o función aptitud a maximizar, será propagado en generaciones sucesivas, buscando la forma de alcanzar el óptimo, que en este caso, correspondería a un individuo con alta probabilidad de ser autista.

Algunas características de los algoritmos genéticos son: 1) trabajan sobre un conjunto de individuos o población que inicialmente es generada aleatoriamente representando posibles soluciones a un problema; 2) un individuo o cromosoma perteneciente a una población está compuesto por un conjunto de genes que representan una característica del individuo, un genotipo corresponde al conjunto de genes de un individuo o la descripción genética del individuo, y cada gen ocupa una posición determinada en el cromosoma llamada locus; 3) a cada uno de los valores que puede tomar un gen se lo conoce como alelo; 4) es necesario especificar una función objetivo a maximizar o minimizar en qué valor de adaptación o ajuste representa la habilidad del individuo para resolver el problema en cuestión; y 5) los operadores genéticos comúnmente utilizados son *selección* para decidir qué individuos contribuirán en la formación de la próxima generación de individuos y el *cruce* y la *mutación* para la reproducción. En este trabajo se utiliza el *framework open source* Pyevolve [9] escrito en lenguaje Python [20]. Para esto, es necesario instalar Python y luego instalar Pyevolve, pudiéndose ejecutar en sistemas Windows y Linux.

2.7 Equipamiento computacional

Computador del tipo *notebook* con un Procesador Intel Core i7-4600M CPU 2.9 GHz., 8GB RAM, sistema operativo de 64 bits, procesador x64, Windows 8 Enterprice, 500 GB en disco y conexión a Internet.

3 Método

3.1 Mapas Conceptuales

La Figura 1 muestra que el concepto Trastorno Autista es parte de un todo conocido como Trastornos Generalizados del Desarrollo. Este trabajo se focaliza en el Trastorno Autista (TEA). La Figura 2 muestra que el trastorno autista o autismo infantil es abordado desde varias dimensiones, por ejemplo, características diagnósticas, criterios de diagnóstico, síntomas y trastornos asociados, etc.

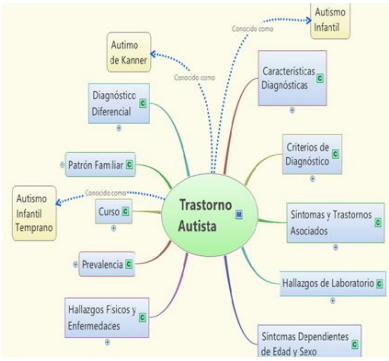


Figura 2: Trastorno Autista (Elaboración propia utilizando XMind [21])

La conceptualización de *Características Diagnósticas* presentado en Figura 3A y Figura 3B establece tres criterios: 1) el criterio A se refiere a la preocupación persistente por parte de objetos, manierismos motores repetitivos y estereotipados, adhesión a rutinas no funcionales, lenguaje estereotipado repetitivo idiosincrásico, alteración de la comunicación, incapacidad para desarrollar relaciones, alteración de la comunicación entre otros aspectos. 2) El criterio B se focaliza en la interacción social y el lenguaje. 3) El criterio C da cuenta que el trastorno autista no se explica mejor por la presencia del trastorno de Rett o el trastorno des integrativo infantil, aspectos que no son abordados en este trabajo.

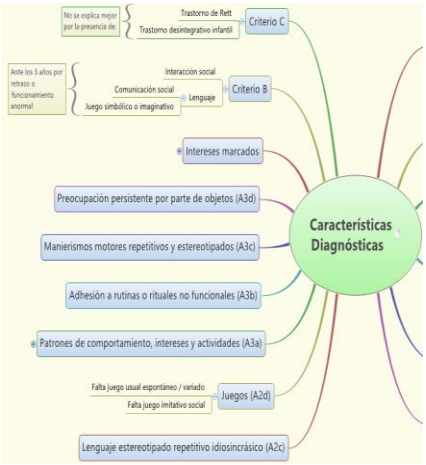


Figura 3A: Características Diagnósticas (Elaboración propia utilizando XMind [21])

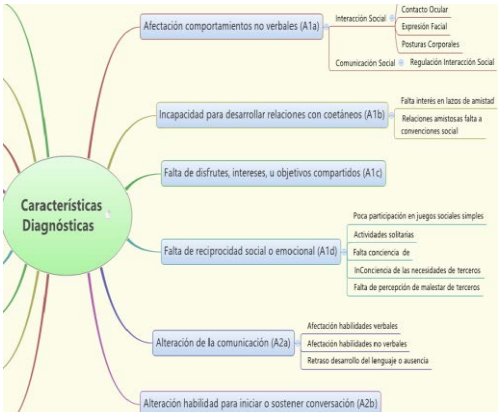


Figura 3B: Características Diagnósticas (Elaboración propia utilizando XMind [21])

Síntomas y trastornos asociados se expone en la Figura 4. Por ejemplo, respecto del humor y la efectividad: se ríe o llora sin un motivo aparente y hay ausencia aparente de reacciones emocionales. Respecto del sueño: hay un despertar recurrente o dificultad para dormir. Respecto la ingestión alimentaria: la dieta es limitada y puede ser a deshoras *picoteando* alimentos.

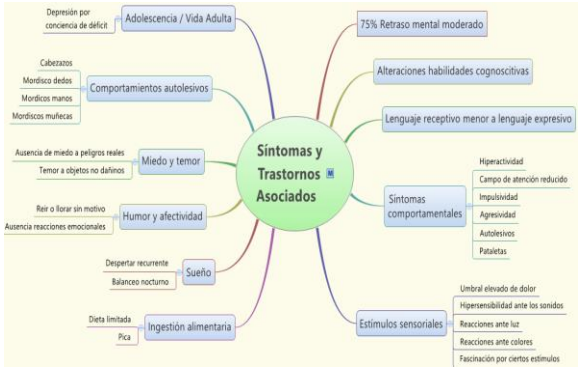


Figura 4: Síntomas y trastornos asociados (Elaboración propia utilizando XMind [21])

Los criterios de diagnóstico DSM se presentan en la Figura 5 y con mayor detalle en las Figuras 6, 7 y 8 para el criterio A.

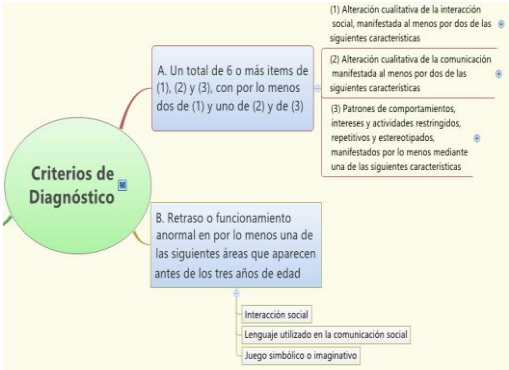


Figura 5: Criterios de diagnóstico (Elaboración propia utilizando XMind [21])

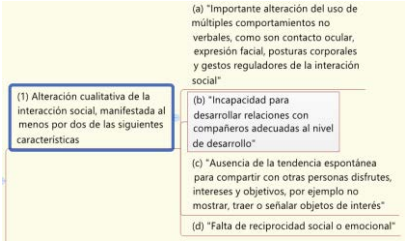


Figura 6: Alteración cualitativa de la interacción social (Elaboración propia utilizando XMind [21])

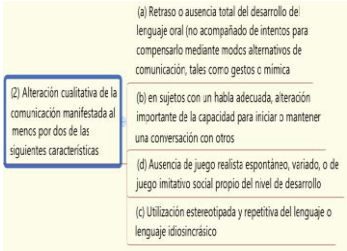


Figura 7: Alteración cualitativa de la comunicación (Elaboración propia utilizando XMind [21])

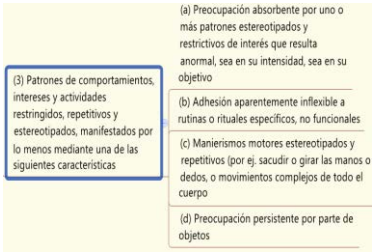


Figura 8: Patrones de comportamientos (Elaboración propia utilizando XMind [21])

El producto de esta etapa es un mapa completo de conceptos (no se muestra la totalidad de mapas elaborados por restricciones de espacio), que se constituye en un insumo para la especificación de taxonomías.

3.2 Taxonomías

En esta sección se especifican taxonomías con el objetivo de precisar la minería de texto a realizar. Se espera que de un determinado término, por ejemplo *comida*, se obtengan textos que traten sobre pizzas, sándwiches y otras frases asociadas. Las taxonomías se definieron como un conjunto T no vacío de conceptos explícitamente definidos en un mapa conceptual M, en este caso el mapa *Trastorno Autista* (Figura 2):

T = {{Características Diagnósticas}, {Criterios de Diagnóstico}, {Síntomas y Trastornos Asociados}, {Hallazgos de Laboratorio}, {Hallazgos Físicos y Enfermedades}, {Síntomas Dependientes de Edad y Sexo}, {Prevalencia}, {Curso}, {Patrón Familiar}, {Diagnóstico Diferencial}}.

El alcance de este trabajo se definió de la siguiente forma:

T* ⊆ T, tal que T* = T - {{Criterio B}, {Criterio C}, {Hallazgos Físicos y Enfermedades}, {Hallazgos de Laboratorio}, {Prevalencia}, {Patrón Familiar}, {Curso}, {Criterios de Diagnóstico}, {Diagnóstico Diferencial}}

Respecto a los conceptos no considerados en T*, el Criterio B y el Curso en Primera Infancia, corresponden a evidencias difíciles de detectar tan tempranamente, al menos por los padres del individuo, quienes acuden a los especialistas médicos cuando el trastorno se hace más evidente. Esto es, cuando la edad de individuo es superior a los 2 años. El Criterio C se refiere a que el trastorno no se explica mejor por la presencia de un trastorno de Rett o de un trastorno des integrativo infantil, aspectos que no corresponden al alcance de este trabajo. Hallazgos físicos, enfermedades y hallazgos de laboratorio pudiesen o no estar presentes en individuos con autismo, y DSM no profundiza lo suficiente como para abordarlos. Similarmente, Prevalencia y Patrón familiar son de carácter informativo. Los Criterios de Diagnóstico se refieren a la forma de diagnosticar autismo propuesto por DSM y no aportan mayor información para efectos de especificar taxonomías, pero sí en aspectos de programación del algoritmo genético. El Diagnóstico diferencial se refiere a trastornos tratados en otros capítulos del DSM, que no se focalizan en el trastorno autista. Es así que:

T* = {{A1a}, {A1b}, {A1c}, {A1d}, {A2a}, {A2b}, {A2c}, {A2d}, {A3a}, {A3b}, {A3c}, {A3d}, {Intereses Marcados}, {Síntomas y Trastornos Asociados}, {Síntomas Dependientes de Edad y Sexo en niños pequeños}}

Del estudio del DSM se incluyeron los conceptos Intereses Marcados y Síntomas dependientes de Edad y Sexo, con el objeto enriquecer el conjunto de taxonomías. Las

taxonomías asociadas al grupo Criterio A, específicamente los sub grupos A1, A2 y A3 se han precisado utilizando sinónimos (expresiones o palabras que tienen un significado similar o idéntico entre sí) y que pudiesen ser de uso común en las fuentes de datos o vivencias sobre las que se realiza el proceso de minería de texto. Para estos efectos se acudió a un diccionario de sinónimos [8].

La notación utilizada para especificar una taxonomía es la siguiente: Nombre genérico de la taxonomía y los conceptos que agrupa. Ejemplo: A1a= {gestos corporales, contacto ocular, expresión facial, posturas corporales, interacción social}, donde A1a denota la categoría o sub grupo DSM y los conceptos que agrupa, tales como *contacto ocular*, *interacción social*, etc. los que se encierran entre "{"y"}". La precisión de conceptos se anota entre ("y"), por ejemplo: Gestos corporales (ademán, mueca, aspaviento, visaje, seña, guiño, tic, rostro, cara, semblante, aire, expresión, aspecto, actitud, postura), que agrupan un conjunto de sinónimos para los conceptos *gestos* y *corporales*. Finalmente, la taxonomía se reduce a la siguiente expresión: A1aGestosCorporales= {gesto, expresión, aspecto, postura, corporal, cuerpo}, donde que A1aGestosCorporales es el nombre de la taxonomía, y los conceptos que agrupa la taxonomía se anotan entre {"y"}. La Tabla 1 lista algunas de las taxonomías a utilizar en la minería de texto, que se explicitan después del símbolo "→". Se considera que en promedio, por taxonomía, se debe ajustar a seis términos con el objeto de no dispersar la minería de texto posterior.

Tabla 1: Taxonomías Grupos DSM A1, A2 y A3

Términos A1a= {**gestos corporales, contacto ocular**, expresión facial, **posturas corporales**, **interacción social**}
Gestos corporales (ademán, mueca, aspaviento, visaje, seña, guiño, tic, rostro, cara, semblante, aire, expresión, aspecto, actitud, postura) → **A1aGestosCorporales= {gesto, expresión, aspecto, postura, corporal, cuerpo}**
Contacto ocular (comunicación, trato, relación, vecindad, familiaridad, intimidad, lente, vidrio, ojo, visual) → **A1aContactoOcular= {contacto, comunicación, ocular, ojo, visual}**
Posturas corporales (posición, actitud, colocación, situación, pose, apariencia, aire, forma, modo, presencia, figura, planta, talle, tronco, torso, complexión, figura) → **A1aPosturasCorporales= {postura, pose, modo, cuerpo, figura}**
Interacción Social (Sociable, social, comunicativo, abierto, amable, confiado, cordial, sencillo, extrovertido, expansivo, accesible, simpático, agradable, conversador) → **A1aInteracciónSocial= {comunicación, comunicar, accesible, socializar, social}**
Términos A1b = {lazos de amistad, **convenciones sociales**}
Convenciones sociales (acuerdo, convenio, tratado, pacto, ajuste, mundo, humanidad, comunidad, colectividad, generalidad, grupo, familia, población, nación, pueblo, estado, ciudadanía, habitantes) → **A1bConvencionesSociales= {ajuste, acuerdo, convención, norma, social, sociedad, sociable}**
Términos A1c= {disfrutes, **intereses compartidos**, objetivos compartidos}
Intereses compartidos (afecto, tendencia, inclinación, disposición, apego, predilección, cariño, repartir, dividir, distribuir, ayudar, auxiliar, colaborar, participar, intervenir) → **A1cInteresesCompartidos= {interés, inclinación, tendencia, compartir, colaborar, participar}**
Términos A1d= {participación juegos, **actividades solitarias**, percepción necesidades, percepción malestar}
Actividades solitarias (ocupación, profesión, tarea, labor, función, trabajo, insociable, huraño, retraído, tímido, misántropo, misógino, huidizo, esquivo, intratable) → **A1dActividadesSolitarias= {actividad, tarea, solitario, retraído, insociable}**
Términos A2a= {**habilidad verbal**, habilidad no verbal, **retraso del lenguaje**, ausencia lenguaje}
Habilidad verbal (capacidad, competencia, aptitud, destreza, pericia, experiencia, gracia, ingenio, soltura, práctica, cualidad, mano, maña, oral, hablado, expresado) → **A2aHabilidadVerbal= {habilidad, capacidad, soltura, oral, habla, hablado}**
Retraso Lenguaje (atraso, demora, aplazamiento, dilación, retardo, lengua, idioma, habla, jerga, jerigonza, verbo, estilo) → **A2aRetrasoLenguaje = {atraso, demora, retardo, lenguaje, habla, palabra}**
Términos A2b= {iniciar conversación, sostener conversación}
Sostener Conversación (sustentar, sujetar, soportar, agarrar, apoyar, aguantar, mantener, cargar, equilibrar, reforzar, consolidar, charla, diálogo, coloquio, palique, cháchara, parrafada, parloteo, rollo) → **A2bSostenerConversación= {sostener, mantener, conversación, charla, diálogo}**
Términos A2c= {**lenguaje estereotipado**, lenguaje repetitivo, lenguaje idiosincrásico}
Lenguaje estereotipado (lengua, idioma, habla, palabra, jerga, jerigonza, verbo, estilo, estándar, invariable, repetido, calcado, clónico) → **A2cLenguajeEstereotipado= {lenguaje, habla, palabra, estereotipado, invariable, repetido, calcado}**
Términos A2d= {juego espontáneo, juego variado, **juego imitativo**}

Juego imitativo (entretenimiento, esparcimiento, diversión, recreo, placer, solaz, distracción, deporte, pasatiempo, descanso, copiar, emular, plagiar, falsificar, repetir, parodiar, fingir, representar, simular, parecer) → **A2dJuegoImitativo= {juego, entretenimiento, diversión, copiado, repetido, parecido}**
Términos A3a= {comportamiento estereotipado, intereses restringidos, **actividades repetitivas**, pautas de interés}
Actividades repetitivas (ocupación, profesión, tarea, labor, función, trabajo, redundante, reiterado, insistente, frecuente, periódico) → **A3aActividadRepetitiva= {actividad, tarea, repetitiva, reiterada}**
Términos A3b= {**adhesión a rutinas**, rituales no funcionales}
Adhesión Rutina (aprobación, conformidad, aceptación, ratificación, consentimiento, solidaridad, hábito, práctica, costumbre, usanza, uso, repetición) → **A3bAdhesiónRutinas= {adhesión, aceptación, rutina, costumbre}**
Término A3c= {manierismos motores repetitivos, **manierismos motores estereotipados**}
Manierismos motores estereotipados (manera, forma, método, procedimiento, peculiaridad, carácter, personalidad, expresión, giro, estándar, invariable, repetido, calcado, clónico) → **A3cManierismosEstereotipados= {manera, expresión, manierismo, estereotipo, estándar, invariable, repetido}**
Términos A3d= {preocupación persistente objetos}
...

La Tabla 2 resume las taxonomías Intereses Marcados (IM), Síntomas y Trastornos Asociados (STA) y Síntomas Dependientes de Edad y Sexo en niños pequeños (SDE).

Tabla 2: Taxonomías IM, STA y SDE

Intereses Marcados
IMAnomaliasPosturales= {puntilla, postura}
IMPreocupaciónObjetosFascinación= {girar, abrir, cerrar, voltear, tirar}
IMRecopilación= {recopilar, datos}
IMAlinear= {alinear, objetos, cosas}
IMResistenciaCambios= {resistencia, cambios}
IMRutinasRituales= {rutinas, rituales}
Síntomas y Trastornos Asociados (STA)
STAComportamientoAutolesivo= {cabezazos, mordiscos}
STAAusenciaMiedo= {ausencia, miedo, temor}
STAHumor= {humor, reír, llorar}
STAAfectividad= {ausencia, emociones}
STASueño= {despertar, balanceo, noche, nocturno}
STASíntomasComportamiento= {hiperactividad, atención, impulsividad, agresividad, lesividad, pataletas}
STAEstímulosSensorialesDolor= {umbral, dolor}
STAEstímulosSensorialesHipesensibilidad= {hipersensibilidad, sonidos, ruidos}
STAEstímulosSensorialesLuz= {reacción, luz, luminosidad}
Síntomas Dependientes de Edad y Sexo en niños pequeños (SDE)
SDEIncapacidadAbrazo= {incapacidad, abrazo, abrazar}
SDEIndiferenciaAfecto= {indiferencia, aversión, afecto}
SDEAversiónContacto= {indiferencia, aversión, contacto, personas}
SDEPersonasIntercambiables= {personas, intercambiables}
SDEAferrarsePersona= {aferrar, aferrarse, persona}

3.3 Minería de texto

Sobre los relatos se realizó un proceso de minería de texto utilizando Semantria, esto es: Análisis de polaridad o sentimiento (expresión o énfasis) con que se ha redactado el relato; descarte de relatos no contextuados en el autismo y análisis de texto detallado buscando que los relatos se encuentren contextuados en una o más taxonomías. La polaridad de cada vivencia (negativa, neutra o positiva) aporta al investigador para decidir la exclusión o no de aquellas negativas pensando que pudiesen no ser objetivas. Pero hay que tener en cuenta el contexto de lo que significa el autismo y lo que posiblemente aportaría una vivencia negativa, tal vez contribuya con más información respecto a causas, efectos o situaciones que una vivencia cuya redacción sea neutral o positiva. En este trabajo no se excluyeron narraciones. La Figura 9 presenta el resultado de análisis de polaridad sobre una muestra de vivencias.

Source text	Document Sentiment	
	Sentiment	Sentiment Polarity
necesito conocer padres de niños	-0,400000006	negative
Acudo a este medio porque me	0,003235723	neutral
En la actualidad el autismo	0,190909088	neutral
Hola me llamo Gloria y tengo una	0,400000006	positive
Hola a todos. Veran, tengo un hijo	-0,612500012	negative
Estos niños pueden ser	-0,299628943	negative
Resulta muy común que los niños	-0,400000006	negative
Una manera de canalizar su	-0,070000023	negative
Gracia Viscasillas presentó una	0,300000012	positive
Mar Calado nos presentó la nueva	-0,200000003	negative
Los psicoanalistas Pilar Foz y Félix	0	neutral
mi hija es autista la verdad es que	-0,074999988	negative

Figura 9: Análisis de Sentimiento o Polaridad (Elaboración propia utilizando Semantria [37])

Para efectos de descartar relatos no contextuados en el autismo, esto es, aquellos que pudiesen no estar focalizados claramente en tal concepto, se programó el *query booleano* "Autismo: (autismo) OR (autist*) OR ("trastorno autista") OR ("autismo infantil") OR (kanner)" y se aplicó la taxonomía "Autismo= {"trastorno autista", autismo, "autismo infantil", kanner, "autismo infantil temprano"}", que concluyo en descartar 3 relatos de un total de 54. Los restantes relatos fueron procesados aplicando el conjunto de taxonomías definidas. Se buscó encontrar evidencia, esto es, que los relatos traten una o más temáticas taxonómicas. La Figura 10 muestra los hallazgos encontrados para el criterio A1a, donde los relatos se asocian a ninguna, una o más taxonomías específicas, lo cual permite calcular un porcentual por taxonomía, valores máximos y mínimos por grupo.

Source Text	Document Sentiment	User Category	User Category Strength
necesito conocer padres de niños	-0,4		
Acudo a este medio porque me	-0,0272846		
En la actualidad el autismo	0,19090909		
Hola me llamo Gloria y tengo una	0,33333334		
Hola a todos. Veran, tengo un hijo	-0,6125	A1aPosturasCorporales	0,527037084
Hola a todos. Veran, tengo un hijo	-0,6125	A1aGestosCorporales	0,503275216
Estos niños pueden ser	-0,2996289		
Resulta muy común que los niños	-0,4	A1aPosturasCorporales	0,507736266
Una manera de canalizar su	-0,07	A1aInteracciónSocial	0,584377289
Una manera de canalizar su	-0,07	A1aPosturasCorporales	0,564095974
Una manera de canalizar su	-0,07	A1aGestosCorporales	0,547212422
Gracia Viscasillas presentó una	0,30000001		
Mar Calado nos presentó la nueva	-0,2		
Los psicoanalistas Pilar Foz y Félix	0	A1aPosturasCorporales	0,569184542
Los psicoanalistas Pilar Foz y Félix	0	A1aGestosCorporales	0,543549716
mi hija es autista la verdad es que	-0,075	A1aPosturasCorporales	0,525200486
Mi hijo se llama Josue, nombre	0,37503988	A1aGestosCorporales	0,494184941
Mi hijo se llama Josue, nombre	0,37503988	A1aPosturasCorporales	0,481544048
el ser padre de un niño autista es	0,05	A1aGestosCorporales	0,515352309
tengo un niño autista a veces	0,40000001		
Los primeros síntomas del	-0,0561364	A1aInteracciónSocial	0,563877285
Los primeros síntomas del	-0,0561364	A1aContactoOcular	0,524870157
Es probable que aquellos recién	-0,25		
Yo tengo un hijo de 5 años	0,14048812	A1aInteracciónSocial	0,604020476
Yo tengo un hijo de 5 años	0,14048812	A1aContactoOcular	0,521916687
me podrían ayudar con	0,12083331		
Es importante entender la	-0,0333333		
El autismo es un enigma	-0,3214286	A1aInteracciónSocial	0,573960304
El autismo es un enigma	-0,3214286	A1aContactoOcular	0,503933966
Una compañera de trabajo	0,215		
Los investigadores no han	-0,5		
Hola, mi hijo precioso se llama	0,25833333		
tengo un niño de ya apunto de	-0,1333333		
Hilda, mi nino es autista y tiene 4	0,30000001		
mi niño se llama mateo y es	0,18000002	A1aPosturasCorporales	0,485514104
Me podian ayudar. Tengo un hijo	0,10714287	A1aPosturasCorporales	0,508609235
Hola yo tengo un hijo de 13 años y	0,38000003		
100%	73,0000		
A1aContactoOcular	0,0822		
A1aExpresiónFacial	0,0274		
A1aGestosCorporales	0,1644		
A1aInteracciónSocial	0,0959		
A1aPosturasCorporales	0,2192		
No categoriza	0,4110		
100%	1,0000		

Figura 10: Minería de texto A1a (Elaboración propia utilizando Semantria [37])

El gráfico radial presentado en la Figura 11 resume la Figura 10. El gráfico radial de la Figura 12 resume la información obtenida para el criterio A2a. Por motivos de espacio no se han incluido la totalidad de los análisis de minería de texto.

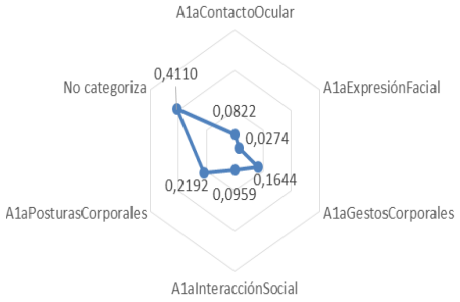


Figura 11: Resumen de hallazgos para Grupo A1a (Elaboración propia)

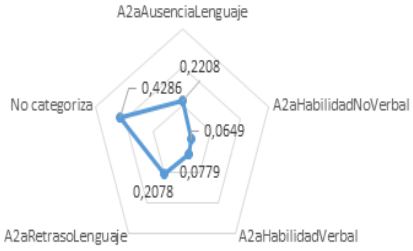


Figura 12: Resumen de hallazgos para Grupo A2a (Elaboración propia)

En la Tabla 3 se listan los valores mínimo y máximo. Con esta estrategia, se estableció el rango de valores que cada gen del cromosoma puede tomar en el proceso de optimización del algoritmo genético.

Tabla 3: Mínimos y Máximos

	Mínimo	Máximo
A1a	0,02739726	0,219178082
A1b	0,166666667	0,203703704
A1c	0,105263158	0,175438596
A1d	0,032258065	0,241935484
A2a	0,064935065	0,220779221
A2b	0,019230769	0,019230769
A2c	0,192771084	0,265060241
A2d	0,036363636	0,072727273
A3a	0,016949153	0,254237288
A3b	0,058823529	0,058823529
A3c	0,120689655	0,120689655
A3d	0,352941176	0,352941176
IM	0,017857143	0,160714286
STA	0,011235955	0,157303371
SDE	0,012658228	0,481012658

3.4 Algoritmo Genético

Los pasos generales del algoritmo genético son los siguientes:

1. [Inicio] Genera una población aleatoria de n cromosomas que son posibles soluciones para el problema.
2. [Fitness] Evalúa la función de fitness f(x) de cada cromosoma perteneciente a la población.

3. [Nueva Población] Crea una nueva población repitiendo los siguientes pasos hasta que la nueva población se complete:
   3.1 [Selection] Selecciona dos cromosomas padres desde una población considerando el fitness de cada uno. Mejores fitness tendrán mayor chance de ser seleccionados.
   3.2 [Crossover] Cruzamiento de los padres con el objeto de generar hijos (offspring).
   3.3 [Mutation] Mutación de los nuevos hijos en cada locus (posición en el cromosoma).
   3.4 [Accepting] Aceptación del nuevo offspring para integrarlo a la nueva población.
4. [Replace] Utilización de la población generada para iteración del algoritmo.
5. [Test] Verificación de la condición de término. Si la condición es satisfecha el algoritmo para y retorna la mejor solución encontrada en la población actual.
6. [Loop] Ir al paso 2.

En las dos secciones siguientes, se explicitan los principales trozos de código Python/Pyevolve (algoritmo genético desarrollado en este trabajo) y el correspondiente comentario previo al código.

Sección 1 (declarativas):

1. La estructura del individuo o genotipo se basa en definiciones de Criterio A, IM, STA y SDE:

   (A1a; A1b; A1c; A1d; A2a; A2b; A2c; A2d; A3a; A3b; A3c; A3d; IM; STA; SDE), en que A1a, A1b, etc. representa un gen.

2. El genotipo corresponde a una lista de 15 números reales definida e inicializada de la siguiente forma:

```
Cromosoma = G1DList.G1DList(14)
cromosoma.initializator.set(Initializators.G1DListInitializatorReal)
```

3. La función de evaluación de cada cromosoma entrega una puntuación o fitness cuando es llamada. El algoritmo genético buscará maximizar la función, esto es, en lo posible encontrar un óptimo. La función a maximizar se llama "Autismo" y la estructura de datos "xlist" representa el cromosoma. La función de evaluación programada, verifica el rango de valores mínimos y máximos listados en Tabla 3 para cada gen. Además, verifica lo indicado por DSM, en que del criterio A1 se deben cumplir dos aspectos al menos, del criterio A2 se deben cumplir un criterio al menos:

```
def Autismo(xlist)
    if 0.02739726 <= xlist[0] <= 0.219178082 :
    DeA1dos+= 1
    if 0.166666667 <= xlist[1] <= 0.203703704:
    DeA1dos+= 1…
    if ((DeA1dos >= 2) and (DeA2uno >= 1) and (DeA3uno >= 1) and (IM >= 1)):
     puntuacion=DeA1dos+DeA2uno+ ...
     else:
     puntuacion= 0
     return puntuación
```

Sección 2 (Cuerpo principal del algoritmo genético):

4. La variable "ajuste" permite controlar que la sumatoria de valores generados para un cromosoma no exceda al 100%:

```
ajuste=0
for gen in xlist:
ajuste+= gen
```

5. En la búsqueda de un óptimo, el algoritmo genético generó Genomas (Poblaciones) en que las soluciones para una población se seleccionan de acuerdo a su fitness con el objeto de obtener poblaciones mejores que las anteriores. Si una solución se puede representar por su valor o fitness, entonces la búsqueda de una solución corresponde a buscar un punto máximo o mínimo:

```
ga.setPopulationSize(10000)
ga.setGenerations(3000)
ga.minimax= Consts.minimaxType["maximize"]
```

6. La operación crossover crea una generación de nuevos individuos (offspring) tomando información de los ancestros. Se puede establecer una probabilidad respecto de esta operación, por ejemplo, si la probabilidad es 100%, todos los offspring son por crossover, si es 0%, la generación entera es construida a partir de copias exactas de cromosomas padres. En caso contrario, crossover toma parte del cromosoma padre:

```
ga.setCrossoverRate(0.7)
```

7. En caso que no se defina mutación, los offspring son tomados después del crossover sin cambio, caso contrario, una parte del cromosoma es cambiado. Si la probabilidad de mutación es 100%, el cromosoma entero es cambiado, si es 0% nada es cambiado:

```
ga.setMutationRate(0.1)
```

8. Los cromosomas de una población son seleccionados para ser padres en crossover. Para tales efectos se puede utilizar el método de la ruleta, en que los cromosomas con mejor fitness tienen mayor probabilidad de ser seleccionados al girar la ruleta:

```
ga.selector.set(Selectors.GRouletteWheel)
```

9. Evolucionar corresponde a encontrar un máximo en las poblaciones que se han generado. La evolución finaliza al encontrar convergencia:

```
ga.terminationCriteria.set(GSimpleGA.ConvergenceCriteria)
ga.evolve(freq_stats=100)
```

10. El óptimo o mejor individuo en la población generada:

```
pop = ga.getPopulation()
conAutismo= ga.bestIndividual()
```

Los resultados obtenidos, después de ejecutar el paso 10, se representan en Tabla 4, lo que corresponde a la caracterización de un individuo con alta probabilidad de padecer autismo, de acuerdo al método expuesto en este trabajo.

Tabla 4: Valores para cada gen generado por PYevolve

- G1DList List size: 15 List:
[0.067400659535971338,
0.16853716426720922,
0.12163811903143572,
0.067400659535971338,
0.068016280911308694,
0.018897659264526318,
0.19992777815106244,
0.067400659535971338,
0.067400659535971338,

```
0.048140578753312152,
0.067400659535971338,
0.067400659535971338,
0.067400659535971338,
0.067400659535971338,
0.067400659535971338]
>>>
```

4 Contrastación y Conclusiones

Para contrastar los resultados expuestos en Tabla 4, se utilizó el cuestionario AQC en versión Web. Para estos efectos, la Tabla 5 presenta una equivalencia porcentual entre los valores óptimos obtenidos por el algoritmo genético (columna "Óptimo") y el valor mínimo y máximo descubierto en el proceso de minería de texto (columnas "Mínimo" y "Máximo") para cada gen. La columna "Respuesta AQ-C" indica el criterio de equivalencia para las opciones 1, 2, 3 y 4 que tabula el cuestionario. La columna "% del Máximo" es el valor porcentual que el algoritmo genético alcanza (Columna Óptimo") respecto la columna "Máximo" que se obtuvo en el proceso de minería de texto. El llenado de la matriz explicita la equivalencia entre columna "% del Máximo" y rangos para columnas 1, 2, 3 y 4 del cuestionario.

Tabla 5: Equivalencia óptimo y respuesta AQ-C

	Mínimo	Máximo	Óptimo	% del Máx	Respuesta AQ-C			
					(1)	(2)	(3)	(4)
					AcuerdoTotal	AcuerdoParcial	Des.Parcial	Des. Total
					75 ≤ x ≤ 100	50 ≤ x ≤ 74	25 ≤ x ≤ 49	0 ≤ x ≤ 24
A1a	0,02739726	0,219178082	0,06740066	30,75155094			3	
A1b	0,166666667	0,203703704	0,168537164	82,73642597	1			
A1c	0,105263158	0,175438596	0,121638119	69,33372804		2		
A1d	0,032258065	0,241935484	0,06740066	27,85893926			3	
A2a	0,064935065	0,220779221	0,068016281	30,80737426			3	
A2b	0,019230769	0,019230769	0,018897659	98,26782935	1			
A2c	0,192771084	0,265060241	0,199927778	75,42729811	1			
A2d	0,036363636	0,072727273	0,06740066	92,67590651	1			
A3a	0,016949153	0,254237288	0,06740066	26,5109261			3	
A3b	0,058823529	0,058823529	0,048140579	81,83898445	1			
A3c	0,120689655	0,120689655	0,06740066	55,84626084		2		
A3d	0,352941176	0,352941176	0,06740066	19,09685356				4
IM	0,017857143	0,160714286	0,06740066	41,93818808			3	
STA	0,011235955	0,157303371	0,06740066	42,84756208			3	
SDE	0,012658228	0,481012658	0,06740066	14,01224238				4

La Tabla 6 muestra las respuestas al cuestionario, considerando la equivalencia presentada en la Tabla 5. La columna "T" corresponden a las preguntas del Test y la columna "Cromosoma" agrupa los genes. El test se respondió seleccionando el gen más acorde con la pregunta, marcando el número 1 o 2 o 3 o 4 de acuerdo a la Tabla 5. Para algunas preguntas del cuestionario, la respuesta se registra marcadas considerando que, por ejemplo, en la pregunta 38 se consulta *es bueno para las conversaciones sociales* y el gen asociado (A2b) indica que es un tema altamente (valor igual a 1) tratado en los foros de discusión, o sea, es una situación que afecta altamente a los individuos con autismo por la falta o ausencia del lenguaje, por lo tanto, corresponde asociar el valor 4 (desacuerdo total).

Los resultados obtenidos luego de aplicar el test AQ-C fueron 111 puntos de 150, resultado superior a 76 puntos que es indicativo de una valoración diagnóstica especializada. Este trabajo tuvo como objetivo apoyar, mediante soluciones tecnológicas, el diagnóstico de autismo en niños. Para estos efectos, los criterios DSM fueron

cuantificados numéricamente, sirviendo de referencia para el diagnóstico de individuos con alta probabilidad de pertenecer al espectro autista TEA. Al utilizar mapas mentales, los supuestos conceptuales (criterios DSM) se hicieron explícitos, facilitando la especificación de taxonomías. En resumen, este trabajo contribuye a cuantificar criterios DSM referidos al autismo, siendo factible adicionar al mapa de conceptos valores óptimos que representan como un todo una caracterización de niños autistas.

Tabla 6: Respuesta Cuestionario AQ-C

T	Cromosoma														
	A1a	A1b	A1c	A1d	A2a	A2b	A2c	A2d	A3a	A3b	A3c	A3d	IM	STA	SDE
1		1													
2											2				
3								1							
4									3						
5														3	
6											2				
7				3											
8								1							
9														3	
10						1									
11	3														
12												4			
13		1													
14					3										
15															4
16										1					
17						1									
18					3										
19													3		
20														3	
21								1							
22		1													
23											2				
24			2												
25													3		
26						1									
27				3											
28												4			
29													3		
30														3	
31		1													
32										1					
33				3											
34			2												
35														3	
36															4
37									3						
38						1									
39							1								
40								1							
41													3		
42															4
43										1					
44			2												
45														3	
46										1					
47		1													
48															4
49															4
50	3														

Referencias

[1] Gallo, A et al. (2008). Dificultades en la evaluación y diagnóstico de autismo: Una discusión. Salud Mental 31(1), pp. 37-44.

[2] DSM (2000). Diagnostic and statistical manual of mental disorders. American Psychiatric Association. Washington: American Psychiatric Press.

[3] Baron, S., Allen. J. & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. The British Journal of Psychiatry 161(6), pp. 839-843.

[4] Goldberg, D. & Holland, J. (1988). Genetic algorithms and Machine Learning. Journal Machine Learning, Kluwer Academic Publishers 3(2-3), pp. 95-99.

[5] EspectroAutista (2015). Cociente de espectro autista, AQC. Test Interactivos.

[6] Auyeung, B. et al. (2008). The autism spectrum quotient: Children's version (AQ-Child). Journal of autism and developmental disorders 38(7), pp. 1230–1240.

[7] Hoekstra, R. et al. (2011). The construction and validation of an abridged version of the Autism-Spectrum Quotient (AQ-Short). Journal of Autism and Developmental Disorders 41(5), pp. 589-596.

[8] WordReference (2015). Online language dictionaries, Diccionario de sinónimos y antónimos.

[9] Perone, C. (2009). PYevolve Genetic Algorithms. PYevolve documentation. Online.

[10] Cuxart, F. (2000). El autismo: aspectos descriptivos y terapéuticos. Online.

[11] Baio, J. (2010). Prevalence of autism spectrum disorder among children aged 8 years — Autism and developmental disabilities monitoring network. USA: National Center on Birth Defects and Developmental Disabilities.

[12] EspectroAutista (2015). Recursos en línea para para los trastornos del espectro autista. Online.

[13] Departamento de Discapacidad y Rehabilitación (2011). Guía de Práctica Clínica de Detección y Diagnóstico Oportuno de los Trastornos del Espectro Autista, TEA. Gobierno de Chile – Ministerio de Salud.

[14] Adams, J. et al. (2012). Advice for parents of young autistic children. Autism Research Institute.

[15] Blogs Autismo (2015). Autismo de la A a la Z. Indice de blogs y foros de habla hispana sobre autismo.

[16] Buzan ,T. & Buzan B. (1996). The mind map book: How to use radiant thinking to maximize your brain's untapped potential. London: Plume Books.

[17] Bernier, R. & Gerdts, J. (2010). Autism spectrum disorders: a reference handbook. USA: ABC-CLIO.

[18] Holland, J. (1992). Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. Massachusetts: American Psychological Association.

[19] Sociología Contemporánea (2015), Once aplicaciones gratuitas para crear mapas mentales. Aplicaciones Online.

[20] Python Software Foundation (2015). Python. Online

[21] XMind (2015). XMind. Online.

[22] Pictogram (2011). Pictogram Room. España: Universidad de Valencia & Fundación Orange.

[23] AZAHAR (2013). Aplicaciones AZAHAR. España: Universidad de Valencia & Fundación Orange.

[24] Proyecto SPecial QR (2011). Aplicaciones SPQR. España: Fundación Orange y BJ Adaptaciones.

[25] Autismodiario.org (2013). Potenciando la comunicación y el lenguaje gracias a la tecnología. Autismo Diario. Online.

[26] Wang, S. (2014). Los autistas son los empleados ideales para ciertos puestos en firmas de software. The Wall Street Journal. Online.

[27] Seshadri, S. (2012). ¡Pad gives voice to kids with autism. Special to CNN.

[28] Rabello, R. (2010). Interacao e autismo: Uso de agentes inteligentes para detectar déficits de comunicacao em ambientes síncronos. Brasil: Universidad Federal do Rio Grande do Sul.

[29] Lozano, J. et al. (2011). Software para enseñar emociones al alumnado con trastorno del espectro autista. España: Universidad de Huelva.

[30] Scassellati, B. et al. (2012). Robots for Use in Autism Research. Annual Review of Biomedical Engineering 14, pp. 275-29.

[31] Duch, W. et al. (2012). Computational Approach to Understanding Autism Spectrum Disorders. Computer Science, 13(2), pp. 47-61.

[32] Ahmadlou, M. et al. (2012). Fuzzy Synchronization Likelihood-wavelet methodology for diagnosis of autism spectrum disorder. Journal of Neuroscience Methods, 211(2), pp. 203–209.

[33] Stahl, D. et al. (2012). Novel Machine Learning Methods for event-related potential Analysis: A Validation From Research on Infants at Risk for Autism. Developmental Neuropsychology 37(3), pp. 274-298.

[34] Pellicano, E. & Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. Trends in Cognitive Sciences 16(10), pp. 504–510.

[35] Santen, J. et al. (2013). Quantifying Repetitive Speech in Autism Spectrum Disorders and Language Impairment. Autism Research 6(5), pp. 372–383.

[36] Predictive Analytics Today (2015). Predictive Analytics &/ Text Analytics. Online.
[37] Semantria (2015). Semantria for Excel Quickstart. Semantria by Lexalytics. Online.
[38] ARC (2015). Downloadable Tets. University of Cambridge. Online.

Reutilización de la investigación científica en Ingeniería Civil usando protocolos experimentales semánticos

Ingrid Torres
Jaime Guzmán L.
Juan Muñoz F.
Universidad Nacional de Colombia – Colombia

1 Introducción

E-Science es una nueva cultura [21] que busca aprovechar completamente los resultados del pasado que han sido previamente logrados desde la investigación [4]. La idea de reutilizar la investigación se origina de la tradicional forma de intercambiar los paquetes de información completos, relacionados con la actividad científica [16]. Entre los paquetes de información más comunes aparecen los protocolos experimentales [10, 26] que, como documentos físicos, son direccionados por una guía de laboratorio documentada en un cuaderno, donde se registran objetivos, materiales, tareas, datos, fechas, resultados, entre muchos otros. Sin embargo, los esfuerzos por hallar información y comprender e interpretar a otros investigadores a partir de los cuadernos, constituyen una tarea altamente desgastante, porque no todos usan la misma organización o vocabulario.

Con el ánimo de reutilizar las investigaciones previamente logradas, pero con el propósito de alivianar las tareas de los seres humanos-investigadores, se hace necesario normalizar en un modelo compartido [6] las prácticas de laboratorio reportadas por los protocolos experimentales. A la fecha existen comunidades académicas que centran su interés en normalizar tales actividades en distintos campos, tales como la genética, la bioinformática, la biología y la astronomía. Sin embargo, las especificaciones existentes de materiales, equipos, objetivos y demás aspectos asociados al protocolo experimental, deben ser atadas al contexto de la investigación. No es lo mismo una investigación en el ámbito de la genética que en el de astronomía. Conscientes de tal limitante, este trabajo otorga inclusión al campo de la ingeniería civil, permitiendo definir la manera en que deben ser modelados los protocolos experimentales, descritos en guías físicas y documentados en cuadernos de laboratorio. Esta documentación resulta clave para realizar control, generar confiabilidad y/o brindar aseguramiento de la calidad de áreas tales como: concreto y pavimento, estructuras, hidráulica, suelos y geología, solamente por citar algunas.

Para dar un mayor detalle de la propuesta, este trabajo se organiza de la siguiente manera: la sección 2 presenta el marco conceptual de la temática; la sección 3 recopila los principales elementos que describen la representación semántica de un protocolo experimental en el campo de la ingeniería civil; la sección 4 presenta un análisis del caso específico del protocolo de humedad de suelos; la sección 5 expone una discusión concluyente y la perspectiva futura, para finalmente presentar las referencias de los trabajos que de una u otra forma hicieron parte inspiradora de este artículo.

2 Marco Conceptual

2.1 Protocolos Experimentales

Un protocolo representa una secuencia de operaciones y habitualmente son escritos en lenguaje natural. Estos protocolos se presentan a manera de *receta* proporcionando

paso a paso la descripción de procesos. En la investigación experimental esta secuencia de tareas y operaciones son consideradas unidades fundamentales de conocimiento [10]. Los investigadores siguen y generan protocolos en sus actividades diarias. Es así como los protocolos experimentales reflejan las habilidades o conocimientos prácticos generados en un laboratorio, que luego son compartidos y adaptados para diversos propósitos, como piezas centrales en los esfuerzos de reproducibilidad. Un protocolo es una especificación planeada que tiene suficiente nivel de detalle y la información cuantitativa para comunicarla entre los expertos de dominio, para que sean capaces de reproducir el proceso de manera independiente [14].

Actualmente, los protocolos experimentales de muchas áreas (entre ellas la civil), están registrados en su gran mayoría en unos documentos físicos, representados por una guía y un cuaderno de laboratorio. La guía se refiere a un documento escrito que contiene instrucciones sobre cómo hacer ciertos procedimientos de laboratorio, con el propósito de conseguir resultados reproducibles. Los cuadernos constituyen un registro completo de todo el trabajo práctico experimental realizado y contienen toda la información necesaria (operaciones, hechos observados y conclusiones, entre otros) para que un investigador pueda reproducir el trabajo realizado. Al ser llevados de forma manual es de esperar que los cuadernos de laboratorio sean escritos de manera legible y ordenada. Estas condiciones operativas son primordiales, por ejemplo, en casos relacionados con reutilización o reproducibilidad del experimento. El uso de los documentos físicos de laboratorio también permite acelerar la estandarización de procesos, mantener un control de calidad y obtener resultados reproducibles [8]. Otras características, tales como portabilidad, facilidad de edición, bajo costo, entre otras, han hecho que el papel siga siendo el medio más utilizado para el registro de datos científicos. Sin embargo, el cuaderno de laboratorio tiene limitaciones [14]:

- *Pérdida de información*: la información en papel puede perderse cuando no se usa un tipo de características físicas mínimas ni funcional en un ambiente de laboratorio.
- *Búsqueda manual de la información*: a pesar de ser un registro actualizado, las búsquedas automatizadas no son posibles.
- *Dificultad para compartir la información*: la información almacenada en los cuadernos no puede ser compartida o distribuida fácilmente entre colegas. Fotocopiar es la manera más práctica, pero legalmente no es válido crear múltiples copias de los registros experimentales consignados en los documentos físicos. Además, si la información no está escrita de manera legible, en ausencia del autor de la información se dificultará la reproducibilidad del trabajo.
- *Captura de datos*: la captura de datos se hace de manera manuscrita. Por ejemplo, los resultados que soportan el cuaderno muchas veces se encuentran en archivos electrónicos, que deben ser impresos y pegados en el documento. Esta es una práctica muy común incluso en laboratorios modernos.
- *Capacidad limitada*: el papel está limitado por su capacidad en número de hojas. Esto hace imposible el incluir, por ejemplo, videos. Y tampoco es posible manejar el volumen de datos en bruto que producen algunos equipos de laboratorio de alta eficiencia, porque esta cantidad puede llegar a ser del orden de los *terabytes*.
- *Mayor tiempo en el planeamiento del trabajo*: se estima que en los laboratorios que emplean los cuadernos de papel los científicos invierten más tiempo en el planeamiento de su trabajo, porque gran parte de ese tiempo se puede atribuir a la escritura y re-escritura de protocolos similares.

- *La organización del cuaderno*: varía de usuario a usuario y no sigue un estándar. Una alternativa para mejorar este aspecto es la construcción de una tabla de contenido, sin embargo, este procedimiento es ignorado por la mayoría de investigadores por el tiempo que debe invertirse en ello.

Bajo las condiciones actuales, reutilizar o reproducir las actividades experimentales de algunos campos, tales como la ingeniería civil, obligan al interesado a invertir mucho tiempo para descifrar el lenguaje usado por cada autor en sus documentos (aunque pertenezcan a la misma temática), y asociarlo a su interés investigativo para poder especificar lo que resulta útil o no. Si a esto le sumamos la cantidad de documentos producidos por todos los investigadores, la tarea resulta tediosa, altamente costosa en recurso humano y muy lenta. Esta serie de desventajas exigen formular un modelo de representación formal que normalice la práctica de los protocolos experimentales de ingeniería civil, haciéndolo comprensible y reutilizable.

2.2 Reutilización

Cuando se habla de reutilización se trata de utilizar algo, bien con la función que desempeñaba anteriormente o con otros fines [17]. En el ámbito de la investigación científica experimental, el concepto resulta clave y altamente atrayente, puesto que permite a los investigadores contar con información verificable [6] en el tiempo, bien sea para reproducir, extender o modificar la investigación experimental original. Desde la investigación, la reutilización es importante porque combate redundancia y perpetúa los métodos *de eficacia probada*, además, transmite la buena práctica científica [1, 24]. También es beneficiosa cuando se requiere: 1) permitir la evaluación de los resultados; 2) validar el método que llevó a una nueva visión científica; o 3) servir como tutoriales, exponiendo cómo tomar ventaja de la infraestructura, los procesos o los datos existentes.

2.3 Ingeniería Civil

La Ingeniería Civil se define como una disciplina que emplea conocimientos de cálculo, mecánica, hidráulica y física para el diseño, la construcción y el mantenimiento de las infraestructuras emplazadas en el entorno, incluyendo carreteras, ferrocarriles, puentes, canales, presas, puertos, aeropuertos, diques y otras [19]. Existen varias ramas y sub-disciplinas dentro de esta ingeniería, sobre las que se analizan distintas características, comportamientos y propiedades de determinados elementos [22, 23], tales como: 1) análisis de estructuras, que se encarga de estimar la resistencia máxima de elementos sometidos a cargas vivas, cargas muertas y cargas accidentales (sismos, vientos, nieve, etc.); 2) análisis geotécnico, se encarga de estimar la resistencia entre partículas de la corteza terrestre de distinta naturaleza, granulometría, humedad, cohesión, y de las propiedades de los suelos en general, con el fin de asegurar la interacción del suelo con la estructura; 3) análisis hidráulico, se ocupa de la proyección y ejecución de obras relacionadas con el agua, sea para su uso, como en la obtención de energía hidráulica, la irrigación, canalización u otras, sea para la construcción de estructuras en mares, ríos, lagos, o entornos similares, incluyendo, por ejemplo, diques, represas, canales, puertos, muelles, rompeolas, entre otras construcciones. También hace referencia a las máquinas hidráulicas; 4) análisis de materiales, se encarga del estudio de las propiedades de los materiales utilizados en la construcción de obras civiles, tales como el concreto hidráulico, el concreto asfáltico, el acero y los polímeros, entre muchos otros. Cada una de las disciplinas implementa laboratorios y actividades experimentales que especifican la observación, el proceso y los resultados de los análisis de la actividad científica, soportados en protocolos experimentales que en gran medida son orientados

por las normas ASTM [3]. La Figura 1 muestra una recopilación de los principales laboratorios del área de ingeniería civil y los correspondientes protocolos experimentales asociados a ellos.

Laboratorio de Resistencia de Materiales:	3. **Laboratorio de Mecánica de Suelos**
1.1. Experimento Ensayo de tracción	3.1. Experimento Contenido de agua
1.2. Compresión experimento de prueba	3.2. Experimento Peso de la unidad
1.3. Experimento ensayo de corte	3.3. Experimento Gravedad Específica
1.4. Experimento de prueba de torsión	3.4. Grano Experimento distribución por tamaño
1.5. Experimento Prueba de Impacto	3.5. Experimento Límites Attenberg
1.6. Experimento ensayo de fluencia	3.6. Experimento Permeabilidad
1.7. Experimento Ensayo de fatiga	3.7. Experimento Compactación de prueba
1.8. Prueba Experimento de compresión	3.8. Experimento Prueba de Consolidación
1.10. Experimento Dureza Rockwell	3.9. Experimento ensayo de corte directo
	3.10. Experimento Prueba Triaxial
2. **Laboratorio Mecánica de Fluidos**	
2.1. El experimento de Bernoulli	4. **Laboratorio Análisis Estructural Básico**
2.2. Experimento Venturi Meter	4.1. Experimento Vigas de Palmo Simple
2.3. Experimento orificios	4.2. Experimento Vigas Continuas
2.4. Boquillas Experimento	4.3. Columna Experimento Análisis
2.5. Experimento Presas	4.4. Experimento Portal Marcos
2.6. Canales Experimento	4.5. Placas Experimento
2.7. Experimento de Reynolds	4.6. Experimento Articulaciones Rijid
2.8. Experimento de Vórtices	4.7. Experimento de Arcos
2.9. Experimento Jets	4.8. Experimento Cerchas
2.10. Turbinas Experimento	4.9. Experimento Muros
	4.10. Experimento rótula plástica

Figura 1: Laboratorios y protocolos experimentales de Ingeniería Civil

En los laboratorios, los protocolos experimentales aportan parte del método científico, validan la teoría y generan nuevos hallazgos [14]. Las principales prácticas de laboratorio del dominio de la Ingeniería Civil están agrupadas en las siguientes categorías:

1. *Laboratorio de materiales*. Se denomina ensayo de materiales a toda prueba cuyo fin es determinar las propiedades mecánicas de un material. En este laboratorio se estudia el comportamiento y propiedades de materiales de construcción comúnmente usados en aplicaciones de Ingeniería Civil, tales como concreto, acero, madera, piezas de mampostería y plásticos. Existen normas y estándares que describen estos materiales y los ensayos para determinar sus propiedades, y protocolos de laboratorio para el ensayo y estudio de diferentes materiales. La ciencia de los materiales abarcan las teorías de falla, el flujo plástico, la relajación, la fatiga, la concentración de esfuerzos, las vibraciones, entre otras.

2. *Laboratorio de mecánica de fluidos*. El objetivo del laboratorio es tratar el tema de los fluidos desde el punto de vista de sus propiedades físicas y su comportamiento mecánico. Estudia específicamente el comportamiento de los fluidos en diferentes situaciones de la Ingeniería, con base en la aplicación de las ecuaciones de la física y de métodos numéricos para facilitar los cálculos. Cita protocolos experimentales de énfasis en las pérdidas por fricción y su efecto sobre el diseño de sistemas de ingeniería relacionados con el manejo del agua. Cubre también análisis experimentales de la estática de los fluidos, la cinemática, las propiedades y el comportamiento de los fluidos reales (esfuerzos cortantes, distribución de velocidades, pérdidas de energía), análisis dimensional y similitud dinámica, flujo en sistemas de tuberías a presión, diseño de tuberías y aplicaciones.

3. *Laboratorio de mecánica de suelos*. Se ocupa de valorar la composición y características de los suelos. Tiene como fin llegar a explicaciones físicas y descripciones matemáticas del comportamiento del suelo, cuando es sometido a cargas monotónicas, haciendo especial énfasis en la respuesta no lineal del suelo. En este laboratorio se registran protocolos experimentales de clasificación del suelo, tales como determinación de humedad, gravedad específica, tamaño del grano, por citar algunos, además del estudio del comportamiento bajo determinadas cargas, donde se reportan protocolos para el estudio de compactación, consolidación y cortes.

4. *Laboratorio de análisis estructural básico*. Es orientado sobre los sistemas estructurales de concreto, las normas, reglamentos y los códigos que rigen su diseño. Los conceptos de seguridad estructural, cargas de diseño, comportamiento y las premisas de diseño de concreto reforzado, propiedades físicas y mecánicas del concreto y del acero de refuerzo, fundamentos de la construcción, también se incluyen. En él se hallan protocolos para el análisis de vigas, columnas, marcos, placas, articulaciones, entre muchos otros.

3 Protocolo experimental en Ingeniería Civil

Como se ha citado recurrentemente, un protocolo experimental en cualquier disciplina de la ingeniería Civil corresponde a una práctica de laboratorio, direccionada por una guía y en su gran mayoría soportada por norma (ASTM). La Figura 2 muestra un ejemplo de un protocolo experimental: *determinación de contenido de humedad*, asociado al laboratorio de suelos. Se destacan las fórmulas matemáticas, un listado de los equipos y materiales de consumo, una descripción enumerada de los pasos que conforman el procedimiento, un formato para registro de resultados, un ítem de presentación de conclusiones, y finalmente una descripción de observaciones.

LABORATORIOS DE INGENIERÍA CIVIL
PROGRAMA DE INGENIERÍA CIVIL
LABORATORIO DE COMPORTAMIENTO DE SUELOS
Código: FESA ITLCS P01 Fecha de emisión: 2013.10.24 Revisión: 00

PRÁCTICA No. 1
DETERMINACIÓN DEL CONTENIDO DE HUMEDAD

1. Objetivo
El alumno determinará el contenido de agua o humedad de varias muestras de suelo.

Es importante la comprensión adecuada de esta práctica para poder entender el comportamiento mecánico de los suelos, además, es necesario conocer cómo se determina, ya que en varias ocasiones del curso se obtiene, como parte de otros ensayes de laboratorio.

2. Introducción teórica
Podemos decir que suelo es todo tipo de material terroso, desde un relleno de desperdicio, hasta areniscas parcialmente cementadas o lutitas suaves. Dentro de los suelos no se encuentran incluidas las rocas sanas de origen ígneo o metamórfico, y los depósitos sedimentarios altamente cementados, que no se ablanden o desintegren rápidamente por acción de la intemperie.

Además el agua contenida juega un papel tan fundamental en el comportamiento mecánico del suelo, que se debe considerarse como parte integral del mismo.

Para entender que es el contenido de agua resulta necesario dibujar el esquema representativo de suelos:

Se conoce como contenido de agua o humedad de un suelo, a la relación entre el peso del agua contenida en el mismo y el peso de su fase sólida, suele indicarse en porcentaje y se expresa con la siguiente fórmula:

$$w(\%) = \frac{Ww}{Ws} \times 100$$

De la observación de la expresión se indica que puede varía teóricamente de 0 a ∞

En la naturaleza la humedad de los suelos varía entre límites muy amplios. En arcillas japonesas se han registrado contenidos de agua de 1,200 a 1,400%, si bien estos valores son excepcionales.

Este documento es propiedad de los Laboratorios de Ingeniería Civil de la Facultad de Estudios Superiores Acatlán, se prohíbe la reproducción parcial o total sin la autorización correspondiente.
FESA GC P01 F04.0
Página 4 de 82

LABORATORIOS DE INGENIERÍA CIVIL
PROGRAMA DE INGENIERÍA CIVIL
LABORATORIO DE COMPORTAMIENTO DE SUELOS
Código: FESA ITLCS P01 Fecha de emisión: 2013.10.24 Revisión: 00

4. Procedimiento
1. Tomar un recipiente, identificarlo con un número y enseguida pesarlo (puede ser una cápsula de porcelana o un recipiente de vidrio o aluminio), obteniendo W_c.

2. Colocar en el recipiente una porción representativa de suelo, la cantidad de material estará en función del tamaño de las partículas que lo constituyan.

3. Pesar el recipiente con la muestra de suelo, en este caso esta el suelo húmedo, obteniéndose Wrsw.

Este documento es propiedad de los Laboratorios de Ingeniería Civil de la Facultad de Estudios Superiores Acatlán, se prohíbe la reproducción parcial o total sin la autorización correspondiente.
FESA GC P01 F04.0
Página 7 de 82

Figura 2: Ejemplo de un protocolo experimental, para el laboratorio de suelos [22]

En esta etapa se aprecia que el protocolo experimental se compone de dos elementos claves: una estructura documental, que representa los pasos del método científico y una descripción del paso a paso, escrito mediante un flujo de procesos (conocido como *workflow* científico) [5, 12]. La descripción anterior es fiel reflejo de la presentada en *SMART Protocol* [15], diseñada para el dominio de la biotecnología de las plantas vegetales. El objetivo es modelar el protocolo experimental en dos partes, separando el modelo de información estático del modelo dinámico. El primero representa un documento científico que registra la información básica de la metodología de la investigación, y el segundo el protocolo como un flujo de procesos, comúnmente llamado *workflow* (científico en este caso). En esta representación es posible modelar la guía del laboratorio como la especificación abstracta del proceso experimental (*workflow*

científico), que debe seguirse para alcanzar los objetivos, propósitos y/o resultados reportados por ese experimento (documento científico).

Ahora bien, si se trata de reutilizar el conocimiento descrito en los protocolos experimentales como documento y como *workflow*, resulta clave la comprensión común de la estructura de información entre personas y agentes de software [9]. En otras palabras, se habla de generar una estructura semántica que soporte los aspectos claves de la reutilización sobre ontologías, y que para este caso se concentra en la especificación del conocimiento de los protocolos experimentales del dominio de la Ingeniería Civil.

4 Representación Semántica

La descripción formal de los experimentos es necesaria para lograr un análisis eficiente de la investigación y permitir el intercambio de resultados científicos, constituyéndose en una parte fundamental de la práctica de la ciencia. Representar semánticamente el dominio de los protocolos experimentales del campo de la ingeniería civil permite contar y modelar la investigación científica del área. Tal representación debe ser univoca, consensuada y contener los elementos comunes, necesarios y suficientes para lograr reutilizar los experimentos, sin importar quién o para qué se requieren.

Con el objetivo de proponer una representación semántica que soporte las necesidades de los protocolos experimentales en Ingeniería Civil, se propone implementar una ontología. Para desarrollar tal implementación se acude a la guía metodológicas más comúnmente aceptada, conocida como *Ontology Development 101* [16], que recomienda: 1) determinar el dominio y ámbito de la ontología, 2) determinar la intención de uso de la ontología, 3) reutilizar ontologías o vocabularios controlados existentes, 4) enumerar los términos importantes del dominio, 5) definir jerarquía de clases, y 6) crear las instancias para determinar las actividades que deben considerarse para generar una nueva ontología.

Aunque aparentemente la cuestión de determinar el dominio estaba clara, el tema de los protocolos experimentales en Ingeniería Civil presenta una amplia categorización de elementos y conceptos, razón por la cual establecer un límite de profundización para cada rama fue una tarea muy cuidadosa, aunque poco justa, porque cada rama podía generar otra ontología completa. La definición de cada categoría temática fue el reflejo de la labor de consenso llevada a cabo entre el análisis de varios protocolos experimentales (basados en la descripción que se definió en la Figura 3), las normas ASTM y los expertos en el tema (ingenieros e investigadores). Como resultado, las áreas temáticas en esta ontología se limitan a la definición de categorías genéricas por elementos constitutivos del protocolo como documento y como *workflow*. La intención de uso de la ontología era de las tareas más claras hasta ahora: la clasificación y reutilización del conocimiento experimental que se genera en los laboratorios del área de Ingeniería Civil. Por ello, la recomendación en la que más se centró la atención fue precisamente en la reutilización de ontologías existentes, así que a la hora de considerar la reutilización de estándares de tipo ontológico en este campo se siguieron propuestas bajo un esquema de modelar experimentos, similares a SMARTP [14], EXPO [19], EXACT [20] y P-PLAN [13].

Es decir, adoptar ontologías de existencia previa que contribuyeran a definir los conceptos más importantes y comunes del dominio experimental. Sin embargo, algunos conceptos resultan repetidos, mientras que otros pueden ser insuficientes o incluso innecesarios para modelar el conocimiento del dominio. Dadas las circunstancias, la primera actividad de esta etapa consistió en determinar los elementos registrados en los

contenidos de los protocolos. Esta tarea estuvo inspirada en la formulación de la Figura 3, en la cual se estableció el detalle de cada protocolo del área de la Ingeniería Civil y su representación bajo la composición de las dos partes: una como documento y otra como *workflow*.

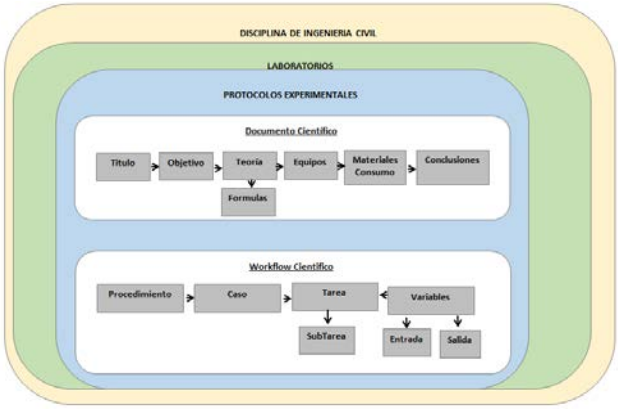


Figura 3: Protocolo experimental como documento y como *workflow* científico

Para enumerar los términos más importantes del dominio se decidió representar cada elemento abstracto de los principales protocolos de la Ingeniería Civil como instancias de la representación, como se observa en la Figura 3. Generándose con ello la imagen que se observa en la Figura 4, que proporciona elementos instanciados para cada clase de las ontologías experimentales reutilizadas. El proceso señalado es repetido para cada uno de los 40 protocolos de la Figura 1. La ontología construida se encuentra conformada por los principales conceptos jerarquizados, asociados a los protocolos de Ingeniería Civil y tal como se muestra en la Figura 5. Sin embargo, se han importado conceptos de las ontologías descritas en la Tabla 1, además de algunas otras, tales como OBI [11] y BFO [2], permitiendo instanciar los elementos asociados a cada protocolo, ya sea como documento o como *workflow*. Sin embargo, estas no han sido las únicas ontologías importadas y dado que resulta inminente la necesidad de que las máquinas comprendan conceptos claves como las fórmulas matemáticas, se deberá acudir a ontologías de sintaxis aritméticas que establezcan los elementos básicos para los cálculos sobre expresiones aritméticas elementales tales como suma, resta, multiplicación, división y sumatoria.

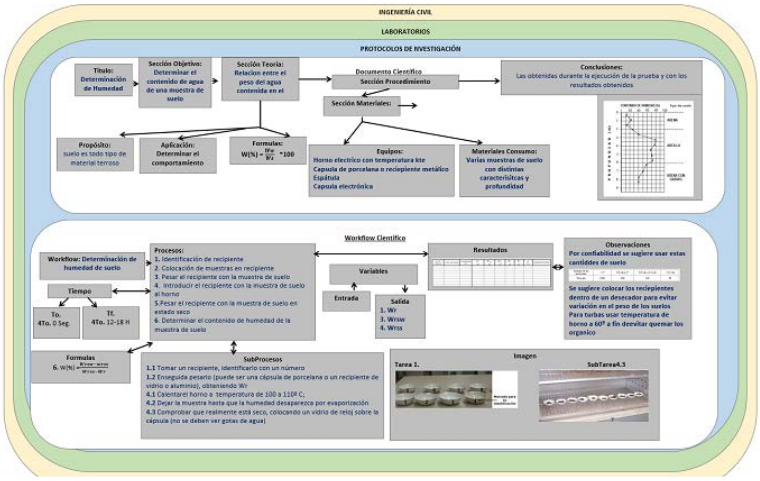


Figura 4: Protocolo experimental como documento y como *workflow* científico

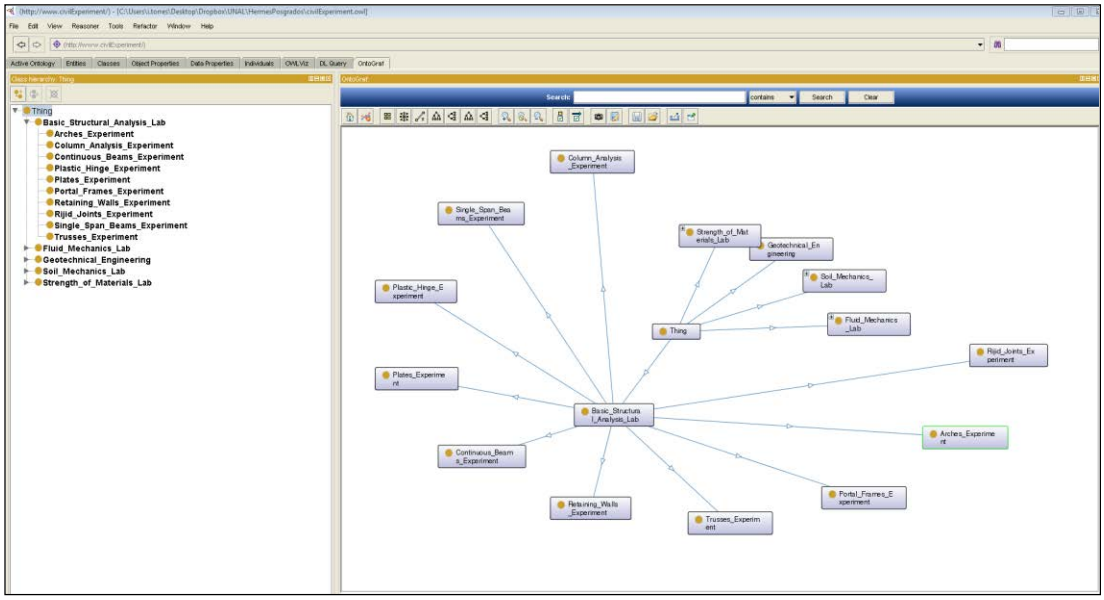


Figura 5: Ontología de Protocolos experimentales en Ingeniería Civil

Tabla 1: Ontologías para representar experimentos

Ontología	Alcance	Términos a reutilizar	Propósito
SMARTP [14]	Biotecnología Vegetal	*sp:protocol* *iao:document, iao:document_part* *sp:introduction_section* *sp:purpose* *sp:application_proco* *sp:materirals_section* *sp:primer_list* *sp:equipment* *sp:supplies_list* *sp:methods*	Describir el dominio de un protocolo como documento. Como workflow, usa P-Plan
EXPO [19]	Bioinformática	*ex:ScientificExperiment* *ex:AdminInfoAboutExp* *ex:TittleOfExperiment* *ex:IDExperiment* *ex:Autor* *ex:Organization* *ex:StatusofDocument* *ex.public* *ex:academic* *ex:Hypothesys* *ex:Results* *ex:Biblioreference* *ex:classificationOfExperiments* *ex:Experimental Design* *ex:SubjectOfExperiment* *ex:ObjectOfExperiment* *ex:ExperimentalEquipment* *ex:ExperimentalTechnology* *ex:PlanExperimentalAction*	Se utiliza para formalizar el conocimiento genérico sobre el diseño científico experimental, la metodología y los resultados de la representación
EXACT [20]	Biológica	*Action* *Assessment Action* *Experiment action* *Mode of Combination* *Mode of separation* *Mode of transformation* *Command action* *Equipment setup action*	Ontología de Acciones Experimentales (EXACT) como base de un método de representación de protocolos de laboratorio biomédicos.

P- PLAN [13]	Biogenética	*p-plan:Entity* *p-plan:Steps* *p-plan:Variable* *p-plan:Activity*	Escribe la procedencia de los objetos (prov: Entity) como un registro de afirmaciones acerca de los pasos (prov: Actividades) que los generaron y las entidades utilizadas.

5 Caso de estudio: Método de prueba estándar para Laboratorio de determinación de agua (humedad) en contenido de suelo y roca

Este método de ensayo cubre el laboratorio de la determinación de agua (humedad) contenido en la masa del suelo, roca y materiales similares, donde la reducción de la masa por secado es debida a la pérdida de agua. Por simplicidad, la palabra *material* se referirá al suelo, la roca o el que sea agregado como más aplicable. Este ensayo involucra tres procesos: 1) selección de muestras de ensayos representativos según tipo, 2) definición de pesos, y 3) determinación de la masa de la muestra seca.

Proceso 1. "Selección de muestras de ensayos representativos"[1] según tipo. Este proceso es_una preparación de la muestra para un ensayo[2]. Cuando la muestra de ensayo es una porción de una cantidad más grande de material, la muestra debe ser seleccionada para ser representativa de la condición del agua de toda la cantidad de material. La manera en que se selecciona la muestra de ensayo depende del propósito y la aplicación de la prueba, el tipo de material que se está probado, la condición del agua y el tipo de muestra (de otra prueba, bolsa, bloque, etc.).

a) Si el material es tal que se puede manipular y ser manejado sin pérdida significativa de humedad y segregación, debe ser completamente mezclado. Se selecciona una parte representativa mediante una bola de un tamaño que se requieran no más de unos pocos *scoopfuls* para obtener el tamaño apropiado del espécimen, como se define en Tabla 1.

   i) Condicional *IF* "*material manipulable*"[3] es_una_sub-clase *"preparación de la muestra para un ensayo*".
   ii) "*scoopfuls*" es_un "suministro de laboratorio"[4]
   iii) "Metodo A" OR "Metodo B"

b) Si el material no se puede mezclar es porque no puede ser completamente mixto o mezclado, entonces que muestrearlo por una cuchara y formar una reserva mezclando tanto como sea posible. Tomar por lo menos cinco porciones de material en lugares aleatorios usando un tubo de toma de muestras, una pala, una paleta, o un dispositivo similar adecuado al tamaño máximo de partícula presente en el material. Combinar todas las raciones para la muestra de prueba.

   i) Condicional *IF NOT* ("*material manipulable*")[5] es_una_subclase *"preparación de la muestra para un ensayo*".
   ii) "*pala*" es_un "suministro de laboratorio"[6]
   iii) "Metodo A" OR "Metodo B"

c) Si el material o las condiciones son tales que no se puede formar, entonces hay que tomar varias porciones del material utilizando lugares aleatorios para representar mejor la condición de humedad. Combinar todas las partes de la prueba de la muestra.

[1] Ontology civil: suministro de laboratorio
[2] OBI Ontology Biomedical Investigación
[3] Ontology civil: suministro de laboratorio
[4] Ontology civil: suministro de laboratorio
[5] Ontology civil: suministro de laboratorio
[6] Ontology civil: suministro de laboratorio

i) Condicional *IF* NOT ("*material manipulable*")[7] **es_una_subclase** *"preparación de la muestra para un ensayo*".
ii) "Metodo A" OR "Metodo B".

d) "muestra de ensayo"[8] es_una "variable_de_salida" [9]

Proceso 2. "Determinación de la masa de la muestra seca"[10] es_precedida[11] por "Selección de muestras de ensayos representativos."[12]

a) "Muestra de ensayo" es_una "variable_de_entrada".

i) "muestra de ensayo" tiene_peso[13].
ii) peso tiene *data_property* (int).

b) "recipiente" [14] es_un "suministro de laboratorio"[15].

i) "recipiente" tiene_peso peso (int).

c) "tapa de recipiente" [16] es_un "suministro de laboratorio"[17].

i) "tapa de recipiente" tiene_peso peso (int).

6 Conclusiones

En este trabajo se ha demostrado la viabilidad modelar los protocolos experimentales como representaciones semánticas, originadas a partir de su caracterización de dos elementos: 1) como parte estática, denominada *documento* y 2) como parte dinámica, identificada como *workflow*. La propuesta aunque inspirada en *SMART Protocol* otorga detalles del proceso de representación semántico, a partir de las prácticas experimentales del dominio de Ingeniería Civil. Este modelo puede constituirse en una guía de trabajo para otros campos de investigación, aunque es evidente la dependencia entre el contexto del dominio y la actividad experimental que se pueda dar en él.

También resulta clave la reutilización de ontologías superiores. Hasta ahora se han identificado las ontologías que pueden ser reutilizadas en la implementación de los protocolos del área de Ingeniería Civil y se han diseñado las estrategias de asociación. Se incluyen algunos apartes de los protocolos analizados a la fecha. Sin embargo, resulta necesario trabajar más en la orquestación de todas las ontologías que pueden ser reutilizadas, apoyadas especialmente en elementos de lógicas descriptivas que faciliten la comprensión e interpretación de las máquinas, generando procesos de razonamiento e inferencia.

Referencias

[1] Alper, P. et al. (2013). Small is beautiful: Summarizing scientific workflows using semantic annotations. Proceedings IEEE International Congress on Big Data (pp. 318-325). June 27 – July 2, Santa Clara, USA.

[2] Arp, R. & Smith, B. (2008). Function, role, and disposition in basic formal ontology. Nature 2(1-4), pp. 1-4.

[7] Ontology civil: suministro de laboratorio
[8] Ontology civil: suministro de laboratorio
[9] P-Plan isPreceededBy
[10] BFO: Basic Formal Ontology
[11] P-Plan isPreceededBy
[12] BFO: Basic Formal Ontology
[13] Manchester OWL Arithmetics Syntax
[14] BFO: Basic Formal Ontology
[15] Ontology civil: suministro de laboratorio
[16] BFO: Basic Formal Ontology
[17] Ontology civil: suministro de laboratorio

[3] ASTM (2015). http://www.astm.org/
[4] Bechhofer, S. et al. (2010). Research objects: Towards exchange and reuse of digital knowledge. Nature 2(5), pp. 1-6.
[5] Bechhofer, S., Soiland, S. & Belhajjame, K. (2011). workflow lifecycle management initial requirements. Technical Report STREP FP7-ICT-2007-6 270192. University of Manchester.
[6] Belhajjame, K. (2007, December). Semantic replaceability of eScience web services. In e-Science and Grid Computing. Proceedings IEEE International Conference on e-Science and Grid Computing (pp. 449-456). December 10-13, Bangalore, India.
[7] Belhajjame, K. et al. (2012). Workflow-centric research objects: First class citizens in scholarly discourse. Proceedings Workshop on the Future of Scholarly Communication in the Semantic Web (pp. 1-12). May 10-12, Heraklion, Greece.
[8] Belhajjame, K., Goble, C. & De Roure, D. (2012). Research object management: opportunities and challenges. Proceedings Data Intensive Collaboration in Science and Engineering workshop (pp. 1-4). February 11–15, Seattle, Washington.
[9] Berners, T., Hendler, J. & Lassila, O. (2001). The semantic web. Scientific American 284(5), pp. 28-37.
[10] IngGroup (2015). E-ciencia semántica. Ontology Engineer IngGroup. Online.
[11] Courtot, M. et al. (2008). OBI: The OWL of biomedical investigations. Proceedings Workshop in the International semantic web conference (pp. 1-12). October 26-27, Karlsruhe, Germany.
[12] Deelman, E. et al. (2009). Workflows and e-Science: An overview of workflow system features and capabilities. Future Generation Computer Systems 25(5), pp. 528-540.
[13] Garrijo, D. & Gil, Y. (2012). P-PLAN Ontology.
[14] O. Giraldo. (2011). Manejo del Conocimiento en los cuadernos de laboratorio. Tesis de Maestría, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, sede Palmira.
[15] O. Giraldo. (2014). SMART Protocols Ontology. Online.
[16] Noy, N. & McGuinness, D. (2001). Ontology development 101: A guide to creating your first ontology. Knowledge Systems Laboratory Stanford University. Online.
[17] RAE (2015). Real Academia Española. Diccionario. Madrid, España.
[18] De Roure, D. et al. (2009). The myexperiment open repository for scientific workflows. Proceedings 4th International Conference on Open Repositories (pp. 1-4). May 18-21, Atlanta.
[19] Soldatova, L. & King, R. (2006). An ontology of scientific experiments. Journal of the Royal Society Interface 3(11), pp. 795-803.
[20] Soldatova, L., Aubrey, W. & King, R. (2006). The EXACT description of biomedical protocols. Oxford JournalsScience & Mathematics Bioinformatics 24(13), pp. 295-303.
[21] Sun, Y. et al. (2011). Research in e-science: Current status and future direction. International Journal of Computer Systems Science & Engineering 26(3), pp. 151-152.
[22] UNAM (2014). Laboratorios de Ingeniería Civil. Programa de Ingeniería Civil. Manual de prácticas de Laboratorio.
[23] Pasmin, G. & Ximena, O. (2011). Manejo del conocimiento en los cuadernos de laboratorio. Disertación Doctoral. Universidad Nacional de Colombia Sede Palmira.
[24] Verdes, L. (2012). e-Science for the SKA, WF4Ever: Supporting Reuse and Reproducibility in Experimental Science. RadioNet Advanced Radio Astronomy, Commissioning Skills and Preparation for the SKA. Manchester, UK.
[25] Yang, X. et al. (2011). Guide to e-Science: next generation scientific research and discovery. USA: Springer.
[26] Zhao, J. et al. (2012). RO-Manager: A Tool for Creating and Manipulating Research Objects to Support Reproducibility and Reuse in Sciences. Proceedings Second International Workshop on Linked Science - Tackling Big Data (pp. 1-4). November 12, Boston, USA.

Desarrollo de una aplicación móvil para la localización en espacios interiores por medio de análisis de radiación de redes WiFi

Carlos Gómez R.
Fundación Universitaria Agustiniana – Colombia

1 Introducción

Actualmente existe una amplia oferta de servicios tecnológicos brindados de forma geolocalizada, es decir, personalizados de acuerdo con la ubicación geográfica del usuario (continente, país, ciudad y calle). Sin embargo, hasta el momento estos servicios se prestan a grandes escalas donde la resolución de la ubicación puede ser un barrio completo o varias calles. La base tecnológica de estos servicios es la localización basada en tecnologías aplicables a escala mundial y de uso en espacios exteriores, como es el caso del Sistema de Posicionamiento Global GPS. Todavía no existe una alternativa estandarizada y de fácil acceso para obtener la localización de dispositivos en interiores, que además ofrezca mayor precisión que los GPS comerciales, tal como se requiere en la localización capaz de ubicar usuarios en pisos, salas y cuadrantes específicos. Por eso es necesario diseñar una solución que aproveche las redes WiFi masivamente desplegadas en entornos urbanos, y que no dependa del desarrollo de nuevas redes o de la instalación de nuevos dispositivos, como una alternativa sustentable y de fácil implementación.

Con la popularización de los dispositivos móviles se han realizado trabajos para crear formas de localización de dispositivos en interiores. Se han estudiado las variables que contienen información relativa a esta localización: 1) el ángulo en que llega la señal (*direction of arrival*, DOA), para estimar la localización relativa de un dispositivo respecto de otro; 2) el instante en que llega la señal (*time of arrival*, TOA), con la que se puede estimar el tiempo que toma la señal para propagarse y así estimar la distancia que separa un dispositivo de otro; y 3) la potencia con la que la señal se recibe (*received signal strength indication,* RSSI), que también sirve para estimar la distancia que separa los dispositivos[1].

DOA y TOA en comparación con RSSI, presentan características de desempeño sensibles a la disponibilidad de la línea de vista (*line of sight*, LOS), por lo que las técnicas de localización basadas en ellas en escenarios sin línea de vista (*non line of sight,* NLOS) están sujetos a errores considerables. El desempeño de RSSI se altera medianamente por la falta de LOS. Los escenarios NLOS pueden generar efectos inesperados en relación con la función potencia-distancia, que pueden ser reducidos aplicando técnicas de filtrado a las mediciones. Por eso es que muchas de las técnicas de identificación, mitigación y localización en escenarios NLOS se diseñan con base en la variable RSSI [2].

Otro de los problemas de la localización basada en el tiempo o ángulo de llegada (TOA y DOA), es la necesidad de instalar hardware adicional y especializado sobre la red WiFi existente, así como la necesidad de un sistema especial de sincronismo. Para el caso de la localización por DOA se requieren arreglos especiales de antenas directivas de alto costo. Todo esto, sumado a la necesidad de un escenario de línea de vista para el uso de TOA y DOA, las hace inviables para ubicar dispositivos móviles en ambientes interiores [3].

El enfoque de este trabajo se basa en la premisa de usar las posibilidades de las tecnologías ampliamente usadas, tales como los teléfonos inteligentes y las tabletas con interfaces WiFi, y las redes WiFi desplegadas masivamente en los entornos urbanos aunque no se posea control y administración sobre ellas. Se hace especial énfasis en la

implementación de una metodología estudiada, sobre una plataforma móvil de un teléfono inteligente, para centrar allí todas las funciones de captura de datos, análisis estadístico y localización de usuarios.

2 Medición de radiación Wifi para localización de dispositivos

El estudio del estado del arte acerca de la localización de dispositivos por redes WiFi permitió identificar los retos que esto implica y algunas de las metodologías planteadas para afrontarlos. Los ambientes interiores se caracterizan porque las ondas de radiofrecuencia se propagan por múltiples caminos, debido a los fenómenos físicos asociados y a la tecnología WiFi, tales como difracción, absorción y reflexión de ondas electromagnéticas, disposición y cantidad de antenas en los equipos activos, etc. Estos ambientes también tienen obstáculos para la comunicación directa entre equipos, constituyendo escenarios NLOS [5]. Estas condiciones, difíciles de controlar y ajenas al diseño tecnológico, constituyen escenarios de incertidumbre y comportamientos aleatorios, donde los esfuerzos no están orientados a modelar la radiación de WiFi en interiores, sino en medirla y analizar su comportamiento. Muchas de las técnicas de identificación, mitigación y localización en escenarios NLOS han sido diseñadas basándose en la variable RSSI [2].

Por otro lado, los cambios en las condiciones atmosféricas como la temperatura también puede afectar la propagación de las ondas de radio y su correspondiente potencia con la que se recibe la señal RSSI. Además, la frecuencia de 2.4GHz que ha sido ampliamente usada para el desarrollo de redes WiFi, es la misma de la resonancia del agua, y debido a esto las personas absorben ondas electromagnéticas en esta banda [4]. Esto implica que factores como su presencia aumentan la incertidumbre en las posibles predicciones de radiación en interiores y se convierten en una fuente de errores.

Debido a la compleja naturaleza y el comportamiento de la propagación de ondas de radiofrecuencia en espacios interiores, su difícil predicción y su comportamiento inestable, no es posible confiar que el nivel de potencia RSSI corresponda directamente con la distancia de separación entre los dispositivos WiFi, condición necesaria para aplicar técnicas de localización como la triangulación. Por eso es que la técnica más usada en espacios internos para la localización es la que se basa en el estudio de la recepción de la potencia de señal RSSI, proceso llamado *fingerprinting*. Los *access points* de las redes WiFi y las fuentes de información relativa a la localización, originadores de las señales WiFi y por ser los transmisores que permiten medir el RSSI, facilitan su análisis abierto y eliminan la necesidad de costos de infraestructura adicional [6].

El primer trabajo relacionado con esta técnica fue el proyecto RADAR [7], y en [8] se propone una metodología para detectar y reducir la distancia de error hasta un promedio cercano a 1 metro lineal, que se basa en las dos etapas de la metodología *fingerprinting*, es decir, levantar la información de radiación de WiFi en un espacio de interiores por medio de colectar los datos de potencia de cada una de las radiaciones incidentes en el espacio, y realizar un análisis estadístico centrado en la caracterización de las señales presentes y en la detección de muestras atípicas características de los ambientes con fenómenos fortuitos, como las redes WiFi. De esta manera se construyen mapas que caracterizan las huellas de radiación de cada red WiFi. Luego se realiza la etapa de localización en la que con las huellas de radiación se hace la comparación en distancia euclidiana de una muestra aleatoria, para determinar la mayor probabilidad de localización de un dispositivo WiFi [8]. Estos filtros logran reducir la distancia de error que presentan métodos similares.

3 Desarrollo del aplicativo móvil

Tomando como base lo propuesto en [8], se fija un espacio en interiores como lugar de pruebas del aplicativo móvil a desarrollar. El ambiente de la Figura 1 corresponde a un entorno urbano de contexto residencial, con alta presencia de redes WiFi hogareñas, una gran cantidad de obstáculos para la radiación de las señales WiFi y elementos como hornos microondas, teléfonos inalámbricos y dispositivos *bluetooth*.

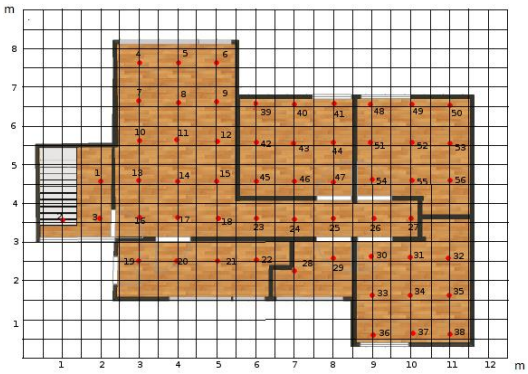


Figura 1: Plano de espacio estudiado y ubicación de los puntos de medición aplicados

El prototipo de aplicación se basó en el sistema operativo de dispositivos móvil Android, debido a su flexibilidad y porque se encuentra en la mayoría de los dispositivos móviles inteligentes. El aplicativo Lazarillo WiFi 1.0 permite importar el mapa arquitectónico del lugar donde se implementará la metodología de calibración *fingerprinting* [8]. Con la importación del plano, en formato de imagen, se especifican sus dimensiones en metros, como se muestra en la Figura 2.

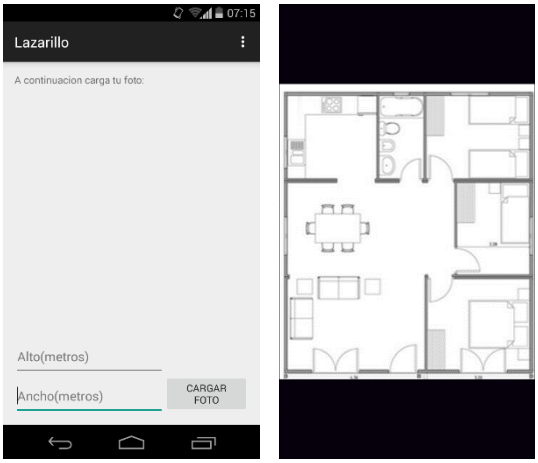


Figura 2: Interfaz de Lazarillo WiFi version 1.0

A partir del plano importado y la información de sus dimensiones, se aplica un algoritmo para calcular la cantidad de cuadrantes de 1m^2 existentes en la superficie, y posteriormente se implementa la segmentación de cada uno por medio de una grilla sobre el plano, creando así cuadrantes como se muestra en la Figura 3. Esta segmentación permite recopilar de forma ordenada y exacta la información de potencias RSSI por cada metro cuadrado.

La interfaz permite que al pulsar cada una de los cuadrantes de la grilla el dispositivo móvil capture la información de todas las redes WiFi incidentes en ese espacio. Aquí se presta atención a la potencia RSSI y a la dirección MAC del *access point* que la radia,

además de permitir que se introduzca un texto descriptivo de esa ubicación. Al mismo tiempo se implementa un algoritmo para indexar las muestras y calcular las coordenadas *X,Y* del centro de la cuadrícula, como se ve en la Figura 4.

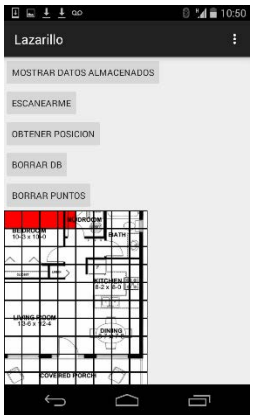


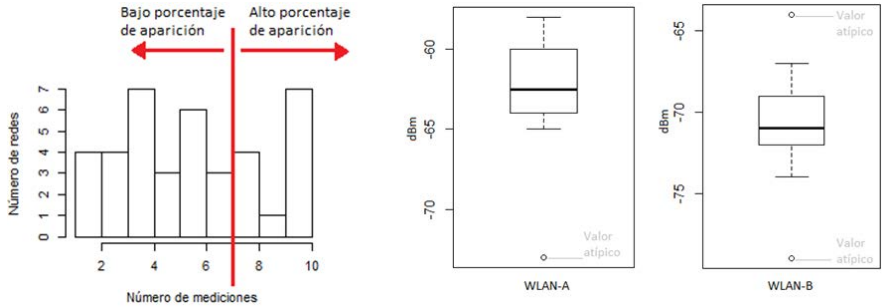
Figura 3: Aplicación de grilla al plano arquitectónico

Figura 4: Interfaz del aplicativo móvil en la etapa de calibración de *fingerprinting*

En la Figura 3 también se observa que los cuadrantes calibrados son deshabilitados y marcados con color rojo, lo cual evita la duplicidad errónea de datos. El aplicativo Lazarillo realiza escaneo de las redes WiFi durante 20 segundos, en los toma 10 muestras de las redes WiFi, es decir, una muestra cada 2 segundos. Esta información se almacena en una base de datos local. Las mediciones hechas se realizaron colocando el teléfono en posición horizontal, siempre con la misma orientación y sobre una superficie de plástico de 1.2 metros de altura, desplazable por todos los espacios del lugar a calibrar. Luego es posible aplicar rutinas de análisis y un filtro estadístico operado por el propio dispositivo móvil.

La metodología propuesta en [8] se aplicó para construir el mapa RSSI y la localización de muestras de usuario en ubicaciones diferentes dentro del espacio. En la mayoría de los puntos coinciden varias redes que solamente son visibles pocas veces de entre todas las posibles, es decir, tienen un porcentaje de aparición bajo. Estas redes son distantes, por lo que se perciben con baja potencia y comportamiento inestable. Por otro lado, las redes con porcentaje de aparición alto corresponden a redes WiFi relativamente estables, cuya potencia de recepción presentan un escenario confiable de medición. Así se implementó el primer criterio de filtrado que se muestra en la Figura 5a, en la que se omiten todas las redes cuyo porcentaje de aparición es menor o igual al 60%. Con este primer criterio de filtrado se evita el uso en la etapa de calibración y en la etapa de localización de señales WiFi altamente inestables y poco confiables [8].

Luego se realizó un segundo filtro estadístico para detectar las medidas atípicas en las redes que superan el primer filtro. Se realizó análisis de *boxplot* a cada una de alto porcentaje de aparición en todos los puntos de medición de la calibración, lo que muestra que las redes analizadas son susceptibles a tener mediciones atípicas, fruto de variables no controladas. Sin embargo, si se omiten esos valores atípicos se obtienen muestras de amplia información que permiten caracterizar cada nodo de medición. El análisis de *boxplot* también permite entender la simetría de las muestras y la dispersión de ellas [8], situación que se muestra en la Figura 5b.



(a) (b)

Figura 5: (a) Definición de redes de bajo y alto porcentaje de aparición según histograma; (b) Ejemplo de análisis de *boxplot* para una red WiFi

Para construir el vector de mediciones RSSI que contiene el valor característico de cada radiación WiFi en un cuadrante específico se aplica el filtro de Kalman, una herramienta para reducir el ruido en las mediciones y predecir una medición estable. Para un modelo estático, como en el caso de una medición individual de la potencias RSSI de una AP en un cuadrante, el filtro es esencialmente un conjunto de ecuaciones que implementa un estimador del tipo predictor–corrector, que es óptimo para minimizar la covarianza del error [9].

Luego de la calibración se diseña la etapa de localización del dispositivo final y según la metodología *fingerprinting* implementada. En el evento en que un dispositivo se quiera localizar dentro del espacio estudiado, se debe realizar la medición de las potencias RSSI y la captura de las direcciones MAC de las redes presentes en esa ubicación, y luego aplicar los filtros estadísticos para tener un vector de potencias RSSI del punto exacto. Con la información del mapa de potencias y el vector de potencias del equipo a localizar, se calculan las múltiples posibles distancias euclidianas en el dispositivo móvil, buscando las de mayor coincidencia. Se implementó un algoritmo que toma las 3 relaciones más coincidentes y calcula la triangulación de posiciones para hallar las posibles coordinadas *X,Y*, y por lo tanto del cuadrante de m^2 donde es más probable que se encuentre el dispositivo. Para calcular la distancia de error se implementa en el dispositivo un algoritmo que usa el cálculo de la distancia euclidiana entre la posición calculada y la posición real:

$$de(P) = \sqrt{(X_p - X_k)^2 + (Y_p - Y_k)^2} \qquad (1)$$

Donde *Xp* y *Yp* indican las coordenadas calculadas del dispositivo, y *Xk* y *Yk* las coordenadas reales de la ubicación del mismo. La precisión en la medición es la diferencia entre la posición calculada y la real.

$$Precision_x = (X_p - X_k) \qquad (2)$$
$$Precision_y = (Y_p - Y_k)$$

Para presentar el resultado de localización en el dispositivo se implementó un algoritmo que ubica la coordenada *X,Y* en el cuadrante de metro cuadrado correspondiente y le informa al usuario el nombre de la ubicación, como se muestra en la Figura 6.

Figura 6: Presentación de información en el dispositivo móvil.

Tomando como elemento de prueba un teléfono inteligente dotado con sistema operativo Android 4.4 e interfaz WiFi de 2.4 GHz, se realizaron pruebas de calibración y cálculo de localización en el espacio cerrado de la Figura 1. La Tabla 1 presenta el resumen de los resultados.

Tabla 1: Resultados de las pruebas

	Distancia de Error (de) con Boxplot y FiltroK
Desv Est σ	0,48
Max	1,87 m
Min	0,24 m
Promedio	0,91 m

De la Tabla 1 se deduce que la metodología propuesta tiene una exactitud en la localización, representada por la distancia de error, de entre 0.2m y 1.87m, y en promedio este error está por debajo de 1m. Dada la configuración por cuadrantes de 1 m^2 se obtiene una experiencia en la cual se ubica al usuario en el cuadrante exacto o en el del lado.

4 Futuros trabajos

El método propuesto permite la reducción del error de distancia en la ubicación y está diseñado para mediciones estáticas en las que el dispositivo a localizar no está en movimiento. Por lo tanto esta metodología se debe acondicionar para rastrear la localización de un dispositivo de movimiento a través del espacio. También se incluirán mediciones adicionales de variables como la presencia de deformaciones del campo magnético de la Tierra en el interior de edificios.

Se están realizando estudios de las técnicas de desarrollo de aplicaciones para uso por parte de personas con discapacidad visual, donde es necesario entender las lógicas de uso de las pantallas táctiles por parte de estas poblaciones, para adecuar la aplicación a sus necesidades y realizar realimentación al usuario de forma audible.

También se desea realizar una versión con una topología de servicio en la nube, donde los cálculos sean realizados por un servidor externo y se puedan centralizar muchas zonas calibradas.

Se proyectan futuros trabajos de investigación asociados a la predicción de desplazamientos y movimientos, sugerencias de rutas a destinos y atención a otros tipos de discapacidad.

En cuanto a las aplicaciones finales se trabajará en soluciones para la ubicación y rastreo de mineros bajo la tierra, la rápida localización de médicos especialistas en grandes hospitales, muestras didácticas e interactivas que dependan de la ubicación del usuario en museos y centros lúdicos.

5 Conclusiones

Por medio del análisis de las huellas de radiación de los diversos *access points* es posible implementar una solución tecnológica, que aproveche las redes WiFi existentes para inferir datos relativos a la localización de un dispositivo en interiores. Asimismo también se puede realizar una implementación totalmente centrada en el uso de la interfaz WiFi de un dispositivo móvil convencional, y usar las capacidades de procesamiento del mismo para realizar los cálculos pertinentes.

Se tomó la metodología de localización de dispositivos móviles por redes WiFi propuesta en [8] y se consiguió desarrollar una solución software que facilita las labores de localización por medio de la metodología *fingerprinting*. El proceso de filtrado de los datos de potencia de RSSI con bajo porcentaje de aparición, el filtro de muestras atípicas por *boxplot* y la aplicación del filtro de Kalman, permiten la estabilización de las muestras y una mejor caracterización del espacio mapeado.

Se obtiene un error en la distancia de localización de aproximadamente 1 metro, lo cual permite establecer con precisión la localización de un dispositivo dentro de un entorno. Se demuestra que es factible implementar una metodología que proporcione información de localización como base para el desarrollo de aplicaciones sociales para muchas necesidades insatisfechas.

El motor de base de datos que utiliza Android esta optimizado para utilizar pocos recursos, sin embargo, en algunas pruebas de campo se mostraban cuadrantes donde incidían hasta 70 redes WiFi. Si en promedio cada una se escanea 7 veces, se tendrán 490 registros por cada cuadrante del plano. Para estos casos se almacena gran cantidad de datos a procesar y el dispositivo móvil estándar toma mucho más tiempo y puede presentar fallos. Las capacidades de un dispositivo móvil estándar permiten procesar gran cantidad de datos de potencias RSSI con los filtros estadísticos de la etapa de calibración de pequeñas zonas, pero para desarrollar estas labores en grandes zonas se recomienda el uso de un equipo de cómputo externo con mayores capacidades.

Referencias

[1] Nuaimi, K. & Kamel H. (2011). A survey of indoor positioning systems and algorithms. Proceedings International Conference on Innovations in Information Technology, pp. 185-190. Abu Dhabi, Emiratos Árabes Unidos.

[2] Zekavat, R. & Buehrer, M. (2011). Handbook of position location. Hoboken: John Wiley & Sons, Inc.

[3] Atia, M.; Korenberg, M. & Noureldin, A. (2012). Particle-filter-based WiFi-aided reduced inertial sensors navigation system for indoor and GPS-denied environments. International Journal of Navigation and Observation 2012, pp. 1–12.

[4] Ladd, A. et al. (2002). Robotics-based location sensing using wireless Ethernet. Proceedings 8th Annual International Conference on Mobile Computing and Networking, pp. 227-238. Atlanta, USA.

[5] Kaemarungsi, K. & Krishnamurthy, P. (2004). Properties of indoor received signal strength for WLAN location fingerprinting. Proceedings First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, pp 14-23. Boston, USA.

[6] Lee, J. et al. (2013). Analysis of location estimation algorithms for Wifi fingerprint-based indoor localization. International Journal of Distributed Sensor Networks 2015, pp. 1-11.

[7] Bahl, P. & Padmanabhan, V. (2000). RADAR: an in-building RF-based user location and tracking system. Proceedings Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, pp. 775–784. Tel Aviv, Israel.

[8] Gómez C. (2015). Localización en espacios interiores de dispositivos móviles por medio de tecnologías WiFi, base para el desarrollo de aplicaciones sociales. Encuentro internacional de educación en ingeniería ACOFI, pp. 1-9. Cartagena, Colombia.

[9] Welch, G. & Bishop, G. (2001). An introduction to the Kalman filter. Technical Report TR 95-041. University of North Carolina.

Adaptative particle swarm optimization algorithm with non-iterative electrostatic repulsion and social neighborhood

Daniel Soto[1]
Wilson Soto[2]
[1]*Universidad de Antioquia – Colombia*
[2]*Universidad Central – Colombia*

1 Introduction

The Particle Swarm Optimization (PSO) is an evolutionary algorithm with particle population that explores a search space in order to find an optimal solution of a problem. This algorithm was proposed in 1995 [3] and it is inspired in the social behavior of the animals living in swarms such as fishes or birds. The particle swarm is composed by simple agents called particles, each particle located into a search space represents a possible solution the given problem. Each particle has three properties: a coordinate vector, a velocity vector and the memory capacity. The trajectory of the particle is influenced by three components: the inertial component, the cognitive component and the social component. Formally, the particle swarm optimization is expressed through the mathematical model:

$$v_i^{t+1} = wv_i^t + c_1 r_1 (x_i^{lbest} - x_i^t) + c_2 r_2 (x^{gbest} - x_i^t) \quad (1)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (2)$$

Where, the variable w is the inertial component. The inertial component was introduced by [7]. Besides, the c_1 is the cognitive component, c_2 is the social component and the variables r_1 and r_2 are random variables between 0 and 1.

The values $lbest$ and $gbest$ are obtained depending of the optimization objective, if the objective is minimization the equations are:

$$lbest_i(t+1) = \begin{cases} lbest(t) & \text{if } f(x_i(t+1)) \geq lbest_i(t) \\ x_i(t+1) & \text{otherwise} \end{cases} \quad (3)$$

$$gbest(t+1) = argmin_{lbest_i} f(lbest(t+1)) \text{ where } 1 \leq i \leq N \quad (4)$$

W.L.O.G the equations are similar for maximization. The variable $gbest$ can be obtained considering all individuals of the population or only the individuals of a neighborhood predefined for each particle. The applications of the PSO algorithm are several and diverse, [5] provides in-depth a state of art on the applications of the PSO algorithm, such as, communication network, combinatorial optimization, financial, entertainment, prediction and robotics, among others. A considerable amount of literature has been published on improvements of PSO algorithm, for example, Constriction Coefficient Particle Swarm, FIPS algorithm (Fully Informed Particle Swarm) and Tribe-PSO [2]. These studies are oriented to the behavior and location of the particles and the variation of the topology.

There are two kind of recent important works that were evaluated with several benchmark functions for global optimization: non-swarm algorithms and PSO variants. Some non-swarm algorithms: Real Genetic Algorithm (RGA) and Gravitational Search Algorithm (GSA) [6]; Teaching Learning Based Optimization (TLBO) algorithm [8]. Some PSO variants: Multi-Swarm PSO (PSO-2S), Multi-Swarm PSO with Diferential Evolution (DEPSO-2S), Efficient Population Utilization Strategy PSO (EPUS-PSO), Comprehensive Learning PSO (CLPSO), among others [2].

The algorithm proposed in this work (A2PSO) is based in the classical PSO algorithm with two improvements: the location of the particles in the search space through of the concept of electrostatic repulsion and the area of interaction of each particle through of the concept of social neighborhood. Our approach is compared with the algorithms: basic PSO, GSA, RGA, PSO-2S, DEPSO-2S, EPUS-PSO, CLPSO and TLBO. Three frameworks were designed with the previous algorithms and differents parameters in order to assess the results in some benchmarking test functions. The structure of the paper is: the Section 2 explains the proposed algorithm. The experimental results are reported in Section 3 and for last the conclusions in Section 4.

2 The proposed algorithm

The proposed algorithm called A2PSO has two important characteristics. The characteristics attempt to improve the location of the particles in the search space. The first characteristic improves the covering of the swarm in the search space with regard to an initialization random. The second characteristic is the influence vector for each particle based in the concept of social neighborhood.

2.1 Initial location

The initial location is based in the concept the electrostatic repulsion [2]. The electrostatic repulsion consists in assign to each particle a repulsion area inside which cannot locate another particle. The PSO algorithm with electrostatic repulsion places random the particles in the search space and iteratively corrects the positions. The graphical representation of the methods are shown in the Figure 1.

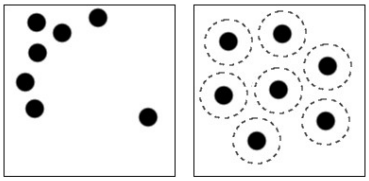


Figure 1: Initialization stage for 7 particles in a space 2D. Initialization random (left) and initialization random with electrostatic repulsion (right)

The A2PSO algorithm uses the electrostatic repulsion eliminating the iterative stage with the calculate of the volume, radio and interaction area of the repulsion of each particle. The volume V of electrostatic repulsion for each particle is calculated by the Equation (5).

$$\mathcal{V} = \frac{L_1 * L_2 * \ldots * L_d}{2 * N} \qquad (5)$$

Where, N is the population, d are the dimensions of the search space and L_i represents the length of the search space on the dimension *i*. With the volume V calculated

we can determine the repulsion radio R which is utilized in the localization of the particles. The Equation (6) calculates the repulsion radio with the volume of a hyper sphere due to be a scheme *n*-dimensional.

$$\mathcal{R} = \left(\frac{\mathcal{V} * \Gamma((d/2)+1)}{(\pi)^{d/2}} \right)^{1/d} \quad (6)$$

For avoid that a particle is in a repulsion area of another particle, the particle must satisfies the Equation (7).

$$\sqrt{(x_1'-x_1'')^2 + (x_2'-x_2'')^2 + \ldots + (x_d'-x_d'')^2} > 2 \cdot \mathcal{R} \quad (7)$$

2.2 Influence vector

An influence vector is a vector based in the concept of the social neighborhood. The social neighborhood is a kind of neighborhood generally without fixed topology, adaptive and static or dynamic behavior. The social neighborhood does not have geographic or distance rules because a particle simply can have neighbors defined in anywhere of the search space [2].

The influence vector in the A2PSO algorithm is built with the Equation (8).

$$z_i^{rand(N)} = 0.2; \mid z_i^n \neq 0 \mid = 5, \forall n \quad (8)$$

The values of the influence vector are iteratively update under the concept of symmetric weakly influences [4]. The updating rule is calculated by the Equation (9) and depends of the nth-neighbor selected. Recent works have determined that inside a birds swarm each individual is in average influenced by 6 or 7 nearest neighbors [1].

$$z_i^n(t+1) = \begin{cases} z_i^n(t) + \dfrac{x_i^{lbest} - x_n^{lbest}}{\sum_{j=0}^{N}(x_i^{lbest} - x_j^{lbest})} & \text{where} \quad 0.2 \leq z_i^n(t) < 1.0 \\ 1 & \text{where} \quad z_i^n(t) \geq 1.0 \\ 0 & \text{where} \quad z_i^n(t) < 0.2 \end{cases} \quad (9)$$

The Equation (10) verifies the previous condition, if the condition is false then is selected a random position with value equal to zero for changing it in the vector. The process is iterative until obtains the predefined number of individuals for creating the social neighborhood.

$$z_i^{rand(N)}(t+1) = 0.2, \text{ if} z_i^{rand(N)}(t) = 0; \mid z_i^n(t+1) \neq 0 \mid = 5, \forall n \quad (10)$$

After of update the values of the influence vector, for each particle i are calculated the location and velocity with the Equations (11) and (12).

$$v_i^{t+1} = wv_i^t + r_1c_1(x_i^{lbest} - x_i^t) + r_2c_2\left(\sum_{n=0}^{N} z_i^n(t)(x_n^{lbest} - x_i^t)\right) \quad (11)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (12)$$

The full A2PSO algorithm is shown in Algorithm 1.

Algorithm 1 A2PSO Algorithm

Require: N ⊳ Number of individuals or particles
Ensure: *gbest* ⊳ Global optimum
1: Locate N particles ⊳ Eqs. 6, 7, 8
2: Evaluate the fitness of N particles
3: Determine the local optimum of N particles (lbest) ⊳ Eq. 3
4: Update the *lbest* value ⊳ Eq. 3
5: Calculate the neighborhood values ⊳ Eqs. 9, 10
6: **while** number of iterations **do**
7: Move the particles ⊳ Eqs. 11, 12
8: Evaluate the fitness N particles
9: Update the *lbest* value ⊳ Eq. 3
10: Calculate the neighborhood values ⊳ Eqs. 9, 10
11: Update *gbest* value ⊳ Eq. 4
12: **end while**
13: **return** *gbest*

3 Results and discussion

The proposed algorithm (A2PSO algorithm) was implemented in language C++ using the MPFR library for the arithmetic compute with decimal multiple precision. Commonly, a set of benchmark test functions is used for evaluating the PSO algorithms, such as the functions in Table 1. The functions are clustered as: unimodal and multimodal.

The results of A2PSO algorithm was compared with three frameworks. Each one contains different classes of algorithms. The first framework (Framework A) includes the basic PSO algorithm (original parameters setting: $\omega \in [0.2, 0.9]$, $c_1 = 2$ and $c_2 = 2$), RGA algorithm and GSA algorithm. The dimension size is 30 for the functions except for the function F7 that is 2. The parameters for the algorithms were: 1000 iterations and 50 individuals. Table 2 shows the results obtained from the Framework A.

Table 1: Test functions. (•) Unimodal function. (∘) Multimodal function

Name	Function	Range	
F1	$f(x) = \sum_{i=1}^{d} x_i^2$	$[-100,100]^d$	∘
F2	$f(x) = \sum_{i=1}^{d} \left(\sum_{j=1}^{i} x_j \right)^2$	$[-100,100]^d$	•
F3	$f(x) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{d} \lvert x_i \rvert$	$[-10,10]^d$	•
F4	$f(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	$[-30,30]^d$	∘
F5	$f(x) = \sum_{i=1}^{n} (x_i + 0.5)^2$	$[-100,100]^d$	•
F6	$f(x) = \sum_{i=1}^{n} -x_i sin(\sqrt{\lvert x_i \rvert})$	$[-500,500]^d$	∘
F7	$f(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1 x2 - 4x_2^2 + 4x_2^4$	$[-5,5]^d$	∘
F8	$f(x) = \sum_{i=1}^{d} \frac{x_i^2}{4000} - \prod_{i=1}^{d} cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	$[-600,600]^d$	∘
F9	Appendix A: Other Functions $(i.)$	$[-32,32]^d$	∘
F10	$f(x) = \sum_{i=1}^{d} [x_i^2 - 10cos(2\pi x_i) + 10]$	$[-10,10]^d$	∘
F11	Appendix A: Other Functions $(ii.)$	$[-100,100]^d$	∘
F12	$f(x) = \sum_{i=1}^{d} i x_i^4 + rand(0,1)$	$[-1.28,1.28]^d$	•

F13	Appendix A: Other Functions $(iii.)$	$[-50,50]^d$	∘
F14	Appendix A: Other Functions $(iv.)$	$[-50,50]^d$	∘
F15	$f(x)=\left(x_2-\frac{5.1}{4\pi^2}x_1^2+\frac{5}{\pi}x_1-6\right)^2+10\left(1-\frac{1}{8\pi}\right)cosx_1+10$	$[-15,15]^d$	∘
F16	Appendix A: Other Functions $(v.)$	$[-2,2]^d$	∘
F17	$f(x)=\sum_{i=1}^{d}\left(\sum_{j=1}^{i}x_j^2\right)^2$	$[-100,100]^d$	•

Table 2: The results Framework A. *b* means the best results and *w* means the worst results

Name		RGA	PSO	GSA	A2PSO
F1	*b*	21.87	1.2×10^{-3}	7.1×10^{-11}	**6.40×10^{-13}**
	w	23.45	5.0×10^{-2}	2.1×10^{-10}	**6.08×10^{-12}**
F2	*b*	5.6×10^{3}	2.2×10^{3}	0.15×10^{3}	**8.81**
	w	5.6×10^{3}	4.1×10^{3}	0.16×10^{3}	**36.95**
F3	*b*	1.07	1.9×10^{-3}	4.03×10^{-5}	**4.82×10^{-7}**
	w	1.13	2.0	**4.07×10^{-5}**	4.93×10^{-4}
F4	*b*	1.0×10^{3}	1.7×10^{3}	25.16	**21.61**
	w	1.1×10^{3}	3.6×10^{4}	25.18	**23.49**
F5	*b*	24.01	1.0×10^{-3}	7.7×10^{-11}	**3.89×10^{-13}**
	w	24.55	0.02	2.6×10^{-10}	**9.84×10^{-13}**
F6	*b*	**-1.2×10^{4}**	-9.8×10^{3}	-2.8×10^{3}	-4.98×10^{3}
	w	**-1.2×10^{4}**	-9.8×10^{3}	-1.1×10^{3}	-4.39×10^{3}
F7	*b*	**-1.0313**	**-1.0316**	**-1.0316**	-1.0316
	w	**-1.0316**	**-1.0316**	**-1.0316**	-1.0316
F8	*b*	1.14	0.0081	0.04	**0**
	w	1.16	0.005	0.29	**4.17×10^{-2}**

The second framework (Framework B) includes the algorithms with better results in [2]: DEPSO-2S, PSO-2S, EPUS-PSO and CLPSO. This framework has two tests. In the test 1 the dimension size is 10 for the functions except for the function F11 that is 2. The parameters for the algorithms were: 40000 iterations and 320 individuals. We decrease the parameters for the A2PSO algorithm: 5000 iterations and 50 individuals. In the test 2 the functions have the dimension size 30. The parameters for the algorithms were: 150000 iterations and 400 individuals. We decrease the parameters for the A2PSO algorithm: 50000 iterations and 150 individuals. Table 3 shows the results obtained from the Framework B test 1 and Table 4 the results from Framework B test 2. The original parameters setting of the algorithms of the framework were: $\omega\approx0.72$, $c_1\approx1.19$ and $c_2\approx1.19$.

Table 3: The results Framework B test 1. *b* means the best results and *w* means the worst results. The character "-" means not evaluated

Name		DEPSO-2S	PSO-2S	EPUS-PSO	CLPSO	A2PSO
F1	*b*	1.3×10^{24}	1.05×10^{-86}	5.55×10^{-153}	4.02×10^{-21}	**7.20×10^{-247}**
	w	3.33×10^{-2}	3.10×10^{-86}	2.44×10^{-134}	6.49×10^{-21}	**8.40×10^{-243}**
F2	*b*	6.18×10^{-28}	6.18×10^{-28}	9.32×10^{-12}	5.04×10^{2}	**1.23×10^{-36}**
	w	**2.13×10^{-27}**	**2.13×10^{-27}**	4.16×10^{-11}	1.99×10^{2}	4.62×10^{-27}
F8	*b*	-	5.3×10^{-5}	-	-	**0**
	w	-	**2.93×10^{-3}**	-	-	1.47×10^{-2}
F9	*b*	4.79×10^{-13}	4.12×10^{-15}	2.72×10^{-15}	1.78×10^{-11}	**-1.06×10^{-14}**
	w	2.53×10^{-2}	**1.21×10^{-15}**	1.5×10^{-15}	1.63×10^{-11}	1.77×10^{-14}
F10	*b*	**0**	**0**	**0**	2.35×10^{-9}	**0**
	w	**0**	**0**	**0**	2.72×10^{-9}	8.9×10^{-1}
F11	*b*	-	1.79×10^{-5}	-	-	**1.20×10^{-49}**
	w	-	**2.45×10^{-3}**	-	-	2.13×10^{-1}

Table 4: The results framework B test 2. *b* means the best results and *w* means the worst results

Name		DEPSO-2S	PSO-2S	EPUS-PSO	CLPSO	A2PSO
F1	*b*	1.25×10^{-25}	3.02×10^{-119}	8.50×10^{-263}	1.34×10^{-25}	$\mathbf{5.05\times10^{-379}}$
	w	3.24×10^{-24}	1.56×10^{-118}	1.02×10^{-262}	1.71×10^{-25}	$\mathbf{5.18\times10^{-371}}$
F2	*b*	3.88×10^{-12}	6.18×10^{-11}	197	2.41×10^{4}	$\mathbf{1.27\times10^{-13}}$
	w	2.50×10^{-11}	2.63×10^{-11}	4.69	6.78×10^{3}	$\mathbf{5.29\times10^{-13}}$
F9	*b*	6.48×10^{-14}	1.63×10^{-1}	$\mathbf{3.91\times10^{-15}}$	9.90×10^{-14}	4.61×10^{-14}
	w	3.33×10^{-14}	3.85×10^{-1}	$\mathbf{1.07\times10^{-15}}$	3.80×10^{-14}	2.16×10^{-13}
F10	*b*	$\mathbf{3.98\times10^{-14}}$	1	0	1.87×10^{-9}	10.9
	w	$\mathbf{2.33\times10^{-14}}$	11.8	0	5.34×10^{-9}	14.9

The third framework (Framework C) has the TLBO algorithm. The dimension size is 30 for the functions. The parameters for the algorithms were in (100, 666, 2000) iterations and (25, 50, 75) individuals depending of the function. Table 5 shows the specific parameters for each function and the results obtained from the Framework C. In the three frameworks, the parameters setting of the A2PSO algorithm were random values in a range: $\omega \in [0.5, 0.8]$, $c_1 \in [0.2, 1.6]$ and $c_2 \in [0.5, 2.1]$.

Table 5: The results framework C. *b* means the best results and *w* means the worst results

Name	Iterations	Individuals		TLBO	A2PSO
F1	2000	25	*b*	2.09×10^{-5}	$\mathbf{1.61\times10^{-37}}$
			w	7.27×10^{-4}	$\mathbf{8.16\times10^{-37}}$
F3	2000	25	*b*	3.53×10^{-5}	$\mathbf{1.12\times10^{-5}}$
			w	9.58×10^{-4}	$\mathbf{4.44\times10^{-4}}$
F4	666	75	*b*	**13.88407**	21.42206
			w	**19.08793**	27.77500
F5	2000	25	*b*	1.81×10^{-5}	$\mathbf{4.04\times10^{-6}}$
			w	8.06×10^{-4}	$\mathbf{7.85\times10^{-4}}$
F6	666	75	*b*	**-12569.49**	-4927.20
			w	**-12158.04**	-3868.12
F7	2000	25	*b*	**-1.031628**	**-1.031628**
			w	**-1.031628**	**-1.031628**
F8	2000	25	*b*	8.93×10^{-5}	$\mathbf{3.55\times10^{-15}}$
			w	$\mathbf{5.27\times10^{-4}}$	8.42×10^{-4}
F9	2000	25	*b*	1.89×10^{-4}	$\mathbf{1.99\times10^{-8}}$
			w	6.21×10^{-4}	$\mathbf{6.22\times10^{-6}}$
F10	2000	25	*b*	$\mathbf{1.78\times10^{-4}}$	33.21
			w	$\mathbf{8.67\times10^{-4}}$	42.78
F12	1000	50	*b*	0.001253	$\mathbf{2.20\times10^{-7}}$
			w	0.014182	$\mathbf{8.93\times10^{-7}}$
F13	666	75	*b*	2.67×10^{-4}	$\mathbf{1.96\times10^{-8}}$
			w	8.27×10^{-4}	$\mathbf{2.00\times10^{-6}}$
F14	666	75	*b*	$\mathbf{2.37\times10^{-8}}$	4.13×10^{-1}
			w	$\mathbf{6.77\times10^{-4}}$	5.08×10^{-1}
F15	2000	25	*b*	0.3978	**0.3978**
			w	**0.3984**	0.4023
F16	2000	25	*b*	3	**2.9999**
			w	3	**2.9999**
F17	2000	25	*b*	3.87×10^{-5}	$\mathbf{6.51\times10^{-6}}$
			w	$\mathbf{8.17\times10^{-4}}$	8.48×10^{-4}

The positive results are due to the two characteristics of our approach (A2PSO algorithm), the initial location and influence vector. These characteristics are meaningful in the distribution of the particles in the search space for each iteration. Another possible explanation for the good results obtained by our approach might be that it has a fast

convergence to solutions and the particles occupy a greater surface in the search space, because of less iterations and less individuals.

4 Conclusions

This work proposed a PSO algorithm. The proposed approach includes two improvements, non-iterative electrostatic repulsion and social neighborhood. The experimental results were promising, these shown that our approach gets better accuracy of some solutions in an extensive benchmarking of unimodal and multimodal functions. Even with regard to the algorithms used in the experimental frameworks our approach achieves positive results with less iterations and less individuals, although the current study has only examined a set of functions and specific values for the parameters.

The A2PSO algorithm does not aspire to be the best, by contrast we want that it will be part of the state of art of the PSO algorithms.

Further research might explore a learning stage to our approach, besides, it would be interesting to assess the effects of a new implementation considered the parallel and distributed paradigms.

References

[1] Ballerini, M. et al. (2008). Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. Proceedings National Academy of Sciences of the United States of America 105(4), pp. 1232–1237.

[2] El Dor, A. (2012). Perfectionnement des algorithmes d'optimisation pas essaim particulaire. Applications en segmentation d'images et en électronique. PhD dissertation. Université Paris.

[3] Kennedy, J. & Eberhart, R. (1995). Particle swarm optimization. Proceedings International Conference on Neural Networks (pp. 1942-1948). Nov. 27 – Dec. 1, Perth, Australia.

[4] Martin, S. (2013). Coordination et robustesse des systèmes dynamiques multi-agents. PhD dissertation. Université de Grenoble.

[5] Poli, R. (2008). Analysis of the publications on the applications of particle swarm optimisation. Journal of Artificial Evolution and Applications (2008), pp. 1-10 .

[6] Rashedi, E., Nezamabadi, H. & Saryazdi, S. (2009). GSA: A Gravitational Search Algorithm. Information Sciences 179(2009), pp. 2232–2248.

[7] Shi, Y. & Eberhart, R. C. (1999). Empirical study of particle swarm optimization. Proceedings International Conference on Evolutionary Computation (pp. 1945-1950). July 6–9, Washington, USA.

[8] Venkata, R. & Patel, V. (2013). Comparative performance of an elitist teaching-learning-based optimization algorithm for solving unconstrained optimization problems. International Journal of Industrial Engineering Computations 4(1), pp. 29–50.

Appendix A: Other functions

$(i.)$	$f(x) = -20exp\left(-0.2\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - exp\left(\frac{1}{d}\sum_{i=1}^{d} cos(2\pi x_i)\right) + 20 + e$
$(ii.)$	$f(x) = \begin{cases} \frac{1-sign(x_2)}{2}(\lvert x_1 \rvert + \lvert x_2 + 50 \rvert) + \frac{1+sign(x_2)}{2}\frac{1-sign(x_1)}{2} \times \\ (1 + \lvert x_1 + 50 \rvert + \lvert x_2 - 50 \rvert) + \frac{1+sign(x_1)}{2}(2 + \lvert x_1 - 50 \rvert + \lvert x_2 - 50 \rvert) \\ \text{where } sign(x) = \begin{cases} -1 \text{ if } x \le 0 \\ 1 \text{ if } x > 0 \end{cases} \end{cases}$
$(iii.)$	$f(x) = \begin{cases} \frac{\pi}{d}\left[10 sin^2(\pi y_1) + \sum_{i=1}^{d-1}(y_i - 1)^2\{1 + 10 sin^2(\pi y_{i+1})\} + (y_p - 1)^2\right] + \sum_{i=1}^{d} u(x_i, 10, 100, 4) \\ \text{where } u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 & -a < x_i < a \\ k(-x_i - a)^m & x_i < -a \end{cases} \\ \text{and } y_i = 1 + ((x_i + 1)/4) \end{cases}$
$(iv.)$	$f(x) = \begin{cases} 0.2\left[sin^2(\pi x_1) + \sum_{i=1}^{d-1}(x_i - 1)^2(1 + sin^2(3\pi x_{i+1})) + (x_p - 1)^2 + (1 + sin^2(2\pi x_d))\right] + \sum_{i=1}^{d} u(x_i, 5, 100, 4) \\ \text{where } u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 & -a < x_i < a \\ k(-x_i - a)^m & x_i < -a \end{cases} \end{cases}$
$(v.)$	$f(x) = \left[1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)\right] \times \left[30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)\right]$

Estudio y análisis de la regulación y normativa en el servicio de datos móviles para tecnologías 4G en Colombia

Jhon Gutiérrez L.
Germán Góez S.
Sergio Serna G.
Instituto Tecnológico Metropolitano – Colombia

1 Introducción

El mercado de banda ancha móvil ha experimentado un vertiginoso crecimiento gracias al desarrollo de nuevas tecnologías implementadas en todo el mundo, nuevas redes desplegadas, nuevos y llamativos dispositivos móviles y más de medio millón de aplicaciones móviles. Por todas estas razones la banda ancha móvil representa en la actualidad la vanguardia en innovación y desarrollo para Computación, Redes, Internet y Software. Según 4G Américas, el crecimiento de conexiones 4G a nivel mundial se ha incrementado de manera exponencial hasta Junio de 2014; resaltando la creciente necesidad de los usuarios por tener conexiones de alta velocidad y servicios de aplicaciones rápidas y sin congestión [1].

A pesar de que en el país existe regulación para garantizar la calidad del servicio en las redes móviles y en ella se habla del monitoreo de algunos de sus servicios dentro del plan Vive Digital, en realidad se observa que en su mayoría se enfocan en las tecnologías 2G y 3G, y la calidad de los servicios de voz; por lo que a partir del crecimiento en las redes 4G en Colombia se hace necesario realizar un estudio desde el punto de vista regulatorio, debido a que se desconoce en detalle la normatividad sobre el monitoreo del desempeño que ofrece esta tecnología.

El desconocimiento de la gestión realizada por los dispositivos de red genera problemas, tales como la imposibilidad de detectar de manera temprana fallas en el sistema, o no conocer el punto exacto donde se presenta una falla. También se tiene la incapacidad de monitorear en tiempo real el comportamiento y el desempeño de la red en parámetros como SNR, potencia, temperatura de los equipos,… por lo cual no se puede garantizar la calidad del servicio a los usuarios del sistema, ni dar una atención oportuna ante quejas originadas en las mismas falencias. Con este trabajo se pretende contribuir al mejoramiento de la calidad del servicio y a su prestación, proponiendo un conjunto de recomendaciones regulatorias que permitan conseguir un eficiente monitoreo y desempeño en las redes 4G.

Este aporte se estructura de la siguiente manera: en la sección 2, el marco regulatorio para comunicaciones móviles en Colombia, Argentina, México y EE.UU., una descripción de las regulaciones actualmente disponibles en estos países. En la sección 3, Organismos internacionales, las disposiciones en 4G por parte de la ITU, 3GPP e IEEE. Para finalizar, en la sección 4, Análisis y resultados, se incluye la actualidad de las regulaciones colombianas y los operadores locales; la síntesis del problema y recomendaciones que servirán como base para un futuro marco regulatorio para el servicio de 4G en el país.

2 Marco regulatorio para comunicaciones móviles en Colombia

En Colombia, el Decreto 1130 de 1999, en su artículo 37 numerales 3 y 4, ordena a la CRT expedir toda regulación general o particular que establezca parámetros de calidad de los servicios e igualmente fijar indicadores y metas de calidad y eficiencia de los servicios,

e imponer índices de calidad, cobertura y eficiencia a uno o varios operadores [2]. También se puede encontrar que los operadores están obligados a reportar trimestralmente los indicadores que presentarán durante el transcurso del siguiente mes calendario a su finalización, y según el servicio que se reporte se definen los indicadores que se deben medir [3].

En 2011, cuando se comenzaron los estudios para la subasta de 4G en el país, se establecieron condiciones con las que se busca cumplir objetivos tales como promover la competencia, favorecer al usuario con mayor cobertura, mejorar la calidad de servicio, promover el desarrollo de Internet y realizar un proceso transparente y participativo. Además de lo anterior, también se habla sobre la asignación del espectro para 4G para los operadores [4]. Dentro de las condiciones se ofrece alternativas para la asignación de espectro con el fin de mejorar la calidad del servicio, incluso se plantea que a mayor espectro mayor calidad [5].

Para diciembre de 2011, la CRC expide la Resolución 3503 donde se modifican las resoluciones CRC 3067 y 3496, definiendo nuevas disposiciones de calidad para el servicio de acceso a Internet móvil, entre ellas que se encuentran:

- Definición de una metodología de medición y reportes, así como fijación de metas de calidad para Internet móvil
- Inclusión de nuevos parámetros para la verificación de la calidad en los servicios de datos a través de redes móviles

También se hace referencia a que el ministerio de las TIC posee la facultad de realizar actividades de monitoreo permanente a las redes, como por ejemplo, a través de verificaciones de campo [6]. En la Figura 1 se observan los indicadores de calidad en 3G.

Clase	Velocidad	Error	Retardo	Variación en el Retardo
Conversacional	La necesaria según aplicación (VoIP, video conferencia, etc...)	Secundario	Parámetro crítico	Parámetro crítico
Streaming	La necesaria según aplicación (video streaming en calidad STD, HD, etc...)	Secundario	Secundario	Parámetro crítico
Interactiva	Parámetro crítico en función del uso recurrente de carga y descarga de archivos	Parámetro crítico, en especial para asegurar la transferencia de archivos	Secundario	Secundario dentro de cierto margen
Background	Secundario	Parámetro crítico, en especial para transferencia de archivos	Secundario	Secundario

Figura 1: Indicadores de calidad para 3G [6]

2.1 De la regulación para 4G

En cuanto a la regulación para 4G en Colombia se ha podido encontrar que para su asignación de bandas IMT a los operadores móviles y su respectiva subasta, fueron las resoluciones 4112 y 449 de 2013 que fijaronn las disposiciones finales para subasta y entrada en funcionamiento.

La resolución 4112 establece las condiciones generales para la provisión de la instalación de Roaming Automático Nacional[18], donde se mencionan elementos tales como la obligación de proveer Roaming Automático Nacional, obligaciones del proveedor, remuneración y reportes periódicos según las disposiciones del Ministerio de las TIC y recomendaciones de la ITU. La resolución 449 de 2013 define la asignación de espectro

[18] Instalación esencial asociada a las redes de Telecomunicaciones con acceso móvil que permite, sin intervención directa de los usuarios, proveer servicios a éstos, cuando se encuentran fuera de la cobertura de uno o más servicios de su red de origen [54].

radioeléctrico para las bandas a utilizar con la tecnología 4G. Se establecen las condiciones de acuerdo con la normatividad del Ministerio de las TIC y algunas recomendaciones de la ITU para la subasta con los operadores, y la regulación para las Fuerzas Armadas de Colombia en cuanto a la utilización de su espectro asignado y las disposiciones sobre los equipos que usarán y los indicadores de parámetros que serán monitoreados [7].

En el sitio web de la CRC se pueden encontrar los informes de calidad del servicio sobre la telefonía móvil, entre los cuales se encuentran parámetros tales como porcentaje de llamadas caídas, porcentaje de llamadas caídas desagregado, porcentaje de intentos de llamada no-exitosas 2G, porcentaje de intentos de llamada no-exitosas 3G, tasa de llamadas nacional e internacional y tiempo medio en reparación de daños, pero no se especifica nada para 4G [8].

A pesar de existir regulación para garantizar la calidad del servicio en las redes móviles, y a que en ella se habla del monitoreo de algunos de sus servicios dentro del plan Vive Digital, en realidad se observa que en su mayoría las regulaciones se enfocan en las tecnologías 2G y 3G y la calidad de servicio de los servicios de voz; por lo que a partir del crecimiento en las redes 4G se hace urgente establecer nuevas normas para esta tecnología, debido a que se desconoce en detalle la normatividad sobre el monitoreo del desempeño que ofrece.

2.2 Disposiciones del Ministerio, la ANE y los operadores que prestan servicio 4G

El Ministerio de las TIC proporciona el listado de las leyes, decretos, resoluciones y artículos constitucionales que se han tenido en cuenta para que ayuden a tener prestación del servicio 4G en Colombia. Entre los contenidos de estas normas se pueden encontrar principalmente definiciones sobre el espectro radioeléctrico, protección al usuario, planes tarifarios y la creación de la ANE, pero no proporcionan muchos detalles técnicos.

La resolución número 3101 de 2011 [9] es la que más resalta los detalles técnicos y menciona recomendaciones internacionales en sus Artículos 19, 20, 22 y 23 determina estas recomendaciones:

- Enrutamiento alternativo y desborde: ITU-T E.170, ITU-T E.521 e ITU-T E.522. Para redes de conmutación de paquetes aplica ITU-T Y.2201 [10-13].
- Protocolos de señalización: protocolo SS7, teniendo en cuenta la recomendación ITU-T Q.704 [14].
- Transmisión: Calidad de servicio, recomendaciones ITU-T G.1000 y ITU-T G.1010 [15, 16]; parámetros de transmisión, recomendaciones ITU-T Q.551 a Q.552, ITU-T G.712, ITU-T G.168, ITU-T G.113, ITU-T G.820, ITU-T P.310, ITU-T P.311 e ITU-T P.341 [17, 24]; transmisión en redes IP, recomendaciones ITU-T Y.1540 y ITU-T Y.1541 [25, 26]; si las redes a interconectar requieren sincronización se permite utilizar ITU-T G.822 con soporte de ITU-T G.811 [27, 28].
- Códecs: recomendaciones ITU-T G.711, ITU-T Q.3401 e ITU-T G.101 [29-31].

Hay que mencionar que las resoluciones listadas para cada asignatario de espectro 4G: Mintic 1177 de 2010, 2623 de 2013, 2624 de 2013, 2625 de 2013, 2626 de 2013, 2627 de 2013 y modificatorias, no se encuentran disponibles para su revisión en el portal de consulta del ministerio de las TIC y tampoco fue posible hallarlas a través de los motores de búsqueda conocidos [32].

Por su parte, la ANE dentro de la agenda para 2015 incluye el Plan de Acción [33] con el objetivo de *"Promover la disponibilidad del espectro necesario para la masificación de*

banda ancha inalámbrica, buscando la ampliación de la cobertura y el mejoramiento de los servicios por parte de los proveedores de redes y servicios". Esto se planea alcanzar por medio del diseño y propuesta de políticas, lineamientos y estrategias, logrando un uso eficiente del espectro en el mediano y largo plazo en pro del interés general. Todo esto teniendo en cuenta las recomendaciones de entidades internacionales como la ITU y participando en la Conferencia Mundial de Radiocomunicaciones de 2015. También aplica el documento Dimensión Estratégica I [34], que en su primera parte define las actividades a realizar para el análisis técnico y económico del uso de las bandas propuestas e identificadas para el uso de las tecnologías IMT en Colombia. Dichas actividades son:

- Estudio técnico y económico de las bandas candidatas a ser identificadas por la Unión Internacional de Telecomunicaciones (UIT) para las IMT en la Conferencia Mundial de Radiocomunicaciones del 2015 (CMR-15) y que han sido seleccionadas.
- Análisis de escenarios de migración y sugerencias de reserva (Primer trimestre).
- Propuesta de una posición del país en las discusiones para la CMR-15 respecto a bandas deseables, prioritarias y convenientes a ser identificadas como IMT (Segundo trimestre).
- Socialización de los resultados del estudio (Tercer trimestre).
- Expedición de actos administrativos para la reserva y cambio de atribución de las bandas que sean definidas y lo necesiten (Cuarto trimestre).

2.3 Operadores de 4G en Colombia

En el estudio de las regulaciones y normativas disponibles en los sitios web de los operadores móviles y las señaladas en las respuestas a las PQR enviadas a éstos, se logra evidenciar que aunque cada resolución aparece identificada con una numeración diferente para identificar a un operador en particular, sus contenidos son similares en cuanto a sus disposiciones para algunas especificaciones técnicas en calidad del servicio, seguridad y a la protección del usuario. Por lo cual es posible determinar que técnicamente Claro, Tigo y Movistar, y por disposición de la CRC, deben cumplir los mismos protocolos, recomendaciones y normas internacionales para operar en el país.

Movistar es el proveedor de servicios 4G que tiene mayor cobertura en el territorio en términos de presencia en los departamentos, aunque no ofrece muchos detalles al respecto; mientras que Claro es el que mayor detalle presenta de las zonas con cobertura sobre las ciudades principales, evidenciando una limitación a las áreas urbanas. Tigo al igual que Movistar no ofrece mayor información respecto a su cobertura, principalmente en los departamentos donde presta los servicios de 4G sin especificar las ciudades con servicio operativo.

3 Regulación para comunicaciones móviles en otros países

3.1 Argentina

Por medio de la Resolución Nº 5 de 2013 [35], la república de Argentina aprueba el Reglamento de Calidad de los Servicios de Telecomunicaciones, y que con base en recomendaciones de calidad de la ITU, tiene como objetivo establecer los requisitos de calidad exigibles para la prestación de los servicios de telecomunicaciones, en el ámbito de las redes de uso público fijas y móviles, y es de cumplimiento obligatorio en todo el país. Esta resolución define elementos más importantes para los servicios de telecomunicaciones en cuanto a la calidad del servicio y la prestación de éste:

- Indicadores de calidad, Artículo 4°
- Sistema de medición, Artículo 5°
- Publicación indicadores de calidad, Artículo 10°

En la Resolución Nº 17 de 2014 [36], y el Decreto 671 de 2014 [37] se define la atribución de las bandas 1710-1770 MHz y 2110-2170 MHz al Servicio Móvil Terrestre con categoría primaria, las cuales corresponden a las bandas IMT designadas por la ITU para la operación de 4G.

Por medio de la Resolución Nº 37 de 2014 [38] se resuelve la aprobación del Reglamento General del Servicio de Comunicaciones Móviles Avanzadas (SCMA), que se define como servicio inalámbrico de telecomunicaciones, y mediante el empleo de tecnología de acceso digital soporta baja y alta movilidad del usuario, altas tasas de transferencia de datos, interoperabilidad con otras redes fijas y móviles, con capacidad para itinerancia mundial y orientadas a la conmutación de paquetes que permiten la utilización de una amplia gama de aplicaciones, incluyendo las basadas en contenido multimedia.

3.2 México

Se destacan documentos como el Espectro Radioeléctrico en México [39], cuyo objetivo es establecer un marco de referencia para la planeación estratégica del espectro, ofreciendo inicialmente un diagnóstico general que abarca las bandas de frecuencias más relevantes desde los 30 MHz hasta los 300 GHz, con la finalidad no solamente de cumplir con el mandato legal a la Comisión, sino como un primer insumo para el cumplimiento de los compromisos 41 y 44 contenidas en el pacto por México. Con base en recomendaciones de la ITU y desde el punto de vista jurídico y técnico, el proyecto busca por medio de varias propuestas que se administre de forma eficiente el espectro radioeléctrico y que se realice una asignación de frecuencias que cumpla con al menos los elementos básicos que aseguren una mejor explotación de este recurso.

Dentro de su marco regulatorio también se encuentran reformas recientes donde se actualizan y adicionan diversas disposiciones de artículos en la constitución política de los Estados Unidos Mexicanos en materia de telecomunicaciones [40]. La Ley Federal de Telecomunicaciones y Radiodifusión [41] presenta su actualización más reciente en julio de 2014, y su objetivo es regular el uso, aprovechamiento y explotación del espectro radioeléctrico, las redes públicas de telecomunicaciones, el acceso a la infraestructura activa y pasiva, los recursos orbitales, la comunicación vía satélite, la prestación de los servicios públicos de interés general de telecomunicaciones y radiodifusión y la convergencia entre éstos, los derechos de los usuarios y las audiencias y el proceso de competencia y libre concurrencia en estos sectores, para que contribuyan a los fines y al ejercicio de los derechos establecidos en los artículos 6, 7, 27 y 28 de la Constitución Política de los Estados Unidos Mexicanos.

3.3 Estados Unidos

En 2009 y por solicitud del congreso, se le solicitó a la FCC desarrollar un Plan Nacional de Banda Ancha con el fin de garantizar el acceso a todos los ciudadanos [42]. Su misión como plan es desarrollar un Estados Unidos de rendimiento más eficiente, creativo y productivo, en el que se pueda acceder a una banda ancha económica desde cualquier lugar y en el que todos posean los medios y la capacidad para utilizar las valiosas

aplicaciones de ella. Según el plan sus objetivos deben fijarse para 2020, de tal manera que sirvan de referencia para la próxima década.

El Plan Nacional de Banda Ancha realiza un estudio al estado actual de esta banda en el país, pasando por el análisis de innovación, competencia de la banda ancha entre las diferentes tecnologías, espectro radioeléctrico, infraestructura, entre otros, y realizando recomendaciones al final de cada tema analizado con el propósito de establecer procesos y aplicaciones eficientes de las tecnologías y espectro.

La regulación establecida por la FCC se encuentra disponible para consulta en el sitio web e-CFR, donde está consignada la normatividad concerniente a telecomunicaciones, espectro radioeléctrico y comunicaciones móviles. Estas normas definen todos los aspectos que deben cumplir los operadores en materia de calidad de servicio y operación, sin hacer énfasis sobre la tecnología 4G.

4 Organismos internacionales

4.1 Unión Internacional de Telecomunicaciones ITU

Con el propósito de conseguir eficiencia, gestión y solución de problemas técnicos del espectro, la Unión Internacional de Telecomunicaciones cuenta con la Recomendación ITU-R M.1036-4, que proporciona una guía en la selección de arreglos de frecuencia de transmisión y recepción para el componente terrestre de sistemas IMT, con la intención de asistir las administraciones sobre problemas técnicos asociados al espectro, a la implementación y al uso del componente terrestre IMT en las bandas identificadas en el Reglamento de Radiocomunicaciones. Esta recomendación es complementada con otras ITU-R y reportes IMT, que dan detalles adicionales en el número de aspectos, incluyendo características de emisión no deseadas para las bandas incluidas en este documento y especificaciones para la interfaz radio [43]. La recomendación contiene aspectos aplicables a los diferentes arreglos de frecuencias:

- Implicaciones de asimetría de tráfico
- Segmentación de espectro
- Arreglo Duplex y separación
- Duplexor dual
- Disponibilidad de frecuencia

Todos los aspectos anteriores aplican para las frecuencias IMT establecidas [43]. Para soportar aspectos técnicos propios de LTE se ofrece la Recomendación ITU-R M.2012-1 que identifica las tecnologías de la interface radio terrestre de IMT-Advanced y proporciona las especificaciones detalladas de la interfaz radio. Estas especificaciones detallan las características y parámetros de IMT-Advanced, incluyendo la capacidad de asegurar compatibilidad mundial, *roaming* internacional y acceso a servicios de datos de alta velocidad [44].

4.2 Third Generation Partnership Project 3GPP

Asimismo, 3GPP describe en detalle las características más recientes para LTE en su documento Overview of 3GPP Release 13 [45], que entre otras incluyen:

- Administración de congestión en plano de usuario
- Mejoras en monitoreo
- Servicio de extremo a extremo de telefonía para extensiones IMS
- Características UTRAN

- Características GERAN
- Estudio de Advanced Wireless Services (AWS) – Banda de extensión para LTE

En situaciones de emergencia es importante tener disponible un medio de comunicación que permita coordinar operativos de rescate y transporte de ayudas a zonas afectadas, por lo que es necesario garantizar los canales y ancho de banda para este propósito. Por tanto, 3GPP también recopila los artículos de trabajo para mejorar LTE con el fin de soportar las comunicaciones de Seguridad Pública, de manera que policía, bomberos, paramédicos, ambulancias y otros servicios puedan usarlo con el fin de remplazar las redes de banda angosta dedicadas a Seguridad Pública [46].

Para ayudar a tener claros los aspectos técnicos aplicables a 4G y que deben ser incluidos en la nueva regulación, 3GPP ofrece varias series de especificaciones determinando parámetros tales como Aspectos de servicio, requerimientos, seguridad y señalización [47]. Es importante mencionar que dentro de las series descritas también se encuentran especificaciones aplicables a la tecnología 3G, las cuales podrían ser tenidas en cuenta para revisar y actualizar aspectos regulatorios sobre la tecnología.

4.3 Institute of Electrical and Electronics Engineers IEEE

Por su parte, con el fin de proporcionar compatibilidad entre tecnologías y buscando la mayor eficiencia posible en sus interfaces, IEEE ofrece algunos estándares dedicados a la banda ancha móvil inalámbrica, que de acuerdo con las necesidades regulatorias en Colombia serán de utilidad en el momento de establecer un nuevo conjunto de normas dedicado a 4G. El estándar IEEE 802.16 especifica la interfaz de aire, incluyendo el medio de la capa de control de acceso y la capa física, de sistemas combinados de banda ancha fija e inalámbrica móvil punto-multipunto proporcionando múltiples servicios. Posibilita un desarrollo rápido de productos de acceso de banda ancha inalámbrica innovadores y de costos accesibles, facilita la competencia proporcionando alternativas telefonía de acceso por banda ancha, alienta la asignación consistente de espectro mundial y acelera la comercialización de sistemas de acceso de banda ancha inalámbrica [48]. El estándar IEEE 802.16.1 y su enmienda IEEE 802.16.1.a, además de las características mencionadas, permite el soporte de los requerimientos para LTE-Advanced de la ITU [49, 50].

5 Análisis y resultados

Para llegar al análisis y síntesis de la problemática y las recomendaciones finales, se estudiaron aspectos importantes como la situación actual de las regulaciones colombianas respecto a telefonía móvil, específicamente para 4G. También la situación de la regulación de otros países para conocer qué tan actualizados se encuentra respecto a la normatividad para ésta tecnología en crecimiento. Por último, se estudiaron las especificaciones, recomendaciones y estándares publicados por entidades internacionales, tales como IEEE, 3GPP e ITU referentes a la tecnología LTE, que puedan ser aplicables a una futura regulación de 4G en el país.

5.1 Actualidad de las regulaciones colombianas y los operadores nacionales

Dado que en el país no se tiene una normatividad dedicada para los servicios de banda ancha inalámbrica móvil a través de LTE, no se puede garantizar en su totalidad la calidad del servicio, además de aspectos técnicos que podrían resultar incompatibles debido a su atraso relativo al avance acelerado en el desarrollo de las nuevas tecnologías, como se puede observar en las caracterizaciones desarrolladas. Sin embargo, con el estudio de las

normas anteriores es posible establecer una base que sirva de modelo para diseñar normativas específicas para la tecnología estudiada.

Algunos estándares y recomendaciones internacionales empleados en las resoluciones actuales que reglamentan las comunicaciones móviles en el país datan de años anteriores a la implementación de la tecnología 4G, lo que podría provocar incompatibilidades con los nuevos sistemas, afectando a la creciente cantidad de usuarios que, según el Boletín Trimestral de las TIC, llega a 466.217 abonados en el cuarto trimestre de 2014 [51]. Las Figuras 2 y 3 muestran el crecimiento exponencial de abonados y suscriptores a 4G en 2014.

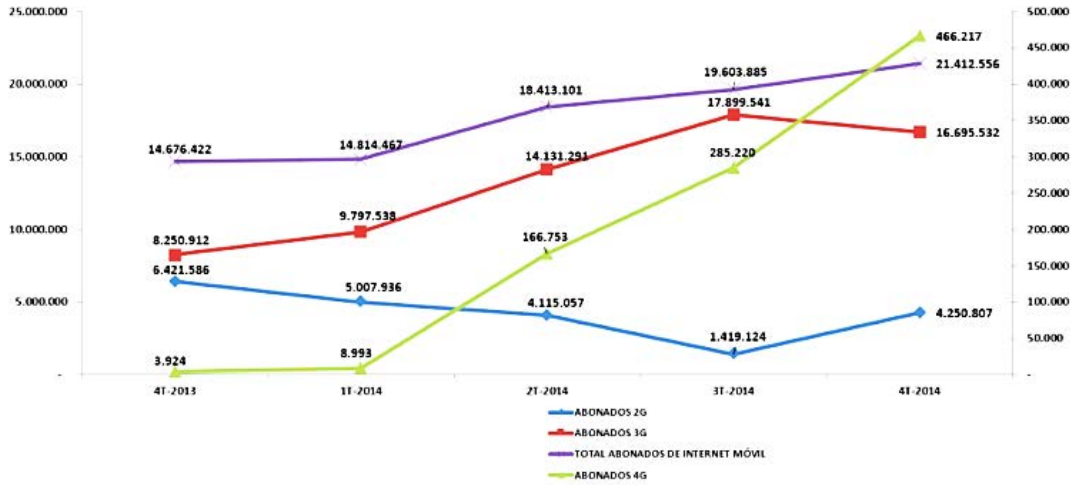


Figura 2: Abonados a Internet móvil por generación en 2014 [51]

Por su parte, el comportamiento de suscriptores[19] llegó a 1.134.987 en el cuarto trimestre de 2014.

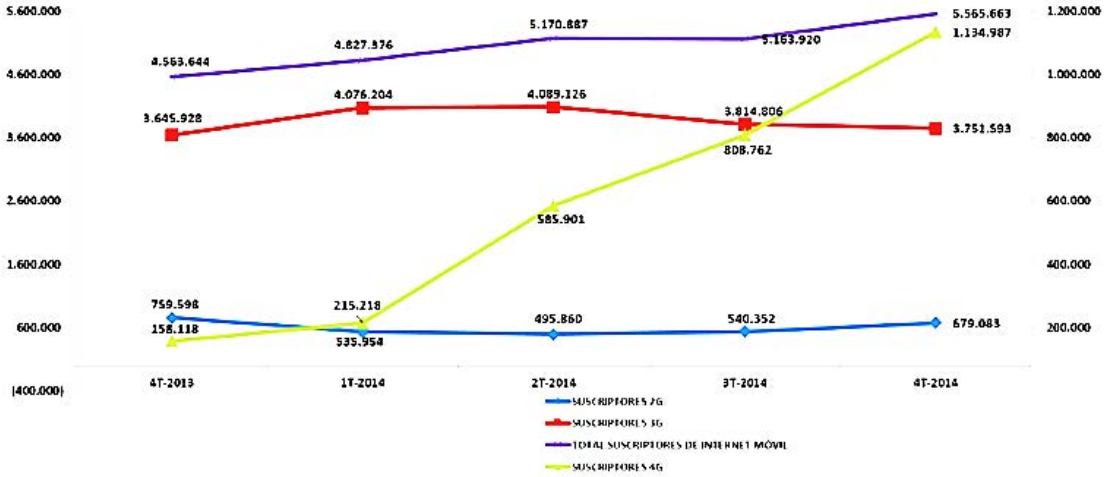


Figura 3: Suscriptores a Internet móvil por generación en 2014 [51]

Según José Otero, director de 4G Américas para Latino América y el Caribe, Colombia lidera el crecimiento en la adopción de 4G LTE en América Latina. A tal punto que se espera para 2019 que el país tenga más de 12 millones de suscriptores con Teléfonos

[19] La diferencia entre Suscriptor y Abonado: Por suscripción hace referencia a aquellos usuarios que pagan un cargo fijo mensual por el servicio de acceso a Internet (Suscriptores), y por demanda hace referencia a cualquier tipo de modalidad de servicio de acceso a Internet en la que no se pague un cargo fijo mensual (abonados) [51].

Inteligentes compatibles con LTE [52]. Para mantener tal posición se revela la urgencia de no solamente crecer en número de abonados y suscriptores sino también de establecer políticas de regulación que garanticen la gestión y el monitoreo de la tecnología, con el soporte de recomendaciones internacionales actualizadas y compatibles con la implementación de los servicios de datos requeridos por los usuarios y empresas.

5.2 Síntesis de la problemática de no contar con regulación propia para 4G

En Colombia existen diversas regulaciones y normas representadas en resoluciones, decretos y leyes con el fin de garantizar la protección al usuario, la cobertura y la calidad del servicio en las telecomunicaciones del país. Específicamente para las comunicaciones móviles se observa énfasis en normativas que buscan asegurar transmisión de datos y voz, mediante la aplicación de recomendaciones de entidades internacionales encargadas de determinar los aspectos técnicos a aplicar.

Las normas para protección al usuario se enfocan en definir y garantizar sus derechos frente a los servicios que ofrecen los operadores móviles. Por otro lado, aunque no se encontraron reglas específicas para cobertura, es posible determinar, especialmente en el caso de Claro, que la cobertura para 4G en Colombia se encuentra concentrada en las zonas urbanas de las principales ciudades.

En el estudio detallado de estos aspectos técnicos incluidos dentro de las normas mencionadas, se revela una significativa cantidad de recomendaciones ITU que cuentan con fechas de publicación y última actualización de hace más de diez años, lo que evidencia que fueron hechas principalmente para satisfacer las necesidades de la generación móvil de la época, y por tanto podrían no ser las más adecuadas para determinar un marco regulatorio para LTE, tecnología que data de años más recientes.

En el Decreto 1130 de 1999, artículo 37 numerales 3 y 4, se ordena expedir toda regulación general o particular que permita establecer parámetros de calidad de los servicios y fijar indicadores, metas de calidad y eficiencia de los servicios, e imponer índices de calidad, cobertura y eficiencia a los operadores [2]; y la Resolución 3503 de 2011 [53] determina la inclusión de nuevos parámetros para la verificación de la calidad en los servicios de datos a través de redes móviles. Lo anterior quiere decir que no se estarían cumpliendo estas disposiciones dictadas por la ley colombiana en lo concerniente a 4G, y que las recomendaciones están desactualizadas en relación al avance tecnológico en comunicaciones móviles (más de diez años de antigüedad), por lo que no se ajustan a los nuevos requerimientos.

Lo anterior deja expuesta la problemática de no tener una normatividad propia para LTE, aun incumpliendo reglas determinando que debe existir tal regulación, sino también que las normas existentes se están quedando obsoletas para garantizar los servicios en las tecnologías anteriores a 4G.

5.3 Recomendaciones base para un marco regulatorio en Colombia

Como se describió antes, Colombia lidera el crecimiento en la adopción de 4G LTE en América Latina. No obstante existe la problemática de la ausencia de una regulación específica para esta tecnología y que su operación se ha estado apoyando en normas existentes pero diseñadas en principio para las generaciones anteriores 2G y 3G. Por lo tanto, en la Tabla 1 se presentan recomendaciones que pueden servir para ayudar a tener una mayor eficiencia en la gestión del espectro, y a formular un nuevo grupo de normas para LTE, con el objetivo de colaborar al mejoramiento de la calidad del servicio.

Tabla 1: Recomendaciones para futura regulación de 4G en Colombia

Colombia	América Latina	Entidades internacionales
Decreto 1130 de 1999	Resolución N° 37 de 2014 Argentina	Spectrum Sharing – 4G Americas
Resolución 3503 de 2011	Plan Nacional de Banda Ancha EE.UU	Reglamento de Radiocomunicaciones I a IV Manual de Comprobación Técnica del Espectro – ITU
Resolución 3101 de 2011	El espectro radioeléctrico en México. Estudio y acciones	ITU-R M.1036-4
Dimensión Estratégica I, II, III y IV		ITU-R M.2012-1 que relaciona Recomendaciones, Resoluciones y Reportes: ITU-R M.1036; ITU-R M.1224; ITU-R M.1645; ITU-R M.1822; ITU-R M.2038; ITU-R M.2072; ITU-R M.2074; ITU-R M.2078; ITU-R M.2079; ITU-R M.2133; ITU-R M.2134; ITU-R M.2135-1; ITU-R M.2198; ITU-R 56-1; ITU-R 57-1
		Overview of 3GPP Release 13 LTE para seguridad pública Series de especificaciones 21.xx a 37.xx – 3GPP
		IEEE 802.16 IEEE 802.16.1 IEEE 802.16.1.a

6 Conclusiones

La ausencia de normas que regulen el monitoreo y desempeño de 4G en Colombia hace urgente la creación de éstas para poder acompañar el crecimiento de esta tecnología, que ya cuenta con millones de usuarios en el país. Por lo cual, para ayudar en el proceso de elaboración de estas reglas se presenta el conjunto de recomendaciones mediante el estudio de la situación regulatoria actual de Colombia y otros países de Latino América, y las regulaciones, estándares y especificaciones de las entidades más importantes en tecnologías de telecomunicaciones.

Se evidencia la necesidad por parte de los gobiernos de los diferentes países de la región y en el resto del mundo con acceso a LTE de incluir normas y planes que ayuden a gestionar y monitorear el desempeño de la tecnología. Algunas entidades encargadas de establecer las normas en telecomunicaciones de diferentes países ya se encuentran en proceso de regular y establecer planes de acción y gestión de espectro radioeléctrico para este propósito, determinando de esta manera el camino que asegurará el futuro de las comunicaciones móviles y la banda ancha inalámbrica, con velocidades y calidad de servicio óptimas.

IEEE no habla específicamente de estándares para 4G, porque cuenta con WiMax como tecnología propia. Sin embargo, para el propósito mencionado, fortalece en gran manera tener en cuenta la aplicación de algunos estándares de esta entidad internacional, con el fin de establecer un marco regulatorio dentro de un país en vía de desarrollo, y que se encuentra en proceso de implementación de LTE.

Algunos de los estándares y recomendaciones internacionales empleados en las normas actuales, que reglamentan las comunicaciones móviles en el país, datan de años muy anteriores a la implementación de la tecnología 4G, lo que podría provocar incompatibilidades con los nuevos sistemas, afectando la creciente cantidad de usuarios. Además, para que Colombia pueda mantenerse a la vanguardia en el crecimiento de 4G LTE en América Latina es urgente no solamente crecer en número de abonados y suscriptores, sino también de establecer políticas de regulación que garanticen la gestión

y el monitoreo de la tecnología en mención, con el soporte de recomendaciones internacionales actualizadas y compatibles con la implementación de los servicios de datos requeridos por los usuarios y empresas.

La problemática que implica no contar con regulación dedicada a LTE en Colombia debe ser abordada mediante la formulación de nuevas normas contenidas en leyes, resoluciones y decretos que permitan asegurar la máxima calidad de servicio, cumpliendo con los reglamentos dispuestos por las entidades nacionales e internacionales.

Agradecimientos

Este trabajo fue apoyado por el grupo de investigación Automática, Electrónica y Ciencias Computacionales, del Instituto Tecnológico Metropolitano en Medellín, y en el marco de los proyectos de investigación P14103 y P14215.

Referencias

[1] Academy ITU. (2014). Curso Básico de LTE (Long Term Evolution). Online.
[2] CRT. (2007). Indicadores de calidad en servicios de telecomunicaciones. Comisión de Regulación de las Telecomunicaciones. Online.
[3] CRT. (2007). Indicadores de calidad para servicios de telcos. Comisión de Regulación de las Telecomunicaciones. Resolución 1740.
[4] Vive Digital. (2013). Proceso Subasta 4G, Condiciones finales. Online.
[5] Vive Digital. (2013). Alternativas Subasta AWS. Online.
[6] CRT. (2012) Modificación de condiciones del régimen calidad en servicios de telecomunicaciones. Comisión de Regulación de las Telecomunicaciones. Online.
[7] MinTIC. (2013). Resolución 449. Ministerio de Tecnologías de la Información y las Comunicaciones. Online.
[8] CRC. (2014) Informes Calidad Julio 2014-Septiembre 2014. Comisión de Regulación de las Comunicaciones. Online.
[9] CRC. (2011). Resolución 3101. Comisión de Regulación de las Comunicaciones. Online.
[10] ITU. (1992). ITU-T E.170. International Telecommunication Union. Online.
[11] ITU. (1988). Cálculo del número de circuitos de un haz utilizado para cursar el tráfico de desbordamiento. International Telecommunication Union. Online.
[12] ITU (1988). Número de circuitos en un haz de gran utilización. International Recommendation Union. Online.
[13] ITU. (2007). UIT-T Y.2701. International Telecommunication Union. Online.
[14] ITU. (1993). ITU-T serie Q.700. International Telecommunication Union. Online.
[15] ITU. (2001). Serie G sistemas y medios de transmisión sistemas y redes digitales calidad de servicio y de transmisión. International Telecommunication Union. Online.
[16] ITU. (2001). G.1010 Categorias de calidad de servicio para los usuarios de extremo de servicios multimedios. Serie G. International Telecommunication Union. Online.
[17] ITU. (2001). ITU-T serie Q.500. International Telecommunication Union. Online.
[18] ITU. (2001). Características de la calidad de transmisión de los canales de modulación por impulsos codificados. International Telecommunication Union. Online.
[19] ITU. (2012). Digital network echo cancellers. International Telecommunication Union. Online.
[20] ITU. (2007). Transmission impairments due to speech processing. International Recommendation Union. Online.
[21] ITU. (2004). Relaciones entre las Recomendaciones sobre la calidad de funcionamiento de la red digital de servicios integrados (RDSI), las redes basadas en IP y la capa física Recomendación. International Telecommunication Union. Online.
[22] ITU. (2009). ITU-T P.310. International Telecommunication Union. Online.
[23] ITU. (2011). Transmission characteristics for wideband digital handset and headset telephones. International Telecommunication Union. Online.
[24] ITU. (2011). Transmission characteristics for wideband digital loudspeaking and hands-free telephony terminals. International Telecommunication Union. Online.
[25] ITU. (2011). IP packet transfer and availability performance parameters. International Telecommunication Union. Online.

[26] ITU. (2011). Network performance objetives for IP-based services. Serie Y: Global information infrastructure, internet protocol aspects and next-generation networks Internet protocol aspects – Quality of service and network performance. International Telecommunication Union. Online.
[27] ITU. (1988). Objetivos de tasa de deslizamientos controlados en una conexión digital internacional. International Telecommunication Union. Online.
[28] ITU. (1997). Sistemas de transmisión digital – Redes digitales – Objetivos de diseño para las redes digitales. International Telecommunication Union. Online.
[29] ITU. (1993). Modulación por impulsos codificados (MIC) de frecuencias vocales. International Telecommunication Union. Online.
[30] ITU. (2007). Signalling requirements and protocols for the NGN – Service and session control protocols. International Telecommunication Union. Online.
[31] ITU. (2003) Conexiones y circuitos telefónicos internacionales – Definiciones generales Plan. International Telecommunication Union. Online.
[32] MinTIC (2015). Resoluciones MinTIC. Ministerio de las Teconologías de las comunicaciones. Online.
[33] ANE. (2015). Plan de Acción 2015. Agencia Nacional del Espectro. Online.
[34] ANE (2015). Dimensión Estratégica I. Agencia Nacional del Espectro. Online.
[35] MPF. (2013). Resolución Nº 5/2013. Ministerio de Planificacion Federal, Inversion Pública y Servicios, Argentina. Online.
[36] MECON. (2014). Resolución Nº 17/2014. Ministerio de Economía y Finanzas Públicas, Argentina. Online.
[37] MECON. (2014). Decreto 671/2014. Ministerio de Economía y Finanzas Públicas, Argentina. Online.
[38] MECON. (2014). Resolución Nº 37/2014. Ministerio de Economía y Finanzas Públicas, Argentina. Online.
[39] IFT. (2012). El espectro radioeléctrico en México. Estudio y acciones. Instituto Federal de Telecomunicaciones, México. Online.
[40] DOF. (2013). DOF: 11/06/2013. Diario Oficial de la Federación, México. Online.
[41] DOF. (2014). Ley federal de telecomunicaciones y radiodifusión. Diario Oficial de la Federación, México. Online
[42] Bray, D. (2015). Modernizing the FCC's IT. Online.
[43] ITU. (2012). Recommendation ITU-R M.1036-4. International Telecommunication Union. Online.
[44] ITU. (2014). ITU-R Detailed specifications of the terrestrial radio interfaces of International Mobile Telecommunications-Advanced (IMT-Advanced). International Telecommunication Union. Online.
[45] 3GPP. (2014). Overview of 3GPP Release 13. Online.
[46] 3GPP. (2014). LTE for Public Safety (authority-to-authority) communications_20140316. Online.
[47] 3GPP. (2015). Series de especificaciones 3GPP. Online.
[48] IEEE. (2012). IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. IEEE Computer Society. Online.
[49] IEEE. (2012). Standard forWirelessMAN-Advanced Air Interface for Broadband Wireless Access Systems. IEEE Computer Society. Online.
[50] IEEE. (2013). Standard for WirelessMAN-Advanced Air Interface for Broadband Wireless Access Systems. IEEE Computer Society. Online.
[51] MinTIC. (2015). Boletín trimestral de las TIC - Cuarto trimestre de 2014. Ministerios de las Tecnologías y las Comunicaciones. Online.
[52] Moss, L. (2015) , Colombia Leading Growth of 4G LTE Adoption In Latin America; Exclusive Interview With 4G Americas' José Otero. Online.
[53] CRC. (2011). Resolución 3503 de 2011. Comisión de Regulación de las Comunicaciones. Online.
[54] CRC. (2013). Resolución 4112. Comisión de Regulación de las Comunicaciones. Online.
[55] FCC. (2015). Title 47 - Telecommunication. Federal Communications Commission. Online.

Capacidad del Ansys Fluent® V 16.0 para determinar la fuerza de arrastre en partículas no esféricas tipo discos cilíndricos

Jorge Sierra[1]
Mohammed Khalij[2]
Anne Tanière[3]
Diego Hincapié[4]
Sebastián Vélez[5]
[1, 4, 5] *Instituto Tecnológico Metropolitano – Colombia*
[2,3] *Université de Lorraine – Francia*

1 Introducción

La capacidad de prever el comportamiento de un fluido multifásico es importante a la hora de realizar un proceso de optimización. Estos fluidos están conformados por una fase dispersa que puede ser líquida, sólida y/o gaseosa y una fase continua compuesta por un líquido y/o un gas [1]. Tanto en los procesos naturales como industriales, la dinámica de las partículas contenidas en la fase dispersa determinan el comportamiento global del fluido multifásico. La dispersión de contaminantes en el ambiente, el transporte neumático, las cámaras de combustión de carbón pulverizado, la sedimentación de minerales en ríos, entre otros, son ejemplos de las aplicaciones naturales e industriales en donde la optimización o caracterización de la dinámica de las partículas juega un papel importante al momento de diseñar u optimizar un proceso de caracterización.

La hipótesis de considerar las partículas como esferas ha sido implementada durante los últimos 20 años, y resulta conveniente debido a la simplicidad en el cálculo de la fuerza de arrastre y debido a la cantidad de experimentaciones realizadas, que describen el comportamiento del fluido cuando interactúa con la esfera, tanto para la zona de flujo laminar como turbulento. La relación entre el coeficiente de arrastre para una esfera y el número de Reynolds se puede encontrar en el trabajo de Schiller y Naummann [2], que ha sido validado numéricamente.

Desde 1920 se han realizado diferentes trabajos teóricos y experimentales relacionados con las fuerzas hidrodinámicas sobre partículas no esféricas. Una compilación de las correlaciones disponibles para el cálculo del coeficiente de arrastre de una partícula no esférica sometida a un flujo de un fluido incompresible se encuentra en [3]. En este trabajo se encontró que los principales inconvenientes a la hora de calcular las fuerzas hidrodinámicas radican en la dificultad que se presenta en la determinación exacta de la superficie de la partícula, así como de la dimensión y la posición instantánea de la partícula respecto a la dirección del fluido. En la literatura existen dos tipos de aproximaciones para determinar el coeficiente de arrastre de una partícula no esférica: 1) el primero consiste en desarrollar expresiones para calcular la fuerza de arrastre para una partícula de geometría única y orientación fija en función del Re, conocidas como correlaciones para $C_D = f$ (Re, geometría). De acuerdo con este tipo de correlación se han realizado algunos trabajos sobre partículas no esféricas y en discos [4, 5]. Las correlaciones obtenidas en estas experiencias ofrecen un margen de error aceptable, sin embargo, no se permite extrapolar hacia otra geometría debido al error que se produce en el cálculo de la fuerza de arrastre. Por esta razón, el primer método resulta insuficiente a nivel general, debido a que no se considera la orientación de la partícula y solamente aplica para una geometría específica. 2) La segunda metodología surge con el fin de suplir

las falencias para el cálculo de la fuerza de arrastre, una correlación que tiene en cuenta la orientación de la partícula y que permite extrapolar hacia otras geometrías. Esta aproximación se realiza mediante la implementación de diferentes parámetros adimensionales, necesarios para determinar la orientación de la partícula, la superficie exterior y la sección transversal respecto a la dirección del fluido. Estas expresiones se conocen como $C_D = f$ (Re, factores de forma). Esta metodología, a pesar de que puede ser utilizada para diferentes geometrías, tiende a ser menos precisa que la primera, debido a la dificultad que se presenta en la medición de la superficie de las partículas no esféricas, sumado a que dichos factores de forma no representan exactamente la geometría irregular de la partícula, ni el comportamiento real del fenómeno de transporte.

La primera correlación propuesta en función del Re y de factores de forma, introdujo el concepto de factor de esfericidad Ø, que es la relación en el área superficial de la esfera de volumen equivalente al de la partícula $s = \frac{\pi}{4} {d_{sph}}^2$, respecto al área superficial de la partícula no esférica S [6]. El cálculo de la fuerza de arrastre para una partícula no esférica sometida a un flujo incompresible, Newtoniano, se calcula mediante la correlación (1):

$$C_D = \frac{24}{Re}(1 + ARe^B) + \frac{C}{1+\frac{D}{Re}} \qquad (1)$$

$$Re = Re_p = \frac{\|u_R\| d_{sph}}{\nu} \qquad (2)$$

$$Ø = \frac{s_e}{S} \qquad (3)$$

Donde:

$A = \exp(2.3288 - 6.4581Ø + 2.4486Ø^2)$
$B = 0.0964 + 0.5565Ø$
$C = \exp(4.905 - 13.8944Ø + 18.42222Ø^2 - 102599Ø^3)$
$D = \exp(1.4681 + 12.2584Ø - 20.7322Ø^2 + 15.8855Ø^3)$
$C_D = Coeficiente\ de\ arrastre.$
$Re_p = Reynolds\ partícula$
$d_{sph} = Diámetro\ de\ la\ esfera\ de\ volumen\ equivalente$
$\|u_R\| = velocidad\ relativa\ entre\ fluido\ y\ la\ partícula$
$\nu = viscosidad\ cinematica\ del\ fluido$

Esta correlación presenta una desviación máxima de 0.024 para el cálculo de la fuerza y el coeficiente de arrastre C_D, en partículas no esféricas con un factor de esfericidad Ø>0.67. Sin embargo, para partículas con Ø<0.67 presenta una desviación entre 0.049 para un factor de esfericidad Ø=0.230 y 0.107 para un factor de esfericidad Ø=0.026, lo cual se considera aceptable. En la Figura 1 se muestra la dependencia del coeficiente de arrastre con el Re.

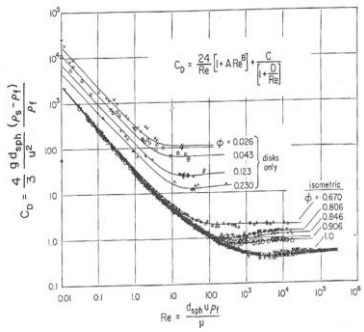


Figura 1: Coeficiente de arrastre de partículas esféricas y no esféricas en función del Re y del factor de esfericidad Ø. Adaptado de [6][6]

Se han realizado clasificaciones de las principales correlaciones de la segunda metodología, con el fin de clasificar las existentes para determinar el coeficiente de arrastre en partículas no esféricas, entre los cuales se destaca la de Hartman et al. [3]. Estos autores realizaron una validación de la confiabilidad de cada una de estas correlaciones, en la determinación de la fuerza de arrastre para diferentes partículas no esféricas, tales como discos, cilindros, elipsoides entre otros y respecto a datos experimentales disponibles en la literatura. Según ellos, la correlación más apropiada para el cálculo de la fuerza de arrastre en discos es la correlación (1), propuesta por Haider y Levenspienl [6], que presenta un error promedio de 14.9% y un error máximo del 45.5% en el cálculo de la fuerza de arrastre en discos.

En 2008 se presenta una correlación general en función del Re y de Factores de forma ($\emptyset, \emptyset_{\perp}$) en las que se considera la orientación de la partícula respecto a la dirección del fluido, la forma y el tamaño, para calcular la fuerza de arrastre en una partícula de forma arbitraria [7].

$$C_D = \frac{8}{Re}\frac{1}{\sqrt{\emptyset_{\perp}}} + \frac{16}{Re}\frac{1}{\sqrt{\emptyset}} + \frac{3}{\sqrt{Re}}\frac{1}{\emptyset^{3/4}} + 0.4210^{0.4(-log\emptyset)^{0.2}}\frac{1}{\emptyset_{\perp}} \qquad (4)$$

Donde la esfericidad Ø es la relación entre el área superficial de la esfera de volumen equivalente respecto al área superficial de la partícula considerada. La esfericidad transversal ($\emptyset_{\perp}$) es la relación entre la sección transversal de la esfera de volumen equivalente respecto a la sección transversal de la partícula considerada. Según los autores, esta correlación es la que ofrece mejor precisión en el cálculo de la fuerza de arrastre de partículas no esféricas, presentando una error promedio de 14.1% respecto a datos experimentales disponibles en la literatura. Sin embargo, para partículas no esféricas como discos, esta correlación presenta una desviación relativa máxima de 88% y una desviación media de 17.7% respecto a los datos experimentales.

En 2012 se presentó un estudio en el que se determinan, mediante simulación computacional, los coeficientes de arrastre, levante y torque de cuatro geometrías axisimétricas, tales como discos, fibras y elipsoides [8]. Este estudio se realizó mediante simulación numérica directa (DNS), que se basa en resolver las ecuaciones de cantidad de movimiento y conservación de masa en el volumen de control discretizado a una escala menor a la escala de Kolmogorov. A pesar de las ventajas que presenta este tipo de simulación, el costo computacional es considerablemente alto, comparado con las técnicas utilizadas en el presente estudio.

El objetivo de esta investigación es determinar la confiabilidad del software comercial ANSYS Fluent® V16.0 mediante el cálculo de las fuerzas hidrodinámicas (arrastre, levante y torque), en partículas no esféricas sometidos a flujos de fluidos Newtonianos. En este trabajo se realiza el estudio de validación del programa para determinar la fuerza de arrastre en discos cilíndricos, como partícula no esférica, a una orientación fija de β=90° respecto a la dirección del fluido. Un segundo estudio será realizado para determinar la influencia de la orientación de la partícula en la fuerza de arrastre experimentada por la partícula.

2 Dinámica traslacional de partículas no esféricas modelo

Existen pocos modelos que describen la dinámica de una partícula no esférica sometida al flujo de un fluido. Estas partículas están sometidas tanto a una rotación como a una traslación simultáneamente en la mayoría de los casos. En un marco de referencia Lagrangiano, la traslación de la partícula se determina por la ecuación de movimiento [8]:

$$m_p \frac{D\vec{v}_p}{Dt} = \overrightarrow{F_D} + V_p(\rho_p - \rho_f)\vec{g} + \overrightarrow{F_{PG}} + \overrightarrow{F_{VM}} + \overrightarrow{F_L} \quad (5)$$

Donde:
$m_p = masa\ de\ la\ partícula$
$V_p = Volumen\ de\ la\ partícula$
$\rho_{p;}\rho_f = Densidad\ de\ la\ partícula - fluido$
$v_p = Vel.translacional\ centro\ de\ masa\ de\ la\ partícula$
$F_D = Fuerza\ de\ arrastre$
$F_{PG} = Fuerza\ por\ gradiente\ de\ presión$
$F_{VM} = Fuerza\ de\ masa\ virtual$
$F_L = Fuerza\ de\ levante$

2.1 Fuerza de arrastre $\overrightarrow{F_D}$

La fuerza de arrastre que experimenta una partícula se debe al gradiente de presión generado por la perturbación del fluido, debido a la partícula y a los esfuerzos viscosos por rozamiento entre el fluido y la superficie de la partícula. Estas fuerzas se representan por la expresión (6):

$$\|\vec{F}_D\| = F_{D,PRESIÓN} + F_{D,FRICCIÓN}$$
$$\|\vec{F}_D\| = -\int_S p\vec{x}\cdot\vec{n}dS + \int_S \vec{n}\cdot\tau\cdot\vec{x}dS \quad (6)$$

Esta fuerza está caracterizada por el coeficiente de arrastre denotado por C_D, que depende físicamente del régimen del fluido cuantificado por $Re_p = \|\overrightarrow{u_R}\|d_{sph}/v$; donde $\|\overrightarrow{u_R}\| = \overrightarrow{u_p} - \overrightarrow{u_f}$ representa la velocidad relativa entre la velocidad de la partícula $\overrightarrow{u_p}$ y la velocidad del fluido $\overrightarrow{u_f}$, d_{sph} representa el diámetro de la esfera de volumen equivalente y v la viscosidad cinemática del fluido. La orientación de la partícula respecto a la dirección del flujo se denotada por el ángulo β, que se forma por el eje de simetría de la partícula y la dirección del fluido. Este criterio aplica para partículas axisimétricas en un marco de referencia simplificado 2D. El coeficiente de la fuerza de arrastre se puede determinar mediante la ecuación (7):

$$C_D = \frac{\|\overrightarrow{F_D}\|}{\frac{1}{2}\rho_f u_R^2 \frac{\pi}{4} d_{sph}^2} \quad (7)$$

3 Análisis numérico computacional

Se plantea el estudio con el fin de validar y determinar los coeficientes de arrastre del disco de la Figura 2. Inicialmente mediante simulación numérica computacional de fluidos (CFD) se calculan las fuerzas y los coeficientes de arrastre con base en la ecuación (7), según los parámetros descritos en la Tabla 1. Luego se realiza una comparación de los valores obtenidos numéricamente respecto de los datos experimentales con los cuales se realizó el ajuste de la correlación (1).

Tabla 1: Parámetros de simulación.

Re_p	Partícula	Rel. Aspecto H/Dp	Factor de esfericidad Ø	Orientación β
1 40 70 100 180 300	Disco cilíndrico	0.028	0.23	90°

3.1 Modelo matemático

En este estudio se considera el fluido isotérmico, incompresible, newtoniano de flujo uniforme y estacionario. No se consideran efectos de pared. Bajo estas consideraciones las ecuaciones que rigen el movimiento del fluido son:

$$\nabla \cdot \vec{u} = 0 \tag{8}$$

$$(\vec{u} \cdot \nabla)\vec{u} = -\frac{1}{\rho_f}\nabla p + v\Delta\vec{u} \tag{9}$$

Donde p representa la presión del fluido.

3.2 Geometría, volumen de control y condiciones de la simulación

La geometría de la partícula analizada es un disco cilíndrico, sus dimensiones y parámetros de forma se muestran en la Figura 2.

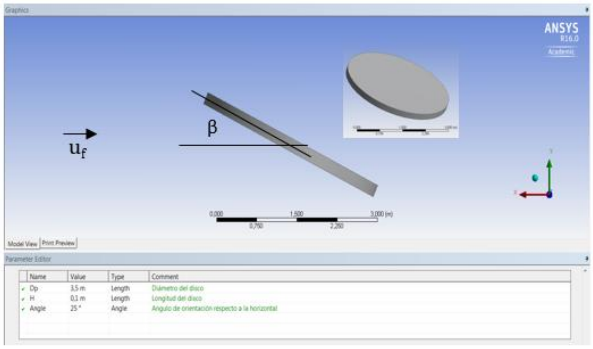


Figura 2: Configuración del disco con factor de esfericidad Ø=0.23

El volumen de control está conformado por un cilindro envolvente de dimensiones previamente establecidas, para evitar efectos de capa límite de la pared del cilindro envolvente sobre el flujo cercano a la partícula. La longitud es lo suficientemente larga para permitir el desarrollo de la estela aguas abajo de la partícula, como se muestra en la Figura 3. El proceso de discretización del dominio (mallado) se conformó por elementos tetrahedrales y se implementó un refinamiento adaptativo por gradiente de presión normalizado de 0.5 cada 20 iteraciones, esto con el fin de garantizar que los elementos tengan el tamaño necesario para capturar los gradientes de presión alrededor de la partícula y así garantizar precisión en el cálculo de la fuerza de arrastre. La independencia del mallado se determinó cuando las variaciones en la fuerza de arrastre eran inferiores al 3% respecto del valor anteriormente calculado, así finalmente se determinó que aproximadamente 650.000 elementos eran suficientes para la realización de las simulaciones. Las simulaciones se realizaron en el programa ANSYS Fluent ® V16.0 versión académica. Las condiciones de frontera consideradas se presentan en la Figura 3.

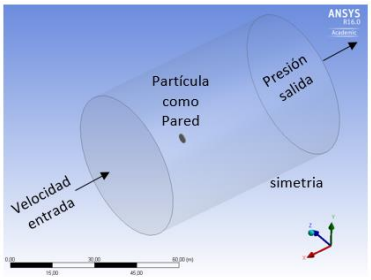


Figura 3: Configuración del volumen de control y condiciones de frontera

4 Análisis de resultados

Los resultados del estudio donde se realiza una comparación de los coeficientes de arrastre C_D para una partícula cilíndrica, con factor de esfericidad de Ø=0.23, orientación β=90° respecto a la dirección del flujo obtenidos mediante simulación numérica computacional, se ajustan satisfactoriamente con los resultados obtenidos para la misma partícula usando la correlación propuesta por (1). Esta correlación es la más apropiada para el cálculo del coeficiente de arrastre en discos con factor de esfericidad Ø=0.23 [3]. El error que se presentó es del 10%, que es aceptable si se considera que el flujo aguas abajo de la partícula a cierto número de Re_p se puede convertir en un flujo inestacionario; no se conoce por parte de los autores el régimen de flujo para el cual la estela aguas abajo de la partícula se convierte en un flujo inestacionario.

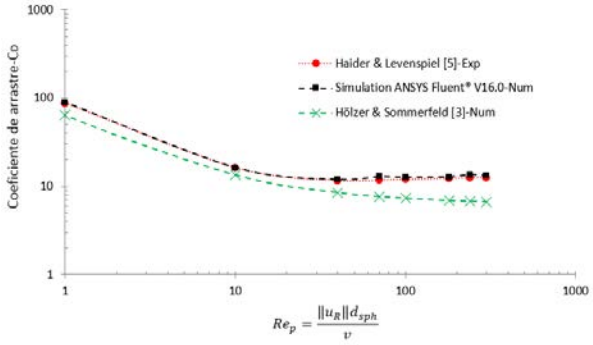


Figura 4: Coeficiente de arrastre en función del Re_p para una orientación de β=90° respecto a la dirección del fluido

En la Figura 4 se realiza una comparación de los resultados obtenidos del proceso de simulación y los valores obtenidos por las correlaciones (1) y (4). La correlación propuesta (4) presenta subvaloración de los coeficientes de arrastre, que se debe a los datos experimentales utilizados en el proceso de ajuste de esta correlación, ya que se utilizó una gran cantidad de partículas con factor de esfericidad superiores a Ø= 0.67 y cercanos a 1 [7]. Caso contrario ocurre con la correlación (1), en el que en su proceso de ajuste de constantes se incluyeron partículas con factor de esfericidad Ø=0.23, igual al de la partícula considerada en este estudio y tal como se puede observar en la Figura 1.

5 Conclusiones

Se calcula la capacidad del programa ANSYS Fluent® V16.0 como satisfactoria para determinar la fuerza de arrastre en una partícula no esférica tipo disco cilíndrico, con factor de esfericidad Ø=0.23 y sometida a un flujo de régimen $1 < Re_p < 300$, donde se presenta un error máximo del 10% que se considera aceptable para este tipo de aplicación.

Se requieren trabajos futuros para validar el programa bajo diferentes configuraciones. Simulaciones en las cuales se tengan como variables la orientación de la partícula respecto a la dirección del fluido, diferentes relaciones de aspecto de la partícula con el fin de variar el factor de esfericidad Ø y aumento en el régimen determinado por el Re_p para encontrar, en función de la orientación, a qué régimen de flujo se vuelve inestacionaria la estela de la partícula aguas abajo. El acompañamiento de estas simulaciones con trabajos experimentales, mediante técnicas de medición no invasivas (PIV, LDV, DLHM), conllevaría a correlaciones numérico-experimentales más precisas a la

hora de calcular las fuerzas hidrodinámicas en partículas no esféricas, y de esta manera optimizar y/o caracterizar tanto procesos naturales como industriales que implican la dinámica de fluidos multifásicos.

Referencias

[1] Oesterlé, B. (2006). Ecoulement multiphasiques-Des fondements aux methods d'ingénierie. Paris: Lavosier.

[2] Ouchene, R. et al. (2015). Drag, lift and torque coefficients for ellipsoidal particles: from low to moderate particle Reynolds numbers. Computers & Fluids 113, pp. 53-64.

[3] Chhabra, R. et al. (1998). Drag on non- spherical particles: an evaluation of availables methods. Powder Technologies 101, pp. 288-295.

[4] Davis, A. (1990). Stokes drag on a disk sedimenting toward a plane or with other disks; additional effects of a side wall or free surface. Physics of Fluids A 2(3), pp. 301-312.

[5] Michael, P. (1966). Steady motion of a disk in a viscous fluid. Physics of Fluids 9(3), pp. 466-471.

[6] Haider, A. & Levenspiel, O. (1989). Drag coefficient and terminal velocity of spherical and nonsphericals particles. Powder technologies 58, pp. 63-70.

[7] Hölzer, A. & Sommerfeld, M. (2008). New simple correlation formula for the drag coefficient of non-spherical particles. Powder technologies 184, pp. 361-365.

[8] Zastawny, M. et al. (2012). Derivation of drag and lift force and torque coefficients for non-spherical particles in flows. International Journal of Multiphase Flow 101, pp. 288–295.

Custom processors design using Python-based high level synthesis

Jaime Parra P.
Institución Universitaria Antonio José Camacho – Colombia

1 Introduction

Embedded systems are reactive systems that interact intensively with their environments and users, must run multiple tasks in parallel and must, at times, react in real time [1]. Added to this is that they should also have low power consumption and occupy little volume, both desirable characteristics to enable portability, key in mobile computation nowadays [2].

The three most popular technologies for implementing embedded systems are Application-specific integrated circuits (*ASIC*), microcontroller, and reconfigurable hardware (*FPGA,* for field-programmable gate arrays) [3]. ASICs are used when the number of units to be manufactured justifies it and/or the application is very particular and its presence in market is estimated to be from medium to long term. Microcontrollers are the usual choice for low production volumes and/or medium or simple applications [4]. FPGAs are chosen when dealing with applications with conflicting requirements and/or in the case of experimental research applications [5].

A widespread way to design solutions for the FPGA case is posing a description using a Hardware Description Language (HDL), which allows the designer to think in terms of high-level abstractions related to meeting the requirements, while provides constructions associated with the hardware and the physical world which are not usually present in conventional programming languages, that are focused on microprocessors programming [6].

While designing in an HDL can achieve much more complex systems than those using only basic digital systems tools such as truth tables or schematics, it is also clear that existing HDLs have certain characteristics that hinder their wider acceptance and use. The four that stand out are: excessive verbosity, strong data typing, little encapsulation, and very explicit connection with hardware [7].

The solution proposed in this paper is the creation of a computing environment, called PyHDL, which allows the designer to describe his/her solution using Python, a contemporary programming language which is valid to represent parallel and interactive efficient solutions. Since Python is a language focused on programming conventional computers it is not directly synthesizable on hardware, hence PyHDL covers this aspect, generating a VHDL equivalent of the description given in Python, which can be easily synthesizable on ASIC or FPGA hardware [8].

PyHDL is therefore an EDA (*Electronic Design Automation*) tool, conceived to assist the designer of complex digital systems, especially specific-purpose processors. The tool is a cross-compiler, whose primary purpose is to automate the tasks associated with the generation of the several files required by tools that make simulation, synthesis and implementation. Thus the tool frees the designer of those details that can be considered of low level, and lets him/her focus on the more abstract aspects of the system and express them through a contemporary high-level language, Python [9]. PyHDL receives a file written in pure conventional Python and processes in order to extract a hierarchy of structures from the usual constructions in a programming language: functions, loops,

decisions and assignments. The outcome generated by PyHDL is a set of files that can be employed directly by EDA tools, for automating the electronic design.

EDA tools focus primarily on one or more of the four fundamental aspects that are part of the design flow: modeling, simulation, synthesis, and implementation [10]. In modeling the designer takes the ideas from which he/she conceives the solution to a given problem and expresses them through some form of representation. In the case of digital systems, there are different types of representations, ranging from the closest to hardware such as schematics to those of algorithmic type, passing through intermediate forms such as truth tables, state diagrams or Petri nets.

Simulation involves running an application that provides an equivalent of the environment in which the solution would run with the same input or excitation vectors, to which the solution generates answers according to the modeling given to its behavior. The simulator presents this information to the user or can take care of carrying out some analysis in order to identify possible flaws in the system behavior. Synthesis consists in mapping the representation of the solution to some kind of hardware. The usual alternatives are either a custom-made device, or ASIC, or a prototype using reconfigurable hardware, or FPGA. In both cases the job of the synthesis is splitting the representation into small modules and fit them in such a way that matches the alternatives present in a component library. Finally, implementation materializes that synthesis, leading to a real hardware on which the solution can be run [11].

Of the four stages mentioned, the last three (simulation, synthesis, and implementation) are fairly mature, to a lesser extent the first one. Moreover, given the enormous complexity of today's digital systems, it is almost impossible to perform these tasks by a human being. That is the reason why improvement is almost an exclusive domain of large manufacturers of EDA tools. That is not the case with the first stage, modeling, which, because of being the closest to human beings, and in fact the most critical, since from its outcome depends the others, still has much improvement work.

PyHDL intervenes in this stage of modeling, providing an interface for the user to translate their ideas into an algorithmic model written in the Python programming language, taking advantage of its remarkable educational characteristics, which have proved to be of great influence on the quality of the model generated [12]. Therefore it seeks that obtaining a functional solution be achieved in less time and that the developed product be more robust and easier to analyze, with a view to future improvements or additions or expansions.

2 Processors

Computation is physically performed by processors. A processor takes certain input data and transforms them into their respective output data according to the rules dictated by the algorithm that directs the activities of the processor. In general, the most widely spread model for the processor represents it as made of two modules: a controller and a datapath [13]. The controller is a finite state machine that follows the algorithm, while the datapath is an arithmetic/logic execution unit (ALU) where data transformations are carried out. According to the way these two modules are related, processors are classified into one of three groups: general purpose, application specific or *custom* or specific purpose [14].

General purpose processors are those processors that are commonly found in commercial computers. Their controller receives as input an instruction from the

algorithm encoded in a particular machine language the processor can recognize, decodes and prepares them, also generic, datapath to transform the data contained or indicated by the instruction (Figure 1).

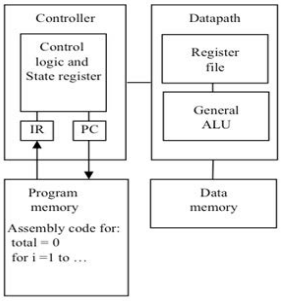


Figure 1: General-purpose processor

Specific application processors are designed for a particular application segment and their architectures are optimized to efficiently carry out typical operations in the implementation in case. Examples of these cases can be audio or video processors, where it is common to use transforms; or processors for scientific applications, where it is often to use numbers in floating point notation and where it is required to maintain accuracy between calculations [15]. The biggest difference with the previous case is that the ALU is optimized to carry out the transformations required in the specific field of application for which the processor was designed (Figure 2).

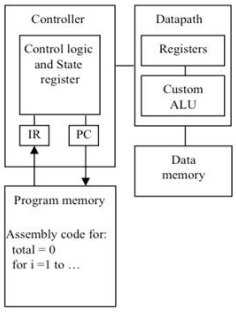


Figure 2: Application-specific processor

Specific purpose processors are designed to explicitly match the requirements of a particular behavior. Their controller executes a single algorithm and their ALU executes only the transformations present in the target behavior [16]. As such, they are optimal in the sense that there is no general-purpose or application-specific processor that can overcome them with respect to the particular behavior defined (Figure 3).

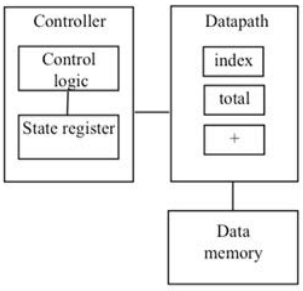


Figure 3: Application-specific processor

3 Cross-compilers

A compiler is a program that transforms a code written in a source programming language into another functionally equivalent code written in object, binary, or machine language. The primary reason for the existence of compilers is to allow human programmers to write programs using high level languages whose primitives are focused on abstract concepts of algorithmic kind, such as assignments, decisions, loops, and functions, but when running on a real machine, it can make use of an optimized version for it, which is written using as primitives the basic operations of the machine, which are actually quite modest: assignment, addition, conditional and unconditional jump.

Since a high level instruction usually translates in several low level instructions, the compiler must use a technique called *parsing*, by which the instruction is broken into logical components. Modern compilers not only translate the code, but can also account for errors in it and make improvements. The tasks typically done by compilers are: lexical analysis, preprocessing, parsing, semantic analysis, code generation, and code optimization [17].

If a compiler is a program that translates code written in a high-level programming language into a low-level or machine equivalent program, a cross-compiler performs a similar task, but has as target another high-level programming language or, as it is in this case, a different-domain language. Figure 4 shows the steps that are commonly carried out by a compiler during its job [18]. The tasks of a cross-compiler usually end with the translation task or, in some cases, with the optimization task; the code generation step is not required, since no low-level code is implied.

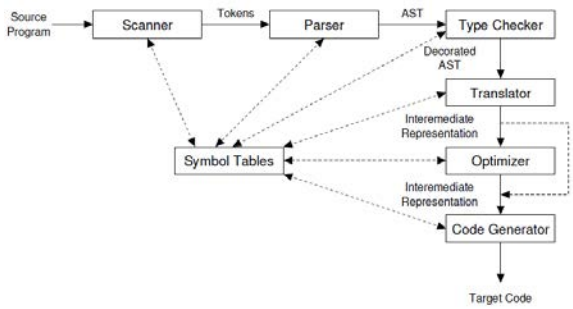


Figure 4: Usual steps carried out by a compiler

The activities of a compiler can be divided into two large stages:

- Analysis of the source code to be compiled
- Synthesis of a target functionally-equivalent code

In more detail, the main steps during a compilation process are:

- *Scanning:* It consists of reading the input text (usually present in a file) and grouping individual characters into tokens such as identifiers, integers, reserved words, or delimiters. The scanning task builds a referential structure, known as AST for *Abstract Syntax Tree*, where the different elements that are identified in the source code are put [19]. The successive steps will enrich the information associated with the elements present in the AST. Since Python employs tab characters as delimiters, PyHDL also stores information about the relative level of each sentence during this step.
- *Parsing:* In this step, the structures created in the previous scanning step related to the organization of tokes are analyzed and distilled to assign meaning and hierarchy

dependences to them. Although during parsing is common to verify the code for correct syntax, PyHDL assumes that the code is syntactically correct. Nonetheless, some simple checking is performed for instance to detect improper tabbing or the use of reserved words as identifiers.

- *Type checking:* It is the checking of the semantics of each AST node. That is, it verifies that the construct the node represents is legal and meaningful, which is mainly associated with determining that all identifiers involved are declared, that their types are correct for the expression being evaluated, among other things. Since data in hardware are stored in predefined registers, PyHDL also checks if the expression implies a change of type for a given variable.
- *Translation:* In this step the abstract representation contained as nodes in the AST is converted into explicit code according to the rules that govern the target language. PyHDL uses for this step a template-driven approach, i.e., the target VHDL code is presented as a generic, static text with some special, replaceable characters that are built on-line as the template is written.
- *Optimization:* PyHDL takes care of possible assignments that can be done in parallel. To do that, it determines the dependencies between the variables involved in assignment instructions immediately followed by other assignment instructions. If no dependencies are detected, those assignments are all put in the same VHDL state.

PyHDL builds the AST identifying the next general syntax construction rules:

```
# identifier := var | fun
#
# header := def fun(parameters)
#
# parameters :=  "" [, var ...]
#
# instruction :=  var = expression   |
#         return expression |
#         if condition:                |
#         else:                |
#         while condition:
#
# expression :=  term |
#         term arithOp term            |
#         fun(expressions)
#
# term :=          constant | var
# arithOp :=       + | - | *
#
# condition :=     term relOp term
# relOp :=         < | <= | == | != | >= | >
```

An *identifier* represents the name of either a variable or a function. Some standard rules apply to the proper construction of an identifier:

- It can only contain letters, numbers and the "_" symbol.
- It cannot have spaces in between.
- It cannot start with a number.

Additionally, VHDL imposes other restrictions that by extension a Python identifier must fulfill:

- It cannot start with the "_" symbol.

- It cannot end with the "_" symbol.
- It cannot have more than one successive "_" symbol.

A *function definition* has the structure: *def fun(parameters)*

Where *def* is the keyword indicating it is a function definition, *fun* is the identifier for that function, and *(parameters)* is an optional list of input variables.

An *instruction* can be any of these statements: assignment, if, else, while, or return. An *assignment* has the form: *var = expression*

Where *var* is a variable identifier and *expression* is an arithmetic expression that produces a number, and that can be a term, an operation between two terms, or a function calling. A *term* is either a variable or a constant. An *operation* is an addition, a subtraction, or a multiplication. A *return* has the form: *return expression*

Where *return* is the keyword and *expression* is as before. An *if* and a *while* have the form:

if condition:
while condition:

Where *if* and *while* are the keywords and *condition* is a logical expression of the form *term relOp term*, in which *relOp* is a relational operator that indicates the kind of comparison to be done (less than, less than of equal, equal, different, greater than or equal, and greater than).

4 PyHDL

PyHDL is a cross-compiler tool conceived to efficiently describe designs of custom processors. Its purposes are: to allow the user writing his/her designs in standard Python code; to translate such a code into HDL code, automatically managing all the low-level aspects such as clocks, state machines, conversion functions, and concurrency; to generate testbenches to simulate such descriptions; and to generate constraint files to ease implementation of the designs in reconfigurable hardware.

PyHDL intentionally hides the particularities of hardware, such as clocks, gates, LUTs, flip-flops, signals, bits, and so on, which, although essential for the professional of computation, are not easy to get for people coming from other disciplines. In this way, the user is invited to think in a high-level domain with respect to his/her solution and let PyHDL take care of the low-level details concerning implementation on physical devices.

PyHDL is a high-level design tool. With it, the designer no longer needs to think in terms of hardware but of the abstract structures of his/her problem. But, why using PyHDL if there are other high-level design tools? The answer is that, in general, most of these tools preserve an explicit link with hardware.

MyHDL [20] is one popular environment at this respect. It is a Python package that allows description in full detail of hardware systems. However, its flavor is still too hardware-based, where concepts such as clock, flip-flop, or edge must be understood. Tools such as Matlab© has, as a part of its set of tools, some interaction with hardware. Simulink, one of such tools, allows designing parametric filters and, for implementation in reconfigurable hardware, permits exportation in VHDL format [21].

A complement is a tool that automates specific aspects of digital design. For instance, PyVHDL [22] is a Python module intended to add scan structures to a hardware design in order to make it testable by standard industrial protocols. In the case of SystemC [23], the

user gets a set of classes and macros written in C++. Since it is capable of modeling systems having concurrent processes, it makes part of some EDA tools. Nonetheless, it still has the unpleasant verbose characteristic.

Many high-level synthesis tools use a similar approach. Chisel [24] builds on the Scala language and is described as a set of special class definitions, predefined objects, and usage conventions within Scala, which, again, are thought of in terms of hardware. PyCPU [25] provides a set of Python modules to generate a general-purpose processor and to program it using a limited subset of the Python language.

In order to show how PyHDL works, let us consider a didactic example, the design of a custom processor that performs the multiplying operation through successive additions. The Python code is shown in Figure 5.

```
def multiplier(num1, num2):
    accum = 0
    count = 0
    while count<num2:
        accum += num1
        count += 1
    return accum
```

Figure 5: Python code multiplier example

In order to generate a custom processor with the desired behavior, it is necessary to associate high-level instructions with low-level machine states. The source code has six instructions, which could be translated to a six-state machine. However, PyHDL does a better job by optimizing the instruction to state mapping. This is done determining what instructions can be executed in parallel. Table 1 presents these concepts applied to the multiplier example.

Table 1: High-level instructions to low-level states mapping

State	Line number	Code
	1	# PyHDL: multiplier example.
	2	
	3	def multiplier(num1, num2):
	4	# set initial values
idle5	5	accum = 0
	6	count = 0
	7	# accum=n1+n1+...+n1; n2 times
while8	8	while count<num2:
assign9	9	accum += num1
	10	count += 1
ret11	11	return accum

The code the tool generates is a VHDL-based state machine. The interface with the module is described as an *entity* (Figure 6). PyHDL automatically adds three signals: start, ready, and result. *Start* is a signal that allows the machine proceeding with its calculation. *Ready* is a signal the machine asserts to indicate it has finished and the result obtained is present in signal *Result*.

```
entity multiplier is
    generic(N : integer := 4);
    port(start : in std_logic;
         clock : in std_logic;
         reset : in std_logic;
         num1 : in signed(N-1 downto 0);
         num2 : in signed(N-1 downto 0);
         result : out signed(N-1 downto 0);
         ready : out std_logic);
end multiplier;
```

Figure 6: VHDL entity multiplier example

PyHDL detects local variables and constants and generates proper *signals* and *constants* for them (see Figure 7). PyHDL also determines the necessary structure for the state machine and proceeds to assemble the state type and internal signals to manage the state sequencing.

```
architecture arch of multiplier is
    type state is (idle5, while8, assign9, ret11);
    signal currentState, nextState : state;
    signal count : signed(N-1 downto 0);
    signal accum : signed(N-1 downto 0);
    constant S0 : signed(N-1 downto 0) := to_signed(0, N);
    constant S1 : signed(N-1 downto 0) := to_signed(1, N);
```

Figure 7: VHDL signals multiplier example

```
-- datapath
process(clock, currentState)
begin
    if (clock'event and clock = '1') then
       if (currentState = idle5) then
           accum <= S0;
           count <= S0;
       elsif (currentState = assign9) then
           accum <= accum + num1;
           count <= count + S1;
       end if;
    end if;
end process;
```

Figure 8: VHDL datapath multiplier example

In order to manage assignments, PyHDL builds a custom datapath taking into account the local variables and constants (see Figure 8). Note the parallelization of instructions fruit of the optimization process performed by PyHDL. The logic the algorithm must follow is provided by the sequencing of instruction in charge of the controller. PyHDL generates the next-state logic based on the kinds of instructions and the conditions present in the source code (see Figure 9). Observe how PyHDL detects what signals should go in the sensitivity list of the process, in this case only *count,* which is the only one that is both modified inside the loop and used in a condition.

```
-- next-state logic
process(currentState, start, count)
begin
    case currentState is
    when idle5 =>
        if (start = '1') then
            nextState <= while8;
        else
            nextState <= idle5;
        end if;
    when while8 =>
        if (count < num2) then
            nextState <= assign9;
        else
            nextState <= ret11;
        end if;
    when assign9 =>
            nextState <= while8;
    when ret11 =>
            nextState <= idle5;
    end case;
end process;
```

Figure 9: VHDL next-state logic multiplier example

The multiplier example is just a proof of concept. More elaborated designs require multiple iterations and block nesting. In order to test the tool capabilities for these

alternatives, three more challenging specifications were presented to two groups of people including students, teachers, and engineers having different levels of expertise. The specifications, in incremental order of complexity were: a 3-number comparator, an integer squared root, and a universal synchronous/asynchronous receiver/transmitter (USART). Figure 10 shows a solution given by a student for the squared root problem.

```
def sqrt(num):
    root = 0
    root2 = 0
    while root2<num:
        root = root+1
        root2 = root*root
    if root2 == num:
        return root
    return root-1
```

Figure 10: Python code squared root example

Figure 11 shows the simulation of that solution after PyHDL automatically generated all the files necessary for that action and invoked a professional simulation tool [26]. The purpose here was not only to test whether the tool translated code into a correct-behavior equivalent, but also to determine the impact of its use on aspects such as hardware resources required for implementation and time for design.

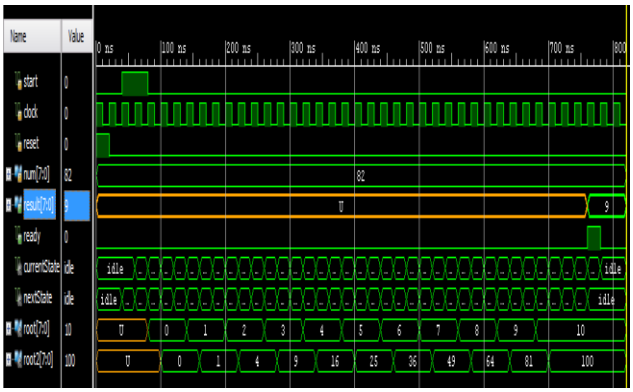


Figure 11: Squared root simulation using PyHDL output files

Figure 12 presents some statistics about the number of *flip-flops* or memory elements extracted by the implementation tool from the files provided by PyHDL for designers with different levels of expertise. As expected, the greater the user knowledge, the lesser the resources required. However, the notable aspect to consider here is the relative comparison between designers with the same level of expertise with respect to such a feature. In all cases, the users that employed PyHDL outperformed those that used VHDL directly for the same task.

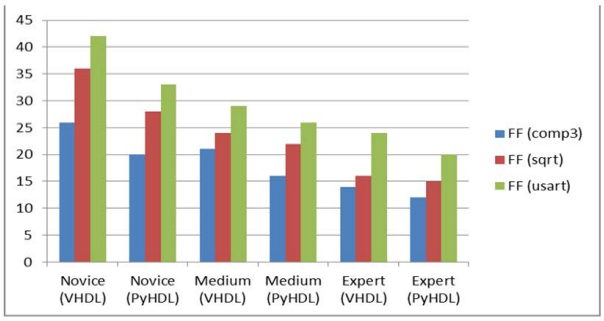


Figure 12: Flip-flops for different specifications

Figure 13 offers another statistics in relation with the time required to obtain a functional prototype. The difference here is more notable than in the case of resources allocated, which is a very interesting result, since the main purpose of the tool is to optimize the time designers require to produce a valid model. On average, the use of PyHDL reduced in around a 50% that time with respect to using conventional VHDL.

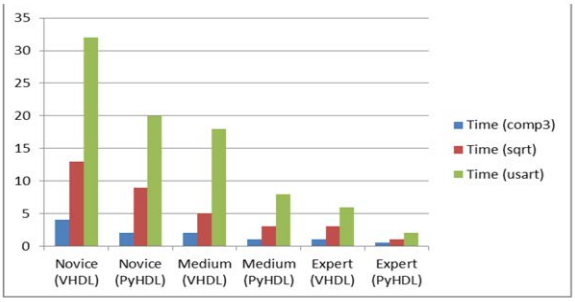


Figure 13: Design hours for different specifications

5 Conclusions

Modern designs in hardware require considering very complex systems. Reconfigurable hardware devices permit a rapid prototyping and exploration of possibilities for such complex digital systems designs. In order to take advantage of the resources available on a reconfigurable hardware device, a powerful description language is mandatory, and languages such as VHDL are expressive enough to meet the requirements.

However, hardware description languages have a very inefficient syntax that does not match modern standards and a very explicit hardware-oriented approach, both factors that become important obstacles for many users to successfully implement their designs. High-level synthesis tools offer a more abstract approach for designing, but most of them are still too focused to hardware constructions, which reduces their target audience to hardare specialists.

PyHDL is a tool that covers such a gap, giving users the opportunity to describe his/her designs in a modern, more friendly language such as Python, and automating generation of the files required by simulation and synthesis tools. PyHDL properly manages Python files containing any arbitrary number of functions, interrelated through any arbitrary hierarchical relationship.

Or course, it is possible to add more functionality. Concurrency may be further optimized by deciding when a condition can be safely put in a transition instead of in a state, reducing the number of states required. Also, data types may be extended to accept fixed-point numbers, taking into account that a general implementation of floating-point numbers is not recommended for reconfigurable hardware devices because of the prohibited amount of resources it requires in absence of built-in floating-point blocks.

The tool can be enhanced by adding libraries of predefined VHDL and/or Python modules, for operations such as division or exponentiation, or others required for particular tasks.

References

[1] Barr, M. (2013). Embedded Systems Glossary. Online [Jan. 2015].
[2] Laplante, P. (2011). Real-time systems design and analysis: Tools for the practitioner. USA: Wiley.

[3] Wolf, M. (2012). Computers as components: Principles of embedded computing system design. London: Morgan Kaufmann.
[4] Kamal, R. (2013). Microcontrollers: Architecture, programming, interfacing and system design. New York: Pearson.
[5] Chua, C., Leong, K. & Lim, C. (2010). Rapid prototyping: Principles and applications. USA: World Scientific Pub.
[6] Mermet, J. (1993). Fundamentals and standards in hardware description languages. USA: Springer Verlag.
[7] Navabi, Z. (2010). Digital system test and testable design: Using HDL models and architectures. Boston: Springer.
[8] Parra, J. (2014). PyHDL: Un cros-compilador de Python estándar a lenguajes de descripción de hardware. Technical Report. Universidad Antonio José Camacho.
[9] Python_Org. (2015). www.python.org/about/quotes/.
[10] Modelsim. (2015). HDL simulation tool. www.model.com.
[11] Sass, R. (2010). Embedded systems design with platform FPGAs: Principles and practices. London: Morgan Kaufmann.
[12] Raspberry_Pi. (2013). Raspberry Pi. www.raspberrypi.org/.
[13] Vahid, F. (2001). Embedded system design: A unified hardware/software introduction. USA: Wiley.
[14] Peckol, J. (2007). Embedded systems: A contemporary design tool. USA: Wiley.
[15] Oshana, R. (2012). DSP for embedded and real-time systems. USA: Newnes.
[16] White, E. (2011). Making embedded systems: Design patterns for great software. New York: O'Reilly Media.
[17] Parra, J. (2007). Lenguaje ensamblador: un enfoque integrado. Cali: Sello Editorial Javeriano.
[18] Fischer, C., Cytron, R. & LeBlanc, R. (2009). Crafting a compiler. USA: Addison-Wesley.
[19] Martin, J. (2003). Introduction to languages and the theory of computation. Boston: McGraw-Hill.
[20] Dannoritzer, G. (2015). MyHDL. www.myhdl.org.
[21] HDL_Coder. (2013). Matlab HDLCoder. www.mathworks.com/products/hdl-coder/.
[22] Huss, N. (2009). Automatic IEEE 1500 wrapper insertion. Master's Thesis. Institutionen för datavetenskap, Dept. of Computer and Information Science. Umeå, Sweden.
[23] OSCI. (2013). Open SystemC Initiative. www.systemc.org.
[24] Bachrach, J. (2012.) Chisel 2.2 Tutorial. Technical report. EECS Department, UC Berkeley.
[25] PyCPU. (2013). PyCPU. http://pycpu.wordpress.com.
[26] Xilinx. (2015). Reconfigurable hardware devices. www.xilinx.com.

Detección de actividad en señales de voz utilizando la transformada *wavelet* discreta

Kevin Rodríguez
Andrés Castañeda
Dora Ballesteros
Universidad Militar – Colombia

1 Introducción

Voice Activity Detection (VAD) o detectores de actividad de voz se implementan en diferentes áreas de procesamiento de voz, tales como comunicación móvil [1], transmisión de voz en tiempo real [2] y reducción de ruido en dispositivos de audición [3]. El objetivo común del VAD en aplicaciones de tecnologías de habla es detectar zonas de voz o silencio sobre una señal de entrada, basándose en sus características de señal y ciertas reglas de decisión basadas en análisis estadístico.

La señal de voz se digitaliza utilizando una frecuencia de muestreo y un número de bits por muestra [4]. Esta señal de audio está conformada por zonas de actividad de voz y de silencio, con ruido de fondo. Convencionalmente, la señal de voz se divide en ventanas de tiempo que oscilan entre 10 y 40 ms, y posteriormente se aplica una etapa de pre-procesamiento [5] cuyo objetivo es reducir el ruido de fondo de la señal. Una opción es utilizar la descomposición wavelet en varios niveles con árbol clásico [6], o por paquetes [4]. Después se hace necesario extraer características que permitan decidir si la zona es de silencio o es de actividad, como la energía [4, 7] o la entropía espectral [8]. Por último se aplica una regla de decisión [5] que permite clasificar las zonas de silencio y de actividad, y como salida se obtiene un pulso alto en la zona de actividad y un pulso bajo en la zona de silencio [7]. Para tener una idea más clara sobre el desarrollo de la programación, en esta investigación se tuvieron en cuenta los algoritmos propuestos en la norma VAD AMR -NB (ETSI, 1999) y en AMR-WB VAD [9]. La secuencia de bloque que implementa el VAD AMR –NB se presenta en la Figura 1.

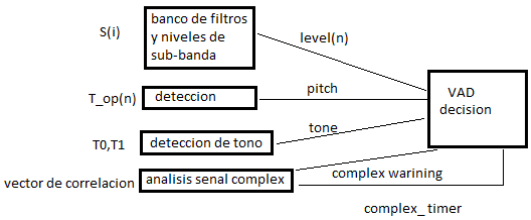


Figura 1: Diagrama de bloques algoritmo VAD [9]

Este sistema presenta cuatro entradas principales conformadas por: *S(i)* que corresponde a las muestras de la trama de entrada, *T_op[n]* que corresponde a un bucle abierto que se calcula por análisis de tono del codificador de voz, *T0* y *T1* que corresponden a la auto-correlación recibidos del análisis de tono y el vector de correlación. La señal *S(i)* se divide en sub-bandas mediante el uso de un banco de filtros tipo IIR, de quinto y tercer orden, cada bloque de filtro divide el sonido de entrada en rangos de diferentes anchos de banda y diezma la frecuencia de muestreo en un factor de dos. Sobre el bloque de detección se realiza la comparación de los valores de retardo de sub-tramas consecutivas. Utiliza una condición sobre el umbral que detecta el tono mediante el uso de una bandera que indica la presencia del mismo. La ganancia del tono

se calcula utilizando valores de auto correlación *T0* y *T1*. Si la ganancia es mayor que un umbral se establece la variable *tone_flag*, con el fin de aumentar la discriminación entre el ruido de fondo y la señal de interés. La detección sobre la caja *Complex signal* calcula una bandera denominada *complex_warning*, que indica la presencia de una señal compleja correlacionada como la música. La bandera *complex_warning* ajustará el proceso de estimación de ruido de fondo, *complex_timer* también se utiliza para iniciar un *hangover* o contador a largo plazo, cuando los valores de correlación compleja filtrados han sido consistentemente grandes durante un tiempo considerable.

El AMR-WB VAD es un algoritmo para encontrar si cada trama de 20 ms contiene las señales que deben ser transmitidas (voz). La salida del módulo de AMR -WB VAD es una bandera booleana denominada *VAD_flag* que indica la presencia de esas señales. El diagrama de bloques simplificado del algoritmo AMR -WB VAD se muestra en la Figura 2. Las tramas de voz *S(i)* es un banco de filtros que permite hacer la descomposición y se implementa la estructura del banco de filtros de 12 bandas. El banco de filtros utilizado en AMR -WB VAD se compone de bloques de filtros IIR de quinto y tercero orden.

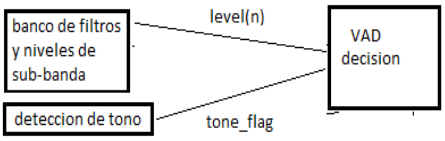


Figura 2: Diagrama de bloques algoritmo AMR-WB VAD [9]

2 Método propuesto para el detector VAD

El método propuesto para la identificación de la actividad de voz utiliza tres módulos: descomposición wavelet, umbralización y optimización.

2.1 Descomposición y umbralización

El objetivo de este apartado consiste en resaltar la zona de actividad en relación con la zona de silencio por medio de la disminución del nivel de ruido de la señal. Este apartado incluye las siguientes etapas:

- Descomposición del audio
- Identificación del nodo con mayor cantidad de ruido
- Estimación del nivel umbralización
- Reconstrucción de los coeficientes umbralizados

Descomposición del audio. Utilizando la base *sym10* se aplica la Transformada Wavelet Discreta (DWT) a la señal para descomponerla. A la salida se obtienen los coeficientes temporo-frecuenciales de tres niveles de descomposición. Se escoge ese número de niveles por las características de ancho de banda de la señal inicial. Se trabaja con los coeficientes de detalle de los tres niveles de descomposición (1,1), (2,1), (3,1) y los coeficientes de aproximación del tercer nivel de descomposición (3,0) (Figura 3).

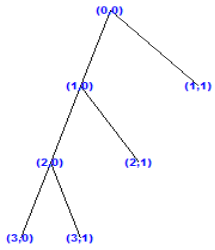


Figura 3: Árbol wavelet para N=3

Identificación del nodo con mayor cantidad de ruido. Con los cuatro nodos de trabajo (3 de detalle y 1 de aproximación) se estima el nodo con mayor nivel. En la Figura 4 se presenta la descomposición DWT de tres niveles para uno de los audios seleccionados en la validación del esquema propuesto.

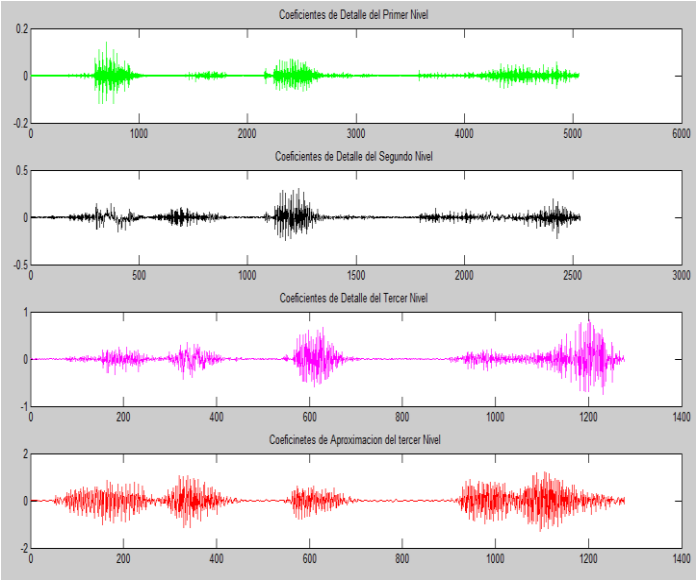


Figura 4: Coeficientes DWT señal de voz para tres niveles de aproximación

Estimación del nivel umbralización. Con el nivel de ruido estimado en cada umbral se procede a aplicar umbralización con la regla de umbral tipo *hard*. Esta regla convierte en cero todos los coeficientes cuyo valor absoluto estén por debajo del umbral y no modifica los coeficientes cuyo valor absoluto estén por encima del umbral. La regla se define como:

$$f(x) = \begin{cases} x & |x| \geq |th| \\ 0 & |x| < |th| \end{cases} \qquad (1)$$

Donde x es el coeficiente de entrada, th es el umbral y $f(x)$ es el coeficiente de salida. Para el caso de los coeficientes de detalle se obtienen los resultados de la Figura 5, y en el caso de los coeficientes de aproximación se obtienen los resultados de la Figura 6.

Reconstrucción de los coeficientes umbralizados. Con los coeficientes umbralizados se reconstruye la señal de voz, obteniendo una señal con menor nivel de ruido que la señal inicial. En la Figura 7 se presenta la señal original (rosado) y la señal umbralizada (negro), la cual tiene menor ruido de fondo.

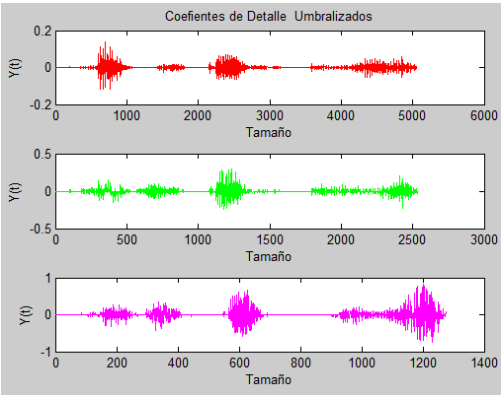


Figura 5. Coeficientes de detalle umbralizados D1-D2-D3 (arriba-abajo)

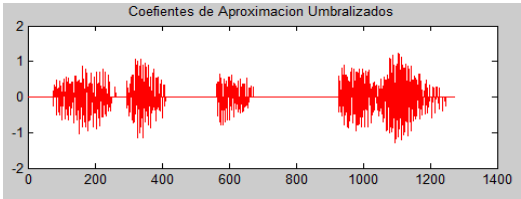


Figura 6: Coeficiente de Aproximación A1 umbralizado

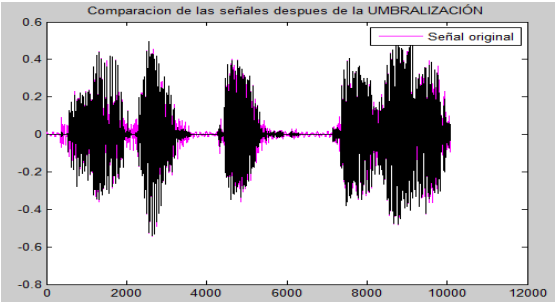


Figura 7: Comparación señal original y reconstruida de los coeficientes

2.2 Decisión VAD: Identificación de Zonas de Silencio y zonas de actividad

El objetivo de este módulo es identificar las zonas de silencio y las zonas de actividad de voz de la señal umbralizada. Para lograrlo se utiliza un umbral que se fija como el nivel de ruido de la señal. Si la magnitud de la amplitud de la señal de voz supera el umbral, se marca como zona de habla, en caso contrario se marca como zona de silencio. En la Figura 8 se presenta con línea roja la identificación de zonas de actividad y de silencio de la señal utilizada como prueba.

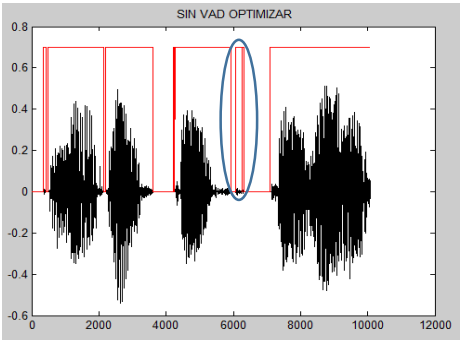


Figura 7: Salida VAD sin optimizar

2.3 Optimización

Debido a que se pueden tener pequeñas ventanas de tiempo con una clasificación de silencio o actividad incorrecta (Ver zona encerrada en la Figura 8), se adiciona esta etapa con el objetivo de minimizar los errores en la identificación. Si la ventana es pequeña y sus ventanas vecinas tienen una clasificación diferente, se actualiza a la clasificación de sus vecinas.

De acuerdo con lo anterior, el procedimiento general consiste en ubicar las ventanas que son muy pequeñas y compararlas con las ventanas vecinas. La definición de una ventana de tiempo pequeña es aquella en la cual no se alcanza a pronunciar un sonido, sea vocal o sílaba. De tal forma que si a una región de silencio le sigue una pequeña ventana de actividad y de nuevo se tiene una región de silencio, la ventana de actividad deberá modificarse y se clasifica como ventana de silencio.

Hay que tener en cuenta que las ventanas pequeñas que se mencionan en la afirmación anterior son en realidad lapsos de tiempo pequeños del orden de los milisegundos, que en realidad no corresponden a información importante en la voz o que tal vez sean componentes diferentes a la voz. Para una señal de voz que ha sido muestreada con una frecuencia (F_s) específica y previamente se ha definido el valor de tiempo de una ventana pequeña (T_w), el total de muestras de la ventana pequeña (L) se define como:

$$L = T_w * F_s \qquad (2)$$

Con el valor de L se compara si el total de muestras consecutivas de una zona clasificada como de actividad es superior a L. Si es inferior la zona se clasifica como de silencio y la salida se convierte en 0. Si es mayor a L la zona se ratifica como de actividad y la salida se mantiene en 1. Aplicando optimización a los resultados obtenidos en la Figura 8 se obtienen los resultados de la Figura 9.

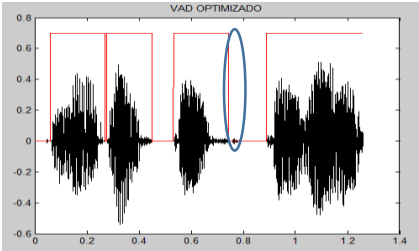


Figura 8: Salida VAD optimizada

Al comparar las Figuras 7 y 8 se aprecia que la parte final de la tercera zona de actividad (que se encierra con un circulo), en la primera era una pequeña zona identificada de actividad, mientras que en la segunda esa pequeña zona se ha identificado como de silencio.

3 Diagrama de flujo del código diseñado

El esquema propuesto se explica por medio del diagrama de flujo de la Figura 10, en el que los pasos para realizar la identificación de la actividad de voz son:

Leer el archivo de voz. Sobre este bloque se carga el archivo de extensión *.wav* que contiene la señal de audio y se asigna a una variable Y.

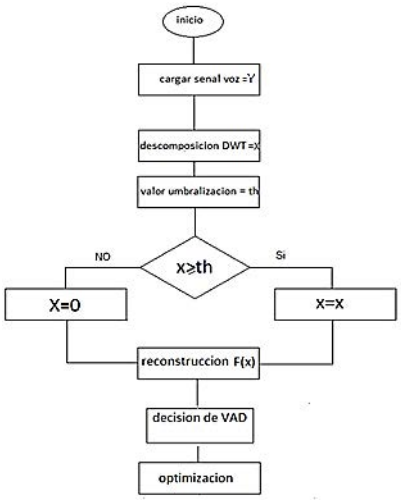


Figura 10: Diagrama de flujo

Descomponer la señal utilizando la DWT. Se realiza la descomposición de la señal que se declaró sobre la variable Y (archivo de audio), y sobre este bloque se determina la longitud para cada coeficiente el de aproximación y los tres de detalle.

Determinar un valor de umbral para cada nodo del árbol (detalles y aproximación).

Aplicar regla de umbralización hard. En este bloque se recorre la longitud de cada coeficiente y se compara la posición del coeficiente con el nivel de umbral, de acuerdo con la regla de umbralización donde no se modifican los coeficientes cuyo valor absoluto estén por encima del umbral y se asigna un cero a todos los coeficientes cuyo valor absoluto estén por debajo del umbral.

Reconstruir los coeficientes wavelet umbralizados. Se reconstruye la señal reuniendo sobre un vector *F(X)* todos los coeficientes umbralizados, implementando la misma base *sym10.*

Identificar las zonas de actividad de voz y zonas de silencio. Teniendo la señal umbralizada, las zonas de voz se fijan con un pulso en alto que se determina a partir del nivel de ruido de la señal umbralizada.

Aplicar optimización en la identificación de zonas. Por último se realiza una comparación que indica si la ventana es pequeña y si las ventanas vecinas son diferentes en longitud, entonces la ventana se actualiza a la clasificación de sus vecinas.

5 Validación del esquema propuesto

Para los resultados que se presentan en esta sección se establece como valor de tiempo de la ventana 30 ms. Si la ventana es menor a ese valor se actualiza con la clasificación de las ventanas vecinas. El esquema se validó con cinco archivos de audio correspondientes al mensaje: *Me gusta estudiar*, presentado por cinco personas diferentes. La frecuencia de muestreo de los audios es 8 KHz y la duración es de aproximadamente un segundo. Se presenta el audio original (azul), la identificación de las posibles zonas de voz sin la optimización y la identificación final de las zonas de actividad de manera optimizada. En los diferentes audios de prueba se simulan los casos en la cual la amplitud es significativamente grande en comparación con la amplitud de la voz, como en el caso del cuarto audio, lo que hace difícil identificar las zonas de silencio.

Como se puede observar del audio original de la Figura 11, la amplitud del ruido en comparación con la amplitud de la voz es muy pequeña, por lo tanto es de esperar que se pueda reconocer y verificar efectivamente las zonas de silencio y de voz. También se puede verificar el efecto que tiene al aplicar la etapa de optimización a la señal VAD, como se observa en la Figura 12, donde hay ventanas en falso que en realidad son componentes de ruido. Al final, en la Figura 13 se corrigen las ventanas falsas de voz.

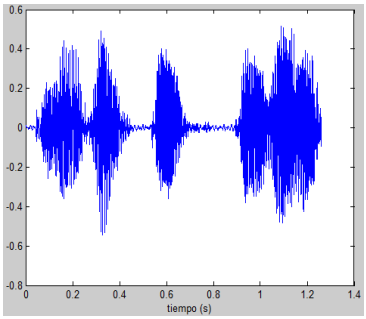


Figura 11: Señal 1 - Audio original

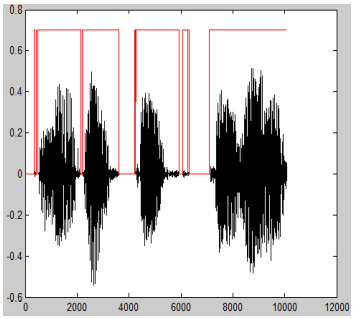


Figura 12: Señal 1 - Decisión de VAD

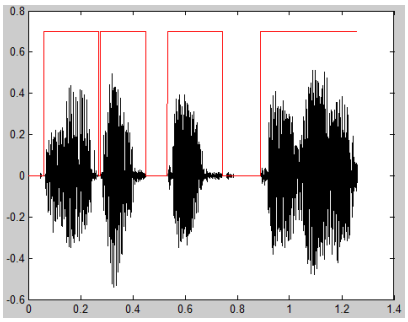


Figura 13: Señal 1 - Decisión de VAD optimizada

En el segundo audio (Figuras 14 a 16) y en el tercer audio (Figuras 17 a 19) se puede verificar la eficiencia de reconocer las zonas de voz para diferentes formas de la señal y diferentes condiciones de ruido, y la importancia de optimizar el programa para un menor margen de error de reconocimiento de voz.

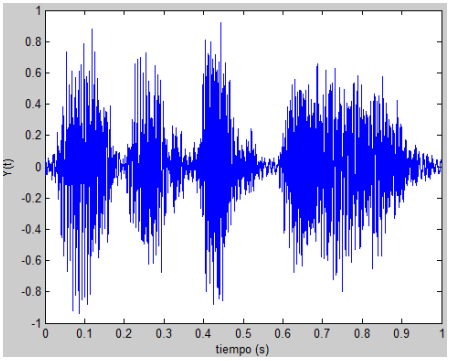


Figura 14: Señal 2 - Audio original

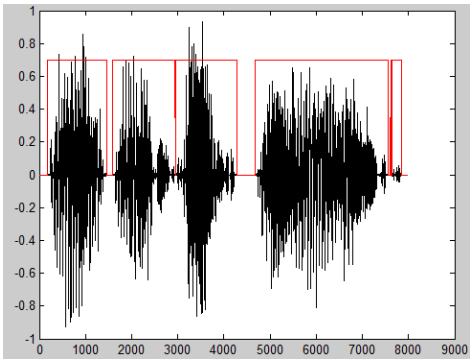


Figura 15: Señal 2 - Decisión de VAD

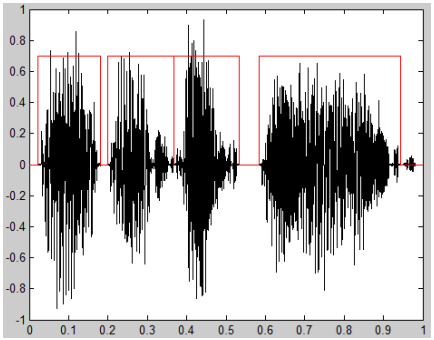


Figura 16. Señal 2 - Decisión de VAD optimizada

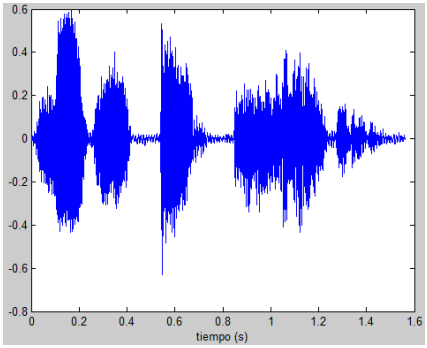


Figura 17: Señal 3 - Audio original

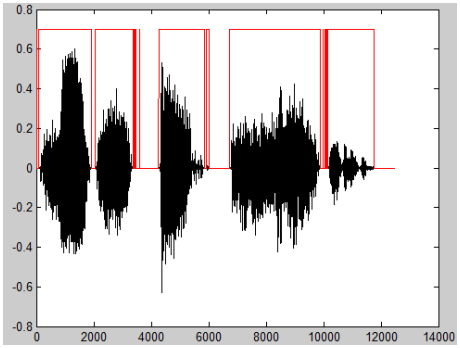


Figura 18: Señal 3 - Decisión de VAD

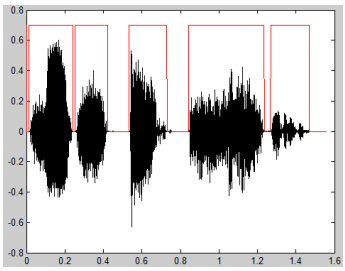


Figura 19: Señal 3 - Decisión de VAD optimizada

En el cuarto audio de la Figura 20 se observa que en la señal original es difícil reconocer las zonas de voz y las zonas de silencio, con lo que se verifica en casos extremos la eficiencia del código.

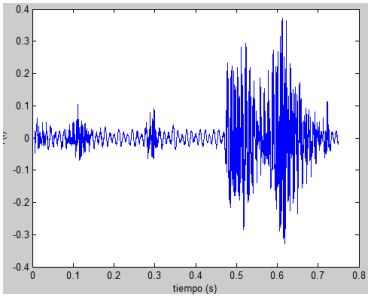


Figura 20: Señal 4 - Audio original

En la Figura 21 se pueden reconocer las zonas de silencio, pero aun existen las ventanas en falso que en realidad son componentes de ruido, de modo que al escuchar específicamente el tamaño de la ventana en falso efectivamente se puede escuchar ruido.

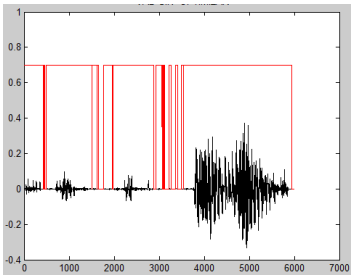


Figura 21: Señal 4 - Decisión de VAD

Luego de la etapa de optimización, en la Figura 22 se puede observar fácilmente que las ventanas en falso desaparecen por completo y quedan las que corresponden a la componente de voz.

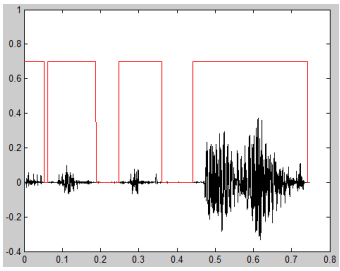


Figura 22: Señal 4 - Decisión de VAD optimizada

De acuerdo con estos resultados se puede observar que el método propuesto es eficiente, debido a que en la primera etapa (decisión de VAD sin optimizar), el algoritmo

filtra la mayor cantidad de ruido de acuerdo con el nivel umbral; sin embargo, sobre la señal resultante se notan pequeños intervalos de tiempo que son relevantes a la voz. Esta condición se mejora al comparar estas pequeñas ventanas con otra de tamaño fijo, obteniendo así el resultado eficiente para la decisión de VAD (decisión de VAD optimizada).

6 Conclusiones

El proceso para construir un algoritmo de reconocimiento de actividad de voz es una herramienta que identifica los tiempos de habla y los tiempos de silencios. El VAD propuesto en este trabajo representa de modo eficaz el funcionamiento de dicho algoritmo, pues para todos los archivos de voz de prueba se analiza que el algoritmo cumple con los requerimientos básicos de analizar los tiempos de voz. Antes del proceso de optimización, el algoritmo reconocía ampliamente componentes diferentes de cero que no eran realmente componentes de voz, de tal forma que el proceso de optimización es un requerimiento importante para el adecuado funcionamiento del algoritmo, el cual ofrece la solución a la necesidad que tiene que ofrecer el VAD como herramienta de procesamiento.

El planteamiento matemático que se emplea para conocer el umbral para los coeficientes de aproximación es un parámetro sencillo y eficiente cuando la amplitud de la componente de voz es significativamente grande, en comparación con la amplitud del ruido. Por lo tanto, sino se cumple la anterior condición es muy probable que se pierda información en las componentes de la voz y el ruido de la señal no se elimina en el proceso de umbralización. Es por esto que no se utiliza el método en el proceso de umbralización de los coeficientes de detalle, porque la amplitud de la voz y la amplitud del ruido son aproximadamente iguales y por tanto no es eficiente aplicarlo a los coeficientes de detalle.

Referencias

[1] Evangelopoulos, G. & Maragos, P. (2006). Multiband modulation energy tracking for noisy speech detection. Transactions on Audio, Speech, and Language Processing 3, pp. 2024 – 2038.

[2] Sangwan, A. & Chiranth, M. C. (2002). VAD techniques for real-time speech transmission on the Internet. Proceedings 5th International Conference on High Speed Networks and Multimedia Communications, pp. 46–50. Jeju Island, Korea.

[3] Nasibov, Z. (2012). Decision fusion of voice activity detectors. Master thesis. School of computing, University of Eastern Finland.

[4] Chen, S. et al. (2005). Robust voice activity detection algorithm based on the perceptual wavelet packet transform. Proceedings International Symposium on Intelligent Signal Processing and Communication Systems, pp. 45–48. The Chinese University, Hong Kong.

[5] Eshaghi, M. & Karami, M. (2010). Voice activity detection based on using wavelet packet. Digital Signal Processing 20(4), pp. 1102-1115.

[6] Hioka, Y. & Hamada, N. (2002). Voice activity detection with array signal processing in the wavelet domain. Proceedings 11th European Signal Processing Conference, pp. 1–4. Toulouse, France.

[7] Chen, S. (2002). A Wavelet-based voice activity detection algorithm in noisy environments. Proceedings 9th International Conference on Electronics, Circuits and Systems, pp. 995–998. Dubrovnik, Croatia.

[8] Shafran, I. & Rose, R. (2003). Robust Speech detection and segmentation for real-time ASR applications. Proceedings International Conference on Acoustics, Speech, and Signal Processing, pp. I-432 - I-435. Hong Kong Exhibition and Convention Centre, Hong Kong.

[9] Chen, S. et al. (2010). Improved voice activity detection algorithm using wavelet and Support vector machine. Computer Speech & Language 24(3), pp. 531-543.

Concentración de redes Wifi en sitios céntricos - Caso de estudio Centro de la ciudad de Medellín Colombia

Roberto Guevara C.
Giovanny Flórez
Corporación Universitaria Remington – Colombia

1 Introducción

El acceso a internet a través de redes inalámbricas WIFI es cada vez más cotidiano, porque su bajo costo sumado a la ventaja de no requerir cables le brinda popularidad, no obstante la vulnerabilidad inherente ante accesos no autorizados. En las grandes ciudades del mundo y principalmente en el centro de éstas se encuentran instalada gran cantidad de estas redes, para usos tales como ofrecer acceso gratis a internet en plazas y centros comerciales; también se usan en pequeños y grandes comercios para la comunicación interna entre sus computadores. En este trabajo se empleará como caso de estudio una zona ubicada en la comuna 10, centro de la ciudad de Medellín Colombia. El estudio se enfoca en caracterizar los protocolos de seguridad empleados para proteger las redes Wifi, determinando el grado de seguridad aplicado; se establecerá la localización GPS de las redes de libre acceso; se implementaran mapas digitales geo-referenciados, y por último se determinará el solapamiento de canales que afectan la velocidad de transferencia.

El proyecto está dividido en dos fases: la primera busca levantar la información para poder cuantificar el problema y ubicar las redes de libre acceso para crear un mapa digital, y la segunda buscará impactar a la comunidad a través del acompañamiento técnico y la capacitación. Este documento se centrará exclusivamente en la primera fase para el diseño de toma de muestras, determinar los protocolos empleados, cuantificar el problema y crear una publicación en línea de mapas geo-referenciados con información de los sitios que provean acceso libre a redes Wifi. Una de las motivaciones con este trabajo, además de cuantificar la problemática de las redes Wifi en sitios céntricos, es dar un primer paso para evitar el acceso no autorizado a las redes Wifi de los comerciantes situados en la zona de estudio, y para estrechar los lazos entre la academia, la industria y la población.

2 Consideraciones generales del desarrollo de la investigación.

El trabajo en áreas como las zonas céntricas de las grandes ciudades, plantea algunos retos que deben ser concretados antes de las pruebas de campo, algunos elementos son de tipo técnico y otros de tipo administrativo y legal.

2.1 Definir lugar geográfico para el caso de estudio

Para el caso de estudio se definió usar una zona relativamente amplia dentro de la ciudad de Medellín. En el estudio se tuvo en cuenta los centros comerciales, bulevares, sitios concurridos, parques, hoteles y universidades. Se estableció la zona comprendida entre las carreras 46 y 52, y entre las calles 50 y 53, como se observa en la Figura 1.

2.2 Zonificar el lugar geográfico para el caso de estudio

Se decidió que es necesario zonificar el área de estudio. En un principio se estudió la posibilidad de zonificarla teniendo en cuenta la densidad de las redes, pero finalmente se determinó usar una zonificación de acuerdo al tamaño por áreas, de tal forma que tuvieran

tamaños similares. Esto dio un resultado de 24 Zonas, delimitadas como se muestra en la Figura 1. El área de estudio tiene alrededor de 170.400 m^2, con un promedio de 7.100 m^2 en cada una de las 24 zonas.

Figura 1: División en zonas del área de estudio

2.3 Asuntos legales y administrativos

En el proceso se estableció que es necesario solicitar permisos legales entre las instituciones y lugares acordados para realizar la toma de pruebas y para acceder a la información requerida, por esto es necesario contar con la participación de una entidad reconocida por los comerciantes implicados, en este caso fue la Corporación Corpocentro. Además, es necesario realizar recorridos exploratorios de toma de muestras con dispositivos más pequeños, tales como teléfonos con software que permitan hacer cálculos iniciales y estimaciones de la cantidad de redes que se pueden encontrar y la densidad de estas, para tomar decisiones sobre cómo hacer el escaneo. Se emplearon tiempos superiores a 10 minutos para *estabilizar* la toma de muestras.

2.4 Instrumentos de medición

Como instrumento de medición se utilizó Wi-Spy DBx, una tarjeta con conexión USB que funciona a modo de antena especializada en captura de trafico Wifi. Esta es una herramienta de hardware que se usa para la identificación de problemas e interferencias en Wifi, también se empleó el software inSSider, que convierte los datos tomados por Wi-Spy en gráficos interactivos.

3 Conceptos previos

Las tecnologías inalámbricas facilitan el acceso a Internet permitiendo la comunicación de equipos inalámbricos. Asimismo, las redes Wifi permiten la movilidad evitando los costos asociados a las conexiones cableadas [8, 10, 12]. Estas redes se protegen mediante claves. Sin embargo, algunas no usan ningún tipo de protocolo de seguridad, esto es válido solamente si se busca que sea de libre acceso [10]. Una red Wifi de libre acceso se considera que no utiliza ningún protocolo de seguridad, es decir, que desde el punto de vista del usuario se puede acceder a ella sin limitaciones [17], y se puede usada eventualmente para acceder a internet en forma libre, como se observa en la Figura 2.

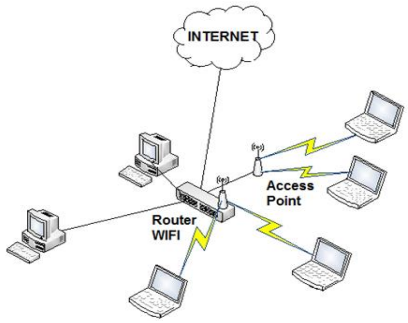


Figura 2: Acceso a internet a través de redes WIFI AP/Router

Un usuario Wifi en general debe autenticarse antes de asociarse con un punto de acceso (AP), o enrutador Wifi. El estándar IEEE 802.11 define la autenticación a nivel de sistema abierto y clave compartida [16].

3.1 Autenticación Wifi

El sistema Wifi emplea básicamente dos tipos de autenticación: de sistema abierto y clave compartida [14]. El primero realmente no es una autenticación, es más bien un mecanismo que consiste en una solicitud realizada por el cliente y que contiene normalmente la dirección MAC del equipo que quiere acceder a la red Wifi. Posteriormente se emite una respuesta del AP/router. En esta forma de autenticación no se requiere una clave de acceso y es la empleada para permitir el acceso a redes Wifi sin que éstas soliciten clave [12]. El segundo requiere que tanto el AP/router como el equipo que quiere acceder a la red intervengan en el proceso de autenticación, porque deben compartir la misma clave [12]. La clave se debe introducir manualmente en ambos y coincidir para poder asociar el nuevo cliente a la red [18]. Existen tres tipos de autenticación de clave compartida WEP, WPA y WPA2, al emplear alguno de estos tipos de autenticación se solicitará una clave para el acceso [13].

Pruebas de campo realizadas confirman la vulnerabilidad del protocolo de seguridad WEP, además varios de autores coinciden en indicar que las medidas que utiliza el protocolo WEP son sustancialmente débiles. Desde décadas los investigadores han mostrado como vulnerar sistemas Wifi, y han concluido que vulnerar las medidas de seguridad WEP, puede ser relativamente fácil utilizando herramientas como Aircrack [1]. La Tabla muestra un comparativo entre los protocolos de autenticación.

Tabla 1: Comparación Tipos de Autenticación

	Tipo de Autenticación			
	Sistema Abierto	**WEP**	**WPA**	**WPA2**
Requiere Clave	No	Si	Si	Si
Grado de Seguridad	Ninguno	Bajo	Media	Alto
Uso	Sistemas públicos, aeropuertos, Parques, Acceso libre a internet	No Recomendado	Redes de Domésticas	Redes domésticas, Redes Oficinas, empleando servidor Radius

3.2 Popularidad de la desencriptación de claves WEP

Una búsqueda en internet puede dar como resultado una amplia cantidad de ayudas que muestran como vulnerar la seguridad de las redes, especialmente si tienen el protocolo WEP. Inclusive en [2] se explica un método paso a paso y lo cataloga de moderadamente fácil. Además, es evidente que no se requiere mayores conocimientos para descifrar estas claves de encriptación con las herramientas *reaver* y *wifiway*. Otro

autor instruye sobre como descifrar una clave WEP en solamente tres minutos [3]. Según el sistema operativo usado existen manuales y tutoriales para Windows y Linux [4], que proponen herramientas como Aircrack para Linux y para sistemas MAC se puede emplear KisMac [5, 9].

Por su parte, la cantidad de información para desencriptar redes WIFI con protocolo de seguridad WAP es alto, de hecho una búsqueda usando la frase *como desencriptar WEP* da como resultado más de 6.000 concurrencias. Si se es más específico, intentando buscar un tutorial son más de 3.800 entre manuales tutoriales y videos. Una búsqueda por videos sobre cómo desencriptar claves Wifi genera aproximadamente 18,600 resultados.

Algunos programas para desencriptar claves son el Aircrack, aunque realmente es una colección de herramientas para la auditoría de redes inalámbricas que incluye las plicaciones Airodump, un programa para la captura de paquetes 802.11, y Aireplay, programa para la inyección de paquetes [1]. El Comsoft Wireless Security Auditor está diseñado para realizar auditorías de seguridad y recuperar claves WPA/WPA2 [6]. Airsnort es una herramienta cuyo propósito es romper la encriptación WEP (obteniendo la contraseña de encriptación) de todas las redes inalámbricas que se encuentren en el radio de alcance del dispositivo inalámbrico que utilice la herramienta [7]. También existen distribuciones de Linux tales como BackTrack, una distribución GNU/Linux en formato LiveCD pensada y diseñada para la auditoría de seguridad [8], y una menos popular, Beini que contiene el modulo Minidwep gtk como herramienta integrada para la desencriptación WEP.

3.3 Mapas Geo-referenciados

Uno de los objetivos de este proyecto es generar un mapa geo-rreferenciado con los puntos Wifi de la zona de estudio, es decir un mapa que muestre la localización de las redes de libre acceso encontradas. Esto se realizó empleando Googlemaps. Esta herramienta permite crear mapas de una región geográfica y referenciarlos con marcación selectiva de un lugar específico con alta precisión, para ser accedido desde la web en cualquier ubicación. Tiene la posibilidad de cargar datos por medio de la dirección o coordenada geográfica de un lugar representando un punto marcado como imagen, que cual es buscado y señalado en el mapa-región. Un grupo de marcaciones sobre distintas coordenadas van formando el mapa georreferenciado con sus diferentes capas de zona

3.4 Solapamiento de canales

El solapamiento de canales consiste en que si dos o más redes Wifi emplean el mismo canal de comunicación generan recíprocamente lentitud en la misma, por ejemplo, en el acceso a internet. El problema radica en que el sistema Wifi solamente ofrece tres canales que no presentan interferencias, entonces, una cuarta red obligadamente tendrá que compartir el mismo canal de comunicación, incurriendo en el solapamiento de canales [10]. La comunicación WiFi se establece en la banda de 2.4Ghz, con 14 canales (ver Tabla 2) y cada canal ocupa 22 MHz de ancho de banda [11]. El estándar IEEE 802.11b/g permite solamente tres canales no interferentes espaciados 3MHz [12]. En la Figura 3 se observa los canales no-interferibles 802.11b/g/n. El sistema Wifi puede transmitir en 13 canales, pero permite que tres de ellos, 1, 6 y 11, no se interfieran entre sí.

Canal 1 = 2,412 Ghz
Canal 6 = 2,437 Ghz
Canal 11= 2,462 Ghz

Tabla 2. Canales IEEE 802.11b/g WIFI adaptado de IEEE

Banda	Frecuencia	Canal
	2412.0 MHz	1
	2417.0 MHz	2
	2422.0 MHz	3
	2427.0 MHz	4
	2432.0 MHz	5
	2437.0 MHz	6
2.4GHz	2442.0 MHz	7
	2447.0 MHz	8
	2452.0 MHz	9
	2457.0 MHz	10
	2462.0 MHz	11
	2467.0 MHz	12
	2472.0 MHz	13
	2484.0 MHz	14

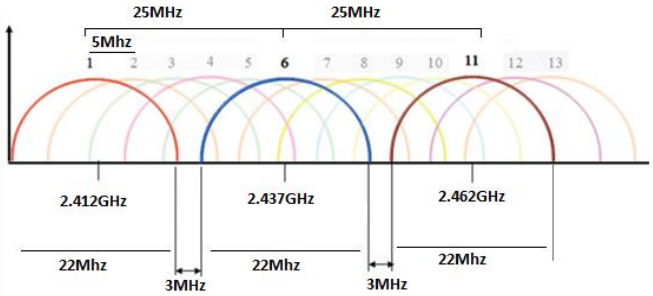


Figura 3: Canales empleados por WIFI

4 Resultados

4.1 Redes de Libre Acceso

Una red Wifi de libre acceso se considera como una red que no usa protocolo de seguridad, y puede ser eventualmente utilizada para acceder a internet de forma libre [12]. Esta investigación determinó que en la zona de estudio existen 103 redes que no cuentan con ningún protocolo de seguridad, frente a 663 que sí usan algún tipo de seguridad, como se observa en la Figura 4. Esto implica que una de cada seis redes no tiene seguridad, bien sea por desconocimiento, instalación inadecuada o porque realmente se tiene la intención de ser un punto de libre acceso.

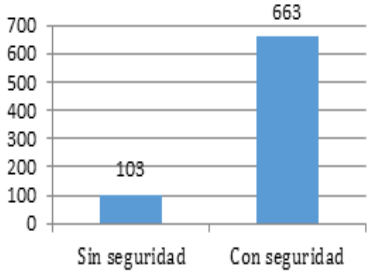


Figura 4: Redes Wifi de libre acceso y redes que requieren autenticación en el área de estudio

Se determinó que en el área de estudio de 170.400 m^2, el 87,5% tiene al menos un punto de libre acceso, solamente el 12,5% (25.560 m^2) no tiene cobertura de alguna red de libre acceso. En la Figura 5 y en la Tabla 3 se muestra la distribución de densidad redes Wifi de libre acceso en cada una de las zonas.

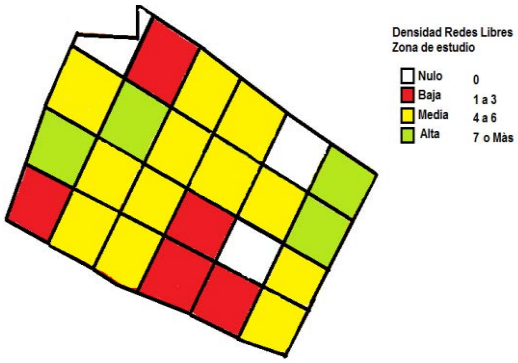


Figura 5: Densidad redes Wifi de libre acceso

Tabla 3: Porcentaje de uso del protocolo WAP en el área de estudio

Densidad WEP	Porcentaje
Nulo	12.5 %
Baja	20.8%
Media	50.0%
Alta	16.7%

4.2 Caracterización de los protocolos de seguridad en la zona de estudio

El estudio indica que pese a los problemas de seguridad muchos usuarios utilizan el protocolo de seguridad WEP para proteger su red Wifi, aunque la mayoría ya han adoptado los protocolos WPA y WPA2 que son más seguro. Se encontró que 663 redes en el área tienen algún tipo de seguridad, y que el protocolo predominante es el WPA2 (41.93%, 278 redes), seguido de WPA con 40.12% (266 redes) y finalmente WEP con 17.95% (119 redes). Aunque WEP se usa en menor porcentaje hay que tener en cuenta que 119 redes corren riesgo de accesos no autorizados. En la Figura 6 se muestra la distribución de uso de protocolos de seguridad.

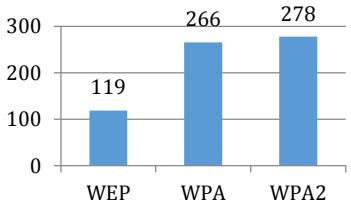


Figura 6: Distribución de los protocolos en las redes con algún tipo de seguridad

El estudio determinó que existen 103 redes de libre acceso que no cuentan con protocolo, frente a 663 que sí usan algún tipo de seguridad. Esto implica que una de cada seis no tiene seguridad, por lo que se consideran redes de libre acceso. Además, se determinó que el 87,5% del área de estudio tiene al menos un punto de acceso libre. En la Tabla 7 se muestra la comparación entre redes con seguridad y sin seguridad.

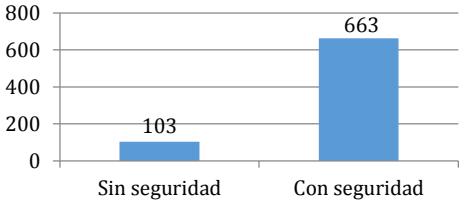


Figura 7: Redes sin seguridad frente a redes con seguridad

También se encontró que en algunas zonas, una de cada tres redes Wifi tiene protocolo WAP, mientras en otras zonas no se tienen redes con este protocolo. La Figura 8 ***Figura***muestra la distribución de densidad de este protocolo en cada una de las zonas

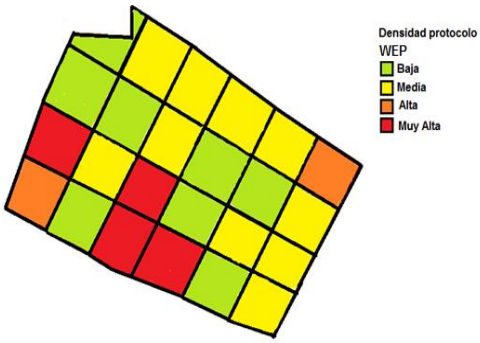


Figura 8: Densidad del uso de protocolo de seguridad WEP

El porcentaje de uso del protocolo WEP en las 24 zonas del estudio se muestra en la Tabla 4. Lo ideal sería que no se empleara el protocolo WEP, o que se tuviera una densidad baja, sin embargo, el estudio muestra que el 66.66% de las zonas tiene una densidad de uso de este protocolo media, alta y muy alta.

Tabla 4: Porcentaje de uso del protocolo WEP en el área de estudio

Densidad WEP	Porcentaje
Baja	33.33%
Media	41.66%
Alta	8.33%
Muy Alta	16.66%

De los resultados obtenidos existen un total de 119 redes que pueden se accedidas en forma fraudulenta, porque usan un protocolo de seguridad débil WEP, y 103 que permite entrar sin restricción alguna. Para la segunda fase del proyecto (acompañamiento técnico y capacitación) se trabajará en las zonas 3, 4, 7 y 9, que se muestran en rojo en la Figura 8. Finalmente, la Tabla 5 muestra el resumen de los resultados de las mediciones.

Tabla 5: Resumen de resultados de mediciones.

Sin seguridad	Con seguridad			Total redes WIFI
	WEP	WPA	WPA2	
103	119	266	278	766

4.3 Ubicación de las redes Geo-referenciados de libre acceso

Como aporte a la comunidad y a los visitantes a la ciudad de Medellín, se entrega finalmente un mapa geo-referenciado en línea donde se puede encontrar la localización de redes Wifi de libre acceso en el centro de la ciudad. En la Figura 9 se muestra el mapa producto de la investigación en el área de estudio.

4.4 Solapamiento de canales

El estudio realizado determinó que las bajas velocidades de transferencia al usar redes Wifi pueden ser causadas por varios factores, pero el más relevante es el solapamiento de los canales empleados. En otras palabras, si dos o más redes Wifi emplean el mismo canal de comunicación, se causaran recíprocamente lentitud. El problema es que el sistema Wifi solamente ofrece 3 canales que no presentan

interferencias, por lo que una cuarta red forzosamente tendrá que compartir el mismo canal de comunicación, incurriendo en solapamiento. Los resultados se muestran en la Figura 10.

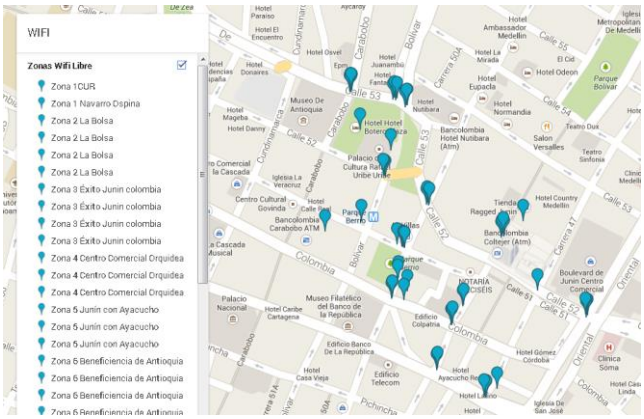


Figura 9: Mapa georreferenciado producto de la investigación

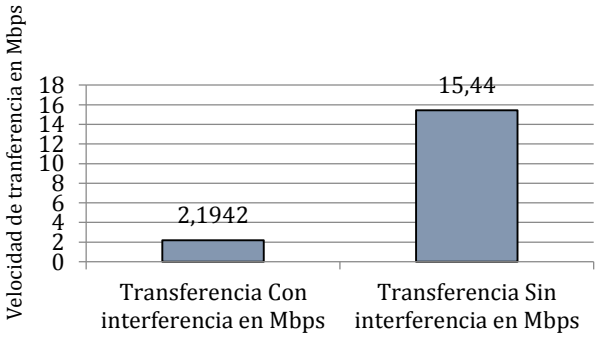


Figura 10: Comparación de envíos WIFI con y sin interferencia de otras redes

5 Conclusiones

La información requerida para vulnerar redes Wifi es abundante y puede ser encontrada fácilmente.

De los resultados obtenidos para el caso de estudio existe un total de 119 redes que pueden se accedidas en forma fraudulenta, debido a que usan un protocolo de seguridad débil WEP, y 103 que permite entrar sin restricción alguna. El estudio también determinó que se puede localizar al menos una red libre en el 87,5% del área.

El Área total del caso de estudio tiene 170.000 m² con 24 zonas de 7.000 m² cada una, es decir, ya que se encontró un total de 766 redes, se puede calcular que cada espacio físico tiene en promedio 31,9 redes, lo que implica un alto solapamiento si se tiene en cuenta que un *router* cubre un área de entre 25 m² y 100 m² en interiores y 300m² o más en exteriores.

En cuanto al solapamiento de canales y apoyados en las pruebas experimentales se puede concluir que la velocidad de trasferencia en una red inalámbrica con interferencia puede disminuir su velocidad de transferencia en 5 veces, si se compara con una libre de interferencia. Esto cobra importancia si tenemos en cuenta cada vez son más las personas y empresas que instalan redes Wifi sin hacer estudios previos. Esto en zonas céntricas con una gran cantidad de redes representa bajas velocidades.

Referencias

[1] Aircrack. Online [Feb. 2015].
[2] Genet. Online [Feb. 2015].
[3] Trucoswindows. Online [Feb. 2015].
[4] Gatewireless. Online [Dec. 2014].
[5] Baccas, P. et al. (2008). OS X exploits and defense. USA: Syngress.
[6] Soft112 (2014). Elcomsoft wireless security auditor.
[7] Backtrack-linux. Online [Feb. 2015].
[8] Blasco, L. (2012). Como descifrar claves Wifi con un Mac. Online [Sept. 2014].
[9] Guevara, R. & Serna, M.E. (2013). Una propuesta de solución al problema de la interferencia entre redes WIFI por solapamiento de canales. Ciencia e Ingeniería Neogranadina 23(2), pp. 7-16.
[10] Jin, A. et al. (2002). Analysis of spectrum channel assignment for IEEE 802.11b Wireless Lan. Proceedings 5th International Symposium on Wireless Personal Multimedia Communications, pp. 1073-1077. Sheraton Waikiki, USA.
[11] Cisco (2006). Fundamentos de redes inalámbricas. Mexico: Prentice Hall.
[12] Ross, J. (2008). The book of Wireless. USA: Starch Press.
[13] Elsevier News (2011). Wifi security cracked. Network Security 2011(1), pp. 2-20.
[14] Upton, N. (2014). Wifi: Amenity or burden? Specialty Coffee Retailer 21(1), pp. 40-44.
[15] Stallings, W. (2004). Fundamentos de seguridad en redes: Aplicaciones y estándares. México: Pearson Educación.
[16] Castro, R. (2005). Avanzando en la seguridad de las redes WIFI. Boletín de RedIRIS 73, pp. 23-33.
[17] Lopez, M. et al. (2009). Algoritmos de protocolos de seguridad en redes de computadoras inalámbricas y el estudio paramétrico de su implementación. Revista Espectro Tecnológico 4, pp. 1-4.

Implementación de un sistema de conteo de estudiantes utilizando redes de sensores inalámbricos

Auris Parody Z.[1]
Jorge Garcia G.[2]
Jorge Chinchilla A.[3]
Jorge Ortiz [4]
[1, 2, 3] *Universidad Popular del Cesar – Colombia*
[4] *Universidad Nacional de Colombia – Colombia*

1 Introducción

Una red de sensores inalámbricos es un conjunto de dispositivos espaciados llamados nodos o motas, cuyo propósito es monitorear condiciones físicas o ambientales de forma rápida, flexible y autónoma para uso de múltiples propósitos [1]. En un principio, sus aplicaciones fueron militares, pero actualmente son ampliamente utilizadas en aplicaciones civiles, monitoreo y protección de propiedades públicas, monitoreo ambiental, control y automatización de procesos industriales, aplicaciones de salud y seguridad, entre otras [1, 2].

En la actualidad, las Redes de Sensores Inalámbricos (REDSI) están siendo utilizadas para desarrollar sistemas de conteo automático en tiempo real [3, 4]. Esto se debe a que no requieren una infraestructura física previa instalada para su funcionamiento, pueden emplear fuentes de energía recargable o solar [4], utilizarse en entornos *indoor* u *outdoor*, requieren menos recursos de procesamiento de la información y presentan bajo consumo de energía.

A nivel internacional se han desarrollado sistemas de conteo utilizando REDSI. Por ejemplo, el implementado en las escaleras de un edificio [5] con el propósito de detectar la dirección del movimiento y el número de transeúntes en un ancho de puerta de 90 cm. En [3] se plantea un sistema de conteo utilizando un tapete formado por sensores piezoeléctricos, que registran las personas que entran al contacto con ellos. Los valores de medición obtenidos por el tapete se transmiten a través de una red inalámbrica de sensores a un computador central que realiza el análisis y el conteo. Está solución fue diseñada para determinar el número de peatones en los cruces de las calles y permite la superposición de varias *FLEXICOUNT*, en el caso que se requiera cubrir intersecciones más grandes. El sistema de conteo en tiempo real implementado en las montañas de Korea con redes de sensores inalámbricas [6] utiliza dos nodos con sensores *photo beam* y un algoritmo de conteo, y permite determinar el número de personas que entran y salen a un área específica de un parque natural. Los nodos sensores se comunican entre ellos y el *gateway* utilizando tecnología inalámbrica.

En el presente trabajo se describe el desarrollo de un sistema de conteo de estudiantes en tiempo real, utilizando una REDSI implementada con sensores tipo pulsador. El objetivo es solucionar el problema de los *estudiantes fantasmas*, que afectan al país desde el año 2010. Con esta red es posible estimar el número de estudiantes presentes en un aula de clases y registrar esta información en una base de datos en internet, para que posteriormente sea consultada por un auditor. El problema de los estudiantes fantasmas se remonta a 2004, cuando el gobierno nacional, con el objetivo de ampliar la cobertura de la educación básica y secundaria, empezó a girar importantes recursos a los departamentos y municipios del país. El objetivo era construir o ampliar la

cobertura en los colegios públicos, y pagarle a los colegios privados para que recibieran un mayor número de estudiantes. Es decir, el gobierno gira recursos por cada estudiante dependiendo de la tipología a la que pertenezca el municipio, por ejemplo, Valledupar se encuentra ubicado en la tipología 5, lo que significa que los estudiantes de básica primaria y secundaria localizados en el área urbana reciben aportes anuales entre 1.141.000 y 1.461.000 pesos respectivamente. De igual manera, los ubicados en el área rural reciben aportes entre 1.256.000 y 1.658.000 pesos, cifras vigentes para el año 2012 [7, 8].

Uno de los casos más nombrados se presentó en Buenaventura, donde según la ministra de educación el 36% de la matrícula de primaria y bachillerato era falsa. Es decir, que de los 111.600 estudiantes que aparecían en los registros, 40.600 no existían. Esto significó que el municipio recibiera 54.500 millones de pesos más [8]. En 2011 el Ministerio de Educación Nacional (MEN), realizó una auditoría a 15 secretarías de educación a nivel departamental y descubrieron 97 mil estudiantes fantasmas. Cifra que representa 132 mil millones de pesos, que se podrían haber invertido en la creación de nuevas escuelas o cupos para niños que aún no gozan del derecho a la educación. Las zonas del país con más estudiantes fantasma son Buenaventura, Turbo y Riohacha [8]. En la actualidad, el gobierno realiza auditorías a las secretarías de educación del país como medio para controlar estos problemas. Aunque ha resultado una herramienta eficiente no es eficaz, porque el estado debe invertir sumas millonarias para poder desarrollarlas.

En Medellín se desarrolló un Sistema Automatizado de Auditoría de Matrícula (SAAM), cuyo fin es verificar que los estudiantes que aparecen matriculados en las instituciones oficiales realmente existan o estén en las aulas, y así evitar los casos de estudiantes fantasmas. Esto lo hace el sistema verificando en línea la lista de estudiantes presentes en clase. La aplicación permite generar indicadores de cada institución educativa y le reporta al MEN cuando uno estudiante no existe [9]. Esta aplicación no tiene ningún costo y se pretende probar en cinco instituciones del municipio de Envigado y en las secretarías de educación de otros municipios [9].

Este trabajo está organizado de la siguiente manera: en la siguiente sección se presenta la metodología utilizada para la evaluación de las tecnologías de sensores; en la sección 2 se presenta un marco teórico referente a los aspectos básicos de las redes de sensores inalámbricos; en la sección 3 se describe la implementación de la red de sensores para el conteo de estudiantes; en la 4 se describe la metodología de desarrollo del tapete sensor; en la sección 5 se presentan las pruebas y resultados; y en la 6 las conclusiones del artículo.

2 Red de Sensores Inalámbricos

Una red de sensores inalámbricos está constituida por nodos, capaces de medir condiciones físicas y ambientales a través de diversos sensores. En una REDSI, los nodos pueden enviar y recibir información de cualquier nodo que se encuentre dentro del rango del radio. Cada nodo sensor puede detectar parámetros del medio ambiente para los que esté preparado, y cada nodo es capaz de comunicar la información detectada a otros nodos o a algún otro dispositivo que agregue y procese la información [10]. El número de nodos y el área a cubrir depende de las necesidades de la actividad a monitorear. La posición de cada de uno puede seguir un patrón específico o tener una disposición aleatoria. Es importante una elección cuidadosa de la topología de comunicación que se utilizará para encaminar y transmitir la información, de tal manera que se asegure un alcance al nodo destino en el menor tiempo posible, evitando al máximo la pérdida de paquetes de información [10]. Las REDSIs pueden ser completamente autónomas o combinarse con

otros tipos de redes, incluso pueden conectarse a Internet utilizando puntos de acceso inalámbricos.

2.1 Arquitectura de una red de sensores inalámbricos

Esta red esta formada por sensores, nodos o motas, un *gateway* y una estación base, como se observa en la Figura 1, en la que el nodo sensor es el elemento básico. Cada nodo es responsable del análisis del medio que lo rodea, el procesamiento de la información y la transmisión de la misma a sus vecinos. Un nodo está formado por uno o más sensores, un controlador, memoria, dispositivos de alimentación, dispositivos de comunicación y una fuente de alimentación. El controlador, también llamado procesador o microprocesador, administra las tareas con uno o más sensores. La memoria almacena los datos temporalmente o durante su procesamiento [10] y puede ser de tipo RAM, ROM o EPROM [11]. Los dispositivos de comunicación son un equipo de transmisión y recepción de radio frecuencia. El nodo puede estar equipado con un sensor o un actuador, o con ambos. El sensor es de bajo costo y tiene capacidades limitadas de monitorización, cómputo y comunicación. Mientras que el actuador tiene una alta capacidad de procesamiento, transmisores más potentes y la batería tiene mayor tiempo de vida. La fuente de alimentación es una batería que le proporciona una energía al nodo de más o menos 10 Julios [10, 11]. Los sensores son transductores que recogen todo tipo de información, tal como temperatura, humedad, movimiento, entre otros [12]. Los nodos tienen la función de tomar los datos del sensor, procesar la información y enviarla a un controlador central o Gateway, que se encarga finalmente de interconectar los distintos nodos con una red de datos o servidor central. La estación base es un servidor que almacena los datos transmitidos por los nodos al *gateway*, para que posteriormente sean visualizados por los usuarios.

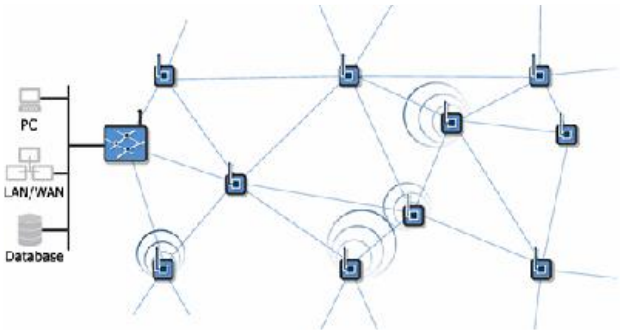


Figura 1: Arquitectura de una red de sensores [1]

3 Diseño e implementación de una red de sensores

El diseño de la red que determina el número de estudiantes que ingresan a un salón de clases es como se muestra en la Figura 2. Los nodos sensores están ubicados en los accesos del salón y el tapete sensor en el piso ocupa 50 cm hacia dentro y hacia fuera en los mismos accesos. Una vez el tapete detecta la presencia de las personas en la entrada, envía una señal a través de un cable al microcontrolador, para que las procese a través del algoritmo configurado y con el fin de estimar el número de estudiantes que ingresan al área de prueba. Posteriormente, la información es enviada por radiofrecuencia a un nodo

gateway, que se encuentra conectado directamente a un servidor, donde se podrá visualizar. La red está conformada por dos nodos sensores ubicados cada uno en un salón de clases, un nodo *gateway* y un servidor. La arquitectura de la red para la transmisión de la información es punto a punto entre el *gateway* y el servidor, y multipunto entre los nodos sensores y el nodo *gateway*.

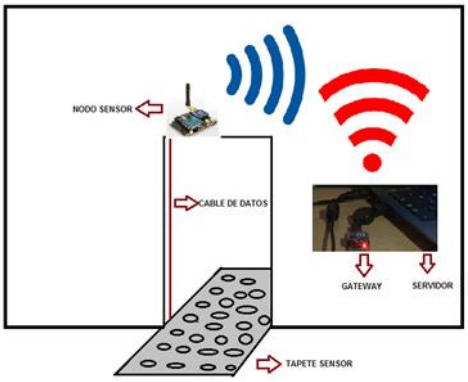


Figura 2: Diseño de la red de sensores

3.1 Implementación de los nodos sensores (nodos finales)

La arquitectura de conexión de los nodos sensores en el acceso del salón se muestra en la Figura 3. El tapete sensor entrega las salidas al microcontrolador PIC 16F877a, que las procesa y entrega al módulo de radio XBEE S1 configurado en nodo final**;** éste tiene una configuración punto-multipunto IEEE 802.14 y se encarga de enviar la información al nodo *gateway*. La batería que alimenta los nodos sensores es una Hi-tronic PS1272 de 12 v.

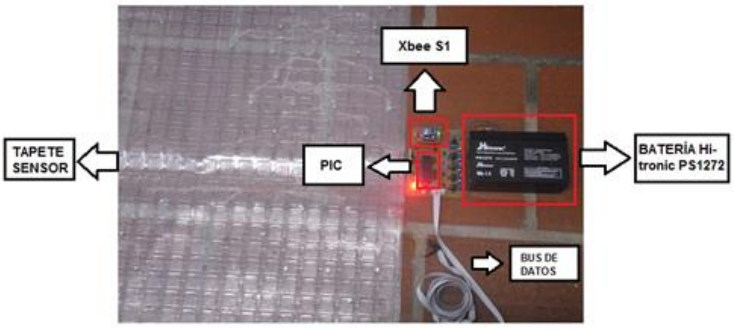


Figura 3. Nodo sensor

PIC 16F877a. Se implementó en este proyecto porque posee la cantidad de interrupciones que se necesitan (5 en total), 4 para el Tapete Sensor y 1 para la detección de nivel bajo de la batería, además, permite la comunicación UART (comunicación serial), dispone de varios temporizadores (3) y la cantidad de pines suficiente para usar una pantalla de visualización LCD en las fases de prueba.

Módulos Xbee S1. Dado que se debe implementar un tipo de red punto-multipunto, se optó por este módulo por su facilidad al momento de la configuración en cualquiera de sus modalidades o modos de funcionamiento, ya sea como Coordinador, Router o Nodo final. Los Xbee S1 son módulos con los que se puede diseñar aplicaciones de comunicación inalámbrica de bajo consumo de energía, como las requeridas en este proyecto. Los módulos Xbee S1 implementan un protocolo propietario de Digi, que permite la configuración de redes punto-multipunto de forma rápida y fácil a través del comando AT.

Batería Hi-tronic PS 1272. La batería es el elemento más sensible al diseño de la red porque es la que marca la autonomía de la misma. Para el desarrollo de este proyecto se utilizó la batería Hi-tronic PS1272. La cual suministra 12 Voltios y 7.2 Amperios, energía suficiente para alimentar los nodos.

Tapete Sensor. En esta etapa se eligió como estructura para montar los sensores de contacto tipo pulsador TacT Switch un tapete plástico transparente de 100x70 cm (Ver Figura 3). Los sensores de contacto fueron distribuidos de forma asimétrica con el fin de ocupar eficientemente cada sección del área que conforma el tapete. Para establecer la conexión entre los sensores de contacto tipo pulsador TacT Switch se seleccionó un hilo de cobre para embobinar de 0,5 mm de diámetro, con una longitud aproximada de 2,5 m por tapete. Para establecer la conexión entre el tapete y el módulo de trasmisión se seleccionó un bus de datos de 3 m de longitud, que fue soldado a las salidas del tapete como entradas del módulo de transmisión.

3.2 Implementación del nodo *gateway* (nodo coordinador)

El nodo *gateway* es un XBEE S1 configurado en modo coordinador, que va conectado al servidor a través de un cable mini USB- USB (Ver Figura 4). A diferencia de los nodos sensores, el *gateway* comunica al servidor la información captada y procesada por los nodos sensores de la red.

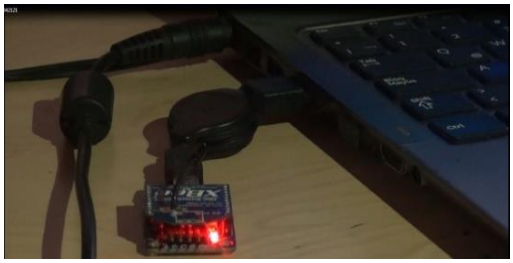


Figura 4: Nodo *gateway*

3.3 Configuración de los nodos finales y el nodo coordinador

Una red de datos RF consistente de un Coordinador y uno o varios dispositivos terminales, que forman lo que se llama una PAN (Personal Area Network). Cada dispositivo en una PAN tiene un identificador llamado ID (ATID), el cual debe ser el mismo para todos los módulos de la PAN. Para configurar esta red, todos los módulos deben tener el mismo canal (ATCH) y la misma PAN (ATID). El módulo coordinador se configura como ATCE = 1 (ATCE1), mientras que los demás, que serán llamados dispositivos terminales, se programan como ATCE = 0 (ATCE0). En la Figura 5 se muestra la configuración en el progama X-CTU del XBEE S1 como nodo coordinador.

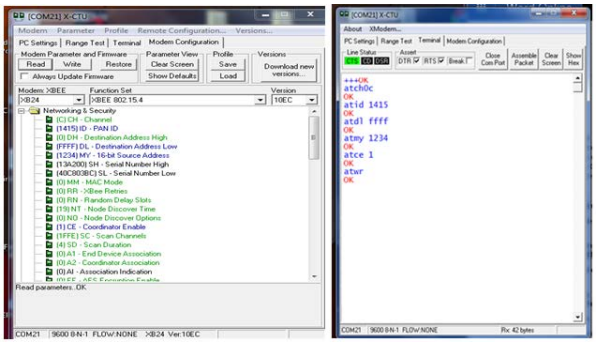


Figura 5: Configuración del nodo coordinador

Para configurar los nodos sensores como dispositivos finales, se debe ingresar el comando CE = 0 (ATCE0). En la Figura 6 se muestran los comandos utilizados para la programación de los XBEE S1 como nodos finales.

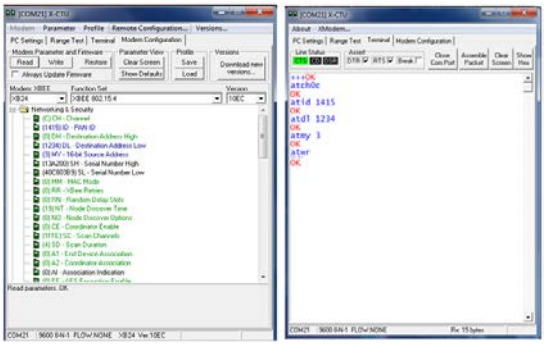


Figura 6: Configuración de los nodos finales

3.4 El servidor

El servidor es un Portátil VAIO con sistema operativo de 64 bits, memoria de 6 GB y procesador AMD A10 5745M APU DE 2.10 GHz. En él se visualiza la información entregada por el nodo *gateway* en un software desarrollado en Visual Basic. La comunicación entre el nodo y el servidor se realiza a través del puerto USB. En la aplicación se administra el funcionamiento de los nodos sensores a través del nodo coordinador, que es el encargado de encenderlos y apagarlos por medio de una transmisión XBEE. Una vez puesta en marcha la red, el *gateway* adquiere la información almacenada en los nodos sensores (cantidad de estudiantes que ingresaron al área de prueba), para finalmente guardarla en una base de datos en internet. El programa permite además algunas funciones administrativas con respecto a la información obtenida, por ejemplo, crear colegios, agregar salones, visualizar el número de estudiantes por salón, entre otras.

3.5 Implementación del Nodo Sensor

El nodo sensor consta de un tapete plástico (100 x 70 cm) con 124 pulsadores, un microcontrolador 16F877A, un Zigbee serie 1 y un servidor. Los pulsadores utilizados fueron Tact Switc, dispositivos simples que se activan o desactivan si se encuentran en contacto con un objeto, por lo que de esta manera se reconoce la presencia de un objeto en un determinado lugar. Estos sensores tienen una vida útil de hasta 80.000 ciclos y operan en un rango de temperatura de entre -20 °C a 70°C.

Se construyó un tapete de 100 x 70 cm de ancho formado por 124 pulsadores de contacto, distribuidos de forma asimétrica en cuatro áreas (47 x 32 cm). Cada una se considera una tecla, y están enumeradas de 1 a 4 con el fin de determinar la dirección de desplazamiento y el número de personas que ingresan o salen de la zona de prueba (Figura 7).

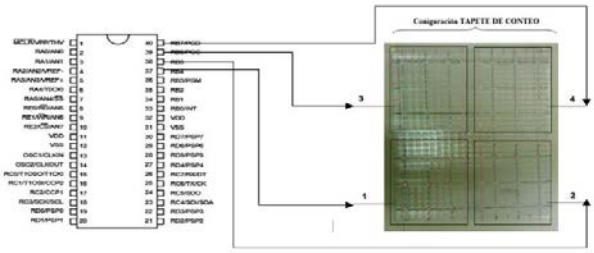


Figura 7: Diagrama de conexión entre puertos de interrupción del PIC y el tapete sensor

Los pulsadores en cada área del tapete están conectados en paralelo formando una tecla, lo que quiere decir que si una persona pulsa más de un sensor de contacto la tecla solamente genera una salida. Para el conteo es necesario que dos teclas estén hacia adentro (50 cm) y dos hacia afuera de la zona de conteo (50cm). Las salidas de las teclas del tapete son enviadas directamente al PIC16F877A a través de un bus de datos. El microcontrolador identifica secuencialmente las teclas que se activaron para determinar la dirección de desplazamiento y el número de personas que ingresaron a la zona de conteo. Finalmente, éste envía a la base de datos del servidor la información a través de una conexión inalámbrica.

3.6 Algoritmo de Conteo

Para el desarrollo de este algoritmo se tuvieron en cuenta tres posibles casos: 1) cuando ingresa o sale un estudiante de la zona de conteo, 2) cuando ingresa o sale más de un estudiante uno atrás del otro, y 3) cuando ingresa o sale más de un estudiante uno al lado del otro.

Primer evento: Una persona ingresa o sale de la zona de conteo. 1) Cuando un estudiante activa una tecla del tapete, esta información es recibida por el PIC que tiene asignado en cada uno de sus puertos de interrupción un valor de 1 a 4. 2) El PIC espera un tiempo aproximado de un segundo la interrupción de una segunda tecla del tapete. Si esta señal llega el microcontrolador, organiza la secuencia recibida con el fin de determinar el sentido de desplazamiento y el número de estudiantes que entraron en contacto con el tapete. Las secuencias para que exista un conteo deben ser:

- Si el primer puerto del PIC que se interrumpe es el 1, para que un estudiante ingrese a la zona de prueba y sea contado, los siguientes puertos que deberían ser interrumpidos deben ser el 3 o el 4, como se muestra en la figura 8.

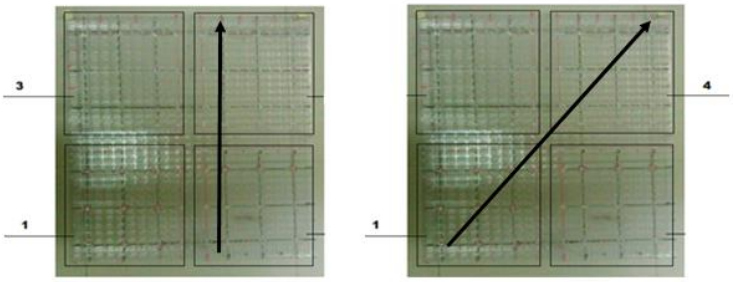


Figura 8: Configuración del tapete sensor, para el ingreso de un estudiante a la zona de conteo

- Si el primer puerto del PIC que se interrumpe es el 3, para que un estudiante salga de la zona de prueba y sea contado, los siguientes puertos que deberían ser interrumpidos deben ser el 1 o el 2, como se muestra en la Figura 9.

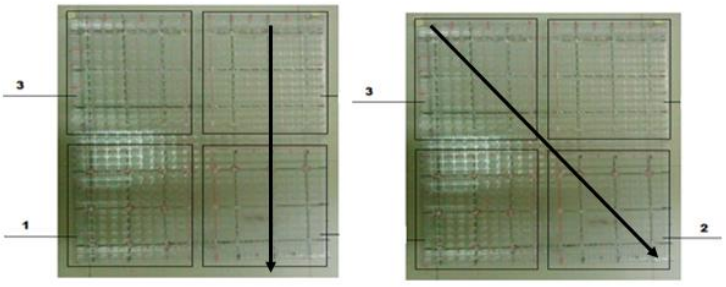


Figura 9: Configuración del tapete sensor, para una persona que sale de la zona de conteo

- Si el primer puerto del PIC que se interrumpe es el 2, para que un estudiante ingrese a la zona de prueba y sea contado los siguientes puertos que deberían ser interrumpidos deben ser el 3 o el 4.
- Si el primer puerto del PIC que se interrumpe es el 4, para que un estudiante salga de la zona de prueba y sea contado, los siguientes puertos que deberían ser interrumpidos deben ser el 1 o el 2.

Segundo caso: Cuando ingresan o salen más de un estudiante uno seguido del otro. Para este caso el algoritmo funciona de la misma forma que para un estudiante que ingresa o sale de la zona de conteo. Es decir, el PIC detecta la interrupción, ejecuta el algoritmo, almacena la información obtenida y se prepara para recibir el próximo registro.

Tercer caso: Cuando ingresan o salen más de una persona una al lado de la otra. 1) Cuando más de un estudiante ingresa a la zona de conteo, se activa más de una tecla del tapete y esta información es recibida por el PIC, que tiene asignado en cada uno de sus puertos de interrupción un valor de 1 a 4. 2) El PIC espera un tiempo aproximado de un segundo la interrupción de dos teclas del tapete. Si estas llegan al microcontrolador se organiza la secuencia recibida, con el fin de determinar el sentido de desplazamiento y el número de estudiantes que entraron en contacto con el tapete. Las secuencias para que exista un conteo deben ser:

- Si los primeros puertos del PIC que se activan son el 1 y el 2 o el 2 y el 1, entonces, para que dos estudiantes que ingresen a la zona de prueba sean contados, los siguientes puertos que deben activarse son el 3 y el 4 o el 4 y el 3, como se muestra en la Figura 10.

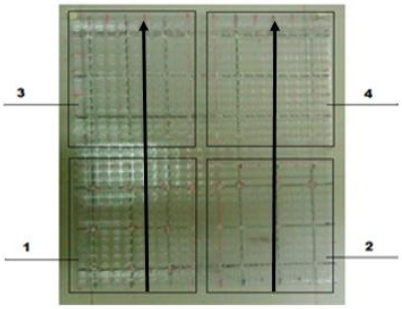


Figura 10: Configuración del tapete sensor para el ingreso de dos estudiantes a la zona de conteo

- Si los primeros puertos del PIC que se activan son el 3 y el 4 o el 4 y el 3 entonces, para que dos estudiantes que salgan de la zona de prueba sean contados, los siguientes puertos que deben activarse son el 1 y el 2 o el 2 y el 1, como se muestra en la Figura 11.

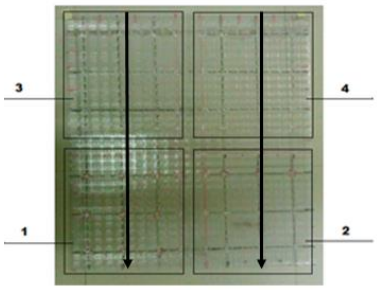


Figura 11: Configuración del tapete sensor para la salida de dos personas a la zona de conteo

4 Pruebas y Resultados

Para evaluar el desempeño del sensor Tact Switc en el prototipo de prueba de conteo expuesto, es necesario que se cumplan dos condiciones: 1) para ser contado, el estudiante debe pisar dos veces el tapete, porque si camina muy lento y la secuencia establecida no es detectada en el tiempo determinado en el PIC, no será contado por el sistema; y 2) que el estudiante que circule por el área de prueba no debe estacionarse sobre el tapete sensor, porque provocará confusión en el sistema sobre la dirección en la que se desplazan las personas, esto no implica que no se realice el conteo.

Las pruebas se realizaron en el Colegio Evangélico Ebenezer y en las instalaciones del Nodo Tecnoparque Valledupar. El tapete se localizó en accesos de 90 cm de ancho, ubicándolo 50 cm hacia dentro y 50 cm hacia afuera de la puerta. En la Figura 12 se muestran los accesos donde se llevaron a cabo los experimentos.

Figura 12: Accesos de prueba colegio Ebenezer y Tecnoparque Nodo Valledupar

Para evaluar el tapete sensor, en el prototipo de conteo se consideró solamente la variable velocidad al caminar y los tres casos planteados en el algoritmo de conteo. Para determinar las velocidades a evaluar durante las pruebas, se efectuó un experimento donde participaron veinte personas, que circularon tres veces (entrando y saliendo) sobre el dispositivo a diferentes velocidades, con el propósito de establecer cuál era la más óptima para obtener los mejores resultados del prototipo de conteo. En la Tabla 1 se registran los resultados alcanzados durante el experimento.

Tabla 1: Pruebas velocidad al caminar

Prototipo de Conteo: Tapete Sensor				
Ubicación del Sensor: Puerta de acceso Tecnoparque				
Hora:8-10 am	Temp: 28-30	Altura: 0,0 m	Población: 20	
Velocidad (m/s)	Personas Entrando	Personas Saliendo	%Precisión Entrando	%Precisión Saliendo
|0,5	13	7	65	35
1	16	20	80	100
2	19	20	95	100

Se puede inferir de la tabla que las velocidades a las cuales el prototipo ofrece una mejor precisión son las de 1 y 2 m/s, ofreciendo la velocidad de 2 m/s el mayor porcentaje de precisión. Por tal razón se puede concluir que la velocidad ideal es aproximadamente 1.5m/s. Una vez los elementos de la red (nodos y *gateway*) están conectados a las baterías y al servidor, se procede encenderlos a través de la interface de Visual Basic para comenzar el conteo de las personas. En el programa, al abrir el formulario de configuración de sensores (administrador de sensores), se establecieron los parámetros necesarios para arrancar el sistema de conteo, por ejemplo, toma de información por días, hora de inicio y fin de captura de datos, salones, entre otros (Figura 13).

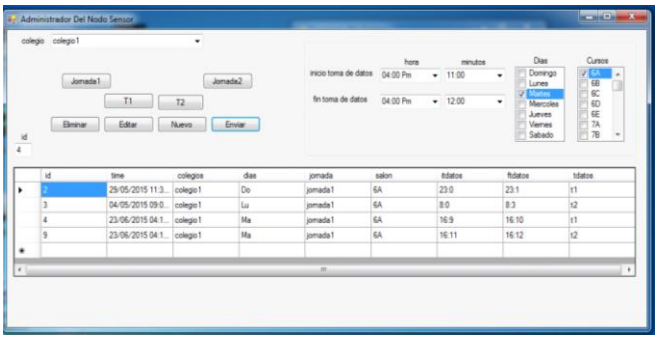


Figura 13: Configuración de arranque del sistema

Una vez el *gateway* envía los parámetros de encendido y apagado y la orden de toma de datos a los nodos finales, estos envían la información obtenida durante el conteo al *gateway* para que finalmente la entregue al servidor y sea almacenada. El sistema fue puesto a prueba durante cuatro sesiones de conteo en tres días, con una duración promedio de 50 minutos con estudiantes y usuarios que ingresaban al área de prueba. Una vez obtenidos los resultados del prototipo junto con los del conteo manual, se realizó un cálculo matemático que determinó la efectividad de desempeño del sistema de conteo.

En la Tabla 2 se registran los resultados de las pruebas realizadas en un salón de clases del Colegio Ebenezer. Los resultados del conteo permiten deducir que la precisión obtenida es aceptable, porque se están considerando las peores variables o condiciones al que puede ser sometido el prototipo, tales como acumulación de personas sobre el tapete, saltar sobre él, estacionamiento de una persona, o que no lo pisen.

Tabla 2: Resultado conteo del tapete sensor en el colegio Ebenezer

Prototipo de Conteo: Tapete Sensor			
Ubicación del Sensor: Acceso Colegio Ebenezer			
Hora: 6:30-8:00 am	Temperatura: ambiente	Altura: 0.0 m	
Pruebas	Real	Dispositivo	% Precisión
1	32	28	87.5
2	32	26	81.25
3	32	25	78.125

En el Nodo Tecnoparque Valledupar se realizaron las pruebas con dos tapetes sensores de igual tamaño. En la Tabla 3 se registra la segunda prueba de conteo, donde el porcentaje de efectividad fue mayor al 85% debido a que la gran mayoría de las personas que ingresaron al recinto lo hacían mirando al piso, provocando una reacción inusual al caminar (aumentaban o disminuían su velocidad de tránsito, alteraban su formar de caminar), dando como resultado para el sistema un mal conteo, porque su forma de caminar no era normal y en otros casos solo pisaban el tapete sensor una vez.

Tabla 3: Resultados de la segunda prueba de conteo realizada en el Nodo TecnoParque

Prueba Tecnoparque				
	V.Real	V.Dispositivo	% Error	% Efectividad
Tapete 1	21	24	14.29	85.71
Tapete 2	18	16	11.11	88.89

En la Tabla 4 se registra la tercera prueba de conteo, que arrojó un porcentaje de efectividad de más o menos 93%, siendo mucho mejor respecto al de la primera y segunda debido a que durante la prueba se les hicieron dos recomendaciones a las personas objeto de experimentación: caminar normalmente y pisar dos veces el tapete sensor. Gracias a

las indicaciones el comportamiento del sistema de conteo mejoró en gran medida, reflejándose en la precisión de los nuevos resultados.

Tabla 4: Resultados de la tercera prueba de conteo realizada en el Nodo TecnoParque

Prueba Tecnoparque				
	V.Real	V.Dispositivo	% Error	% Efectividad
Tapete 1	41	44	7,32	92,68
Tapete 2	38	36	5,26	94,74

En la cuarta prueba de conteo se mantuvieron las mismas condiciones sugeridas a las personas que ingresaban a las instalaciones del recinto Tecnoparque (Ver Tabla 5), arrojando un porcentaje de efectividad de más o menos 92%, superando significativamente la primera prueba.

Tabla 5: Resultados de la tercera prueba de conteo realizada en el Nodo TecnoParque

Prueba Tecnoparque				
	V.Real	V.Dispositivo	% Error	% Efectividad
Tapete 1	41	42	2,44	97,56
Tapete 2	48	54	12,50	87,50

En cuanto a la conectividad de los nodos y la transmisión de la información, los Xbee en el entorno de prueba permitieron una distancia efectiva de transmisión de 15m, existiendo línea de vista entre los módulos. En el evento en que no exista línea de vista entre los nodos finales y el nodo coordinador, la transmisión de los datos se ve afectada, al extremo de ser casi nula. En los casos donde la red de sensores coexista con una red Wifi de alta densidad de usuarios, se debe asignar otro canal de transmisión a los módulos Xbee, de lo contrario se presentará interferencia co-canal entre ellos, ocasionando un menor desempeño del sistema de conteo.

5 Conclusiones

En este trabajo se desarrolló una REDSI para el conteo automático de estudiantes que ingresan o salen a un salón de clases. El sistema permite estimar el número de estudiantes presentes en un aula, en una jornada, hora y fecha establecida. Esto constituye una herramienta útil para auditar las aulas de clases de colegios oficiales, porque la toma de datos puede no estar preestablecida y el auditor puede iniciar la captura de la información en el momento que lo requiera emitiendo la orden desde el servidor.

El elemento principal del sistema de conteo es un tapete sensor, implementado con sensores tipo pulsador y que fue diseñado para accesos de un metro de ancho. El tapete permite estimar la cantidad de estudiantes en un salón de clases y la dirección en la que se mueven. Ofrece un 97% de precisión en el conteo, siempre y cuando las personas transiten sobre él con una velocidad promedio de 1.5 m/s y no alteren su forma natural de caminar al pasar, también es necesario que el tapete sea pisado por lo menos dos veces. El porcentaje de precisión mas bajo durante las pruebas fue de 78%, posiblemente porque no se cumplieron las condiciones explicadas anteriormente.

La transmisión de la información obtenida por los nodos sensores se hizo a través de tecnología inalámbrica, implementando los módulos de transmisión XBEE S1, que se configuraron como nodos finales en los nodos sensores y como nodos coordinador en el *gateway*. El área efectiva de cobertura de la red fue de 15 metros, debido al espesor de las paredes en los espacios de prueba y de la distribución de los entornos *indoor* donde se evaluó el desempeño de la red. Para áreas sin obstáculos los XBEE s1 ofrecen alcances hasta de 30 metros con línea de vista entre el *gateway* y los nodos sensores.

Referencias

[1] Mourelle, A. & Bravo, A. (2011). Detección y seguimiento de personas en entornos indoor empleando redes de sensores con tecnología imote 2. Madrid: Universidad Carlos III.

[2] Akyildiz, I. et al. (2002). Wireless sensor networks: A survey. Computer Networks 52(12), pp. 393–422.

[3] Kutschera, C. et al. (2011). A flexible sensor-mat to automate the process of people counting. Proceedings Fourth International Conference on Advances in Circuits, Electronics and Micro-electronics, pp. 13–16. Nice, France.

[4] Chan, A. & Vasconcelos, N. (2008). Privacy preserving rowd monitoring: Counting people without people models or tracking. Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-7. Anchorage, USA.

[5] Hashimoto, K. et al. (1997). People count system using multi-sensing application. Proceedings International Conference on Solid State Sensors and Actuators, pp. 1291-1294. Chicago, USA.

[6] Son, B. et al. (2007). Implementation of the real-time people counting system using wireless sensor networks. International Journal of Multimedia and Ubiquitous Engineering 2(3), pp. 63–80.

[7] MEN. (2012). Ley 715 de 2001. Ministerio de Educación Nacional.

[8] Revista Semana. (2012). Así se perdió la plata de la educación. Online [Feb. 2014].

[9] David, C. (2014). Presentamos cuatro de las principales innovaciones creadas por jóvenes en apps Medellín. Diario ADN. Mayo 9 de 2014.

[10] Ramirez, L. (2012). Diseño de una arquitectura para redes de sensores con soporte para aplicaciones de detección de eventos. Tesis Doctoral. Universitat Politècnica de València.

[11] Mayteh, F. (2012). Simulación de un algoritmo de enrutamiento para redes de sensores inalámbricas. Tesis de Maestría. Universidad de Colima.

[12] Lopez, J. (2008). Desarrollo de un demostrador para evaluar técnicas Cross-Layer en sistemas de comunicaciones inalámbricos. Tesis de Maestría. Universitat Politècnica de Catalunya.

Cifrado de imágenes en fase usando la transformada fraccionaria de Hartley y operaciones de truncamiento

Juan Vilardy O.
Lenin Nuñez R.
Cesar Torres M.
Universidad Popular del Cesar – Colombia

1 Introducción

El cifrado de imágenes es diferente de la criptología tradicional, porque tiene presente las características espaciales, frecuenciales y de redundancia de la imagen a cifrar [1]. En el proceso se emplean algunas herramientas del procesado de señales, tales como las transformadas de Fourier, Coseno, Seno, Hartley, Hankel, Hadamard, Wavelet, entre otras, con el objetivo de codificar la imagen a cifrar en un ruido aleatorio [1]. También, existen sistemas de cifrado de imágenes que emplean las transformadas mencionadas en versiones fraccionarias. Por ejemplo, en [2, 3] se emplea la transformada fraccionaria de Fourier (fractional Fourier transform, FrFT) y en [4, 5] se utiliza la transformada fraccionaria de Hartley (fractional Hartley transform, FrHT), con el objetivo de incrementar la seguridad del sistema de cifrado, porque dichas transformadas fraccionarias adicionan nuevas claves representadas por los ordenes fraccionarios [1-5]. FrFT [6] y FrHT [7] son generalizaciones de las transformadas de Fourier y de Hartley convencionales, que adicionan un nuevo grado de libertad, dado por el orden fraccionario de las transformadas fraccionarias mencionadas. Además, son empleadas en aplicaciones de análisis y procesamiento de señales e imágenes [6, 7].

El cifrado de imágenes en fase fue inicialmente propuesto en [8], con el objetivo de mejorar la seguridad de los sistemas cifradores que codificaban las imágenes a cifrar en módulo. En otros trabajos se ha empleado la transformada de Hartley [9] y FrHT [10] en los sistemas cifradores de imagen, con el propósito de obtener una imagen cifrada de valores reales e incrementar el nivel de seguridad de dicha imagen cifrada. Las operaciones de truncamiento de módulo y de fase introducen efectos no-lineales sobre los sistemas cifradores que las emplean, permitiendo mejorar su seguridad [11].

En este trabajo se propone la utilización de FrHT, las máscaras aleatorias de fase (*random phase masks*, RPMs) y las operaciones de truncamiento de módulo y fase para el cifrado de imágenes digitales en fase. La imagen a cifrar se codifica en fase porque este esquema de codificación permite una mayor protección de la imagen cifrada en contra de ataques de seguridad [1, 3]. Con respecto al cálculo computacional de FrHT, se introduce una simplificación que permite reducir la carga computacional en el cálculo. Las RPMs contienen imágenes de códigos aleatorios empleadas como llaves de seguridad del sistema de cifrado. Las operaciones de truncamiento de módulo y fase son operaciones matemáticas no-lineales que permiten la obtención de nuevas llaves de seguridad para el sistema de cifrado-descifrado, dichas llaves son dependientes de los valores de la imagen a cifrar y de los códigos aleatorios correspondientes a las RPMs. Las llaves de seguridad que dispone el sistema de cifrado propuesto son: dos órdenes fraccionarios asociados a las FrHTs aplicadas, dos imágenes de códigos aleatorios usadas en las RPMs y dos imágenes de código seudo-aleatorios generadas por las operaciones de truncamiento.

El resto del documento está organizado de la siguiente forma: la sección 2 muestra la definición de FrHT e introduce una simplificación en su cálculo. Las operaciones de

truncamiento de módulo y fase son descritas en la sección 3. Los métodos de cifrado y de descifrado con sus respectivos resultados experimentales son presentados en la sección 4, y finalmente en la sección 5 se describen las conclusiones.

2 La transformada fraccionaria de Hartley

La transformada fraccionaria de Hartley (FrHT) es un operador matemático integral que puede ser definido por medio de la transformada fraccionaria de Fourier (FrFT) [6]. La FrFT de orden fraccionario p, es un operador lineal e integral que transforma una función dada $g(x)$ para obtener una función $G^p(u)$ [6]:

$$G^p(u) = \mathcal{F}^p\{g(x)\} = \int_{-\infty}^{+\infty} g(x)K_p(u,x)\mathrm{d}x \qquad (1)$$

Con:

$$K_p(u,x) = C_p \exp\{i\pi[(u^2+x^2)\cot\alpha - 2ux\csc\alpha]\} \qquad (2)$$

$$C_p = \frac{exp\{i[\pi \mathrm{sgn}(\alpha) - 2\alpha]/4\}}{\sqrt{|\sin\alpha|}}, \alpha = \frac{\pi}{2}p, |\alpha| \leq \pi, |p| \leq 2 \qquad (3)$$

Donde K_p es el kernel fraccionario de Fourier y *sgn* es la función signo. Para $p = 0$ ($\alpha = 0$), la FrFT se corresponde a la identidad transformada. Para $p = 1$ ($\alpha = \pi/2$), FrFT se reduce a la transformada de Fourier directa, y para $p = 2$ ($\alpha = \pi$), es el operador paridad. Para $p = -1$ ($\alpha = -\pi/2$) FrFT se corresponde a la transformada de Fourier inversa. La FrFT inversa corresponde a la FrFT con orden fraccionario $-p$. El operador FrFT es aditivo con respecto al orden fraccionario $\mathcal{F}^{p_1}\mathcal{F}^{p_2} = \mathcal{F}^{p_1+p_2}$. La FrFT tiene un periodo 4 con respecto al orden fraccionario p.

La FrHT de orden fraccionario p para una función $g(x)$ se define como [7]:

$$g_p(u) = \mathcal{H}^p\{g(x)\} = \frac{1+\exp\left\{i\frac{\pi}{2}p\right\}}{2}G^p(u) + \frac{1-\exp\left\{i\frac{\pi}{2}p\right\}}{2}G^p(-u) \qquad (4)$$

Donde $\mathcal{H}^p$ es el operador fraccionario de Hartley, $g_p(u)$ es la FrHT de orden fraccionario p para la función $g(x)$ y $G^p(u)$ es la FrFT de orden fraccionario p para la función $g(x)$. Debido a que la FrHT puede ser definida en términos de FrFT, satisface todas las propiedades matemáticas de la misma [7]. Para $p = 1$ ($\alpha = \pi/2$), la FrHT se reduce a la transformada de Hartley convencional. La FrHT tiene un periodo 2 con respecto al orden fraccionario p.

El cálculo computacional de la FrHT para una determinada función por medio de la ecuación (4) implica la evaluación computacional de dos FrFT. Con el fin de simplificar el cálculo computacional de FrHT se propone la siguiente reducción de la ecuación (4):

$$\mathcal{H}^p\{g(x)\} = \frac{1+\exp\left\{i\frac{\pi}{2}p\right\}}{2}G^p(u) + \frac{1-\exp\left\{i\frac{\pi}{2}p\right\}}{2}G^p(-u)$$

$$\mathcal{H}^p\{g(x)\} = \frac{1+\exp\left\{i\frac{\pi}{2}p\right\}}{2}\mathcal{F}^p\{g(x)\} + \frac{1-\exp\left\{i\frac{\pi}{2}p\right\}}{2}\mathcal{F}^p\{g(-x)\} \qquad (5)$$

$$\mathcal{H}^p\{g(x)\} = \mathcal{F}^p\left\{\frac{1+\exp\left\{i\frac{\pi}{2}p\right\}}{2}g(x) + \frac{1-\exp\left\{i\frac{\pi}{2}p\right\}}{2}g(-x)\right\}$$

Por lo tanto, al emplear la ecuación (5) para el cálculo computacional de la FrHT solamente se necesita la evaluación de una FrFT. De esta manera se consigue que el cálculo computacional de la FrHT sea más rápido y eficiente en comparación con los algoritmos de cálculo numérico de la FrHT empleados en las referencias [4, 5, 10].

3 Operación de truncamiento de módulo y fase

Las operaciones de truncamiento que se describen a continuación son aplicadas a funciones de valores complejos. Sea $f(x) = a(x)\exp\{i2\pi\phi(x)\}$ una función de valores complejos, donde $a(x)$ y $\phi(x)$ representan el módulo y la fase de la función $f(x)$, respectivamente. El módulo $a(x)$ siempre es una función de valores reales positivos y la fase $\phi(x)$ es una función real que puede contener valores positivos o negativos.

El truncamiento de módulo (*TM*) se define como una operación matemática que permite la extracción de la fase de una función de valores complejos, eliminando o truncando el módulo de la función de valores complejos a la cual se le aplica la operación de *TM*. Por lo tanto, al aplicar la operación de *TM* sobre la función $f(x)$, se obtiene:

$$TM\{f(x)\} = TM\{a(x)\exp\{i2\pi\phi(x)\}\} = \phi(x) \qquad (6)$$

El truncamiento de fase (*TF*) permite la extracción del módulo, eliminando o truncando la fase de la función de valores complejos a la cual se le aplica la operación de *TF*. La operación de *TF* sobre la función $f(x)$ es:

$$TF\{f(x)\} = TF\{a(x)\exp\{i2\pi\phi(x)\}\} = a(x) \qquad (7)$$

La *TM* y la *TF* son operaciones matemáticas no lineales que no conservan la energía de las funciones a las cuales son aplicadas. Nótese que la función $f(x)$ de valores complejos puede ser representada de la siguiente forma:

$$f(x) = TF\{f(x)\}\exp\{i2\pi TM\{f(x)\}\} \qquad (8)$$

4 Métodos de cifrado y descifrado

La imagen de valor real con valores en el intervalo de [0, 1] a ser cifrada es representada por medio de la función $g(x)$ (la notación matemática es escrita en una dimensión por simplicidad) y las dos RPMs $r(x)$ y $t_\alpha(u)$ son dadas por:

$$r(x) = \exp\{i2\pi m(x)\}, t(u) = \exp\{i2\pi n(u)\} \qquad (9)$$

Donde $m(x)$ y $n(u)$ son funciones aleatorias, normalizadas, positivas, estadísticamente independientes y uniformemente distribuidas en el intervalo (0, 1). A continuación se describen todos los pasos del método de cifrado. En el primer paso se codifica la imagen $g(x)$ a ser cifrada en fase [3]:

$$g_{ph}(x) = \exp\{i2\pi g(x)\} \qquad (10)$$

En el segundo paso se toma la FrHT de orden fraccionario p_1 del siguiente producto de funciones:

$$h_{p_1}(u) = \mathcal{H}^{p_1}\{g_{ph}(x)r(x)\} = q_{p_1}(u)\exp\{i2\pi\phi_{p_1}(u)\} \qquad (11)$$

La imagen $h_{p_1}(u)$ es una función de valores complejos con módulo $q_{p_1}(u)$ y fase $\phi_{p_1}(u)$. Las características de las funciones módulo $q_{p_1}(u)$ y fase $\phi_{p_1}(u)$ son dependientes de las características de la RPM $r(x)$ y la imagen a cifrar codificada en fase $g_{ph}(x)$.

Para el tercer paso del sistema de cifrado se aplican las operaciones de truncamiento de módulo y fase sobre la imagen $h_{p_1}(u)$:

$$q_{p_1}(u) = TF\{h_{p_1}(u)\}, \phi_{p_1}(u) = TM\{h_{p_1}(u)\} \qquad (12)$$

En el cuarto paso se toma la FrHT de orden fraccionario p_2 del siguiente producto de funciones:

$$s_{p_2}(v) = \mathcal{H}^{p_2}\{q_{p_1}(u)t(u)\} = e(v)\exp\{i2\pi\theta_{p_2}(v)\} \qquad (13)$$

Finalmente, para el quinto paso se aplican nuevamente las operaciones de truncamiento de módulo y fase sobre la imagen $s_{p_2}(v)$:

$$e(v) = TF\{s_{p_2}(v)\}, \theta_{p_2}(v) = TM\{s_{p_2}(v)\} \qquad (14)$$

La imagen cifrada es la función $e(v)$ y las seis llaves de seguridad del método de cifrado son: los órdenes fraccionarios p_1 y p_2 de las FrHTs, las RPMs $r(x)$ y $t(u)$, y las dos imágenes pseudoaleatorias $\phi_{p_1}(u)$ y $\theta_{p_2}(v)$.

Para el proceso de descifrado se parte de la imagen cifrada $e(v)$, luego se realiza el procedimiento inverso del sistema de cifrado empleando los valores correctos de las seis llaves de seguridad con el fin de recuperar la imagen que inicialmente se ha cifrado, es decir, la imagen de valor real $g(x)$. La imagen $g(x)$ se puede recuperar en el proceso de descifrado como la imagen descifrada $d(x)$, usando los valores correctos de las seis llaves de seguridad de la siguiente manera:

$$q_{p_1}(u) = t^*(u)\mathcal{H}^{-p_2}\left\{e(v)\exp\{i2\pi\theta_{p_2}(v)\}\right\} \qquad (15)$$

$$g_{ph}(x) = r^*(x)\mathcal{H}^{-p_1}\left\{q_{p_1}(u)\exp\{i2\pi\phi_{p_1}(u)\}\right\} \qquad (16)$$

$$d(x) = TM\{g_{ph}(x)\} = g(x) \qquad (17)$$

Siguiendo los pasos descritos en esta sección, los resultados experimentales de los métodos de cifrado y descifrado se obtuvieron por medio de la herramienta de cálculo numérico de MATLAB [12]. Todas las imágenes empleadas en las simulaciones numéricas tienen una resolución de 512×512 píxeles ($M = 512$). La imagen a ser cifrada (imagen original) y el código aleatorio $m(x)$ de la RPM $r(x)$ se presentan en la Figura 1(a) y la Figura 1(b), respectivamente. La imagen del código aleatorio $n(u)$ de la RPM $t(u)$ presenta la misma distribución aleatoria de la Figura 1(b), con diferentes valores en los píxeles. Los órdenes fraccionarios usados en las FrHTs de las ecuaciones (11), (13), (15) y (16) fueron seleccionados como: $p_1 = 1.77$ y $p_2 = 0.43$.

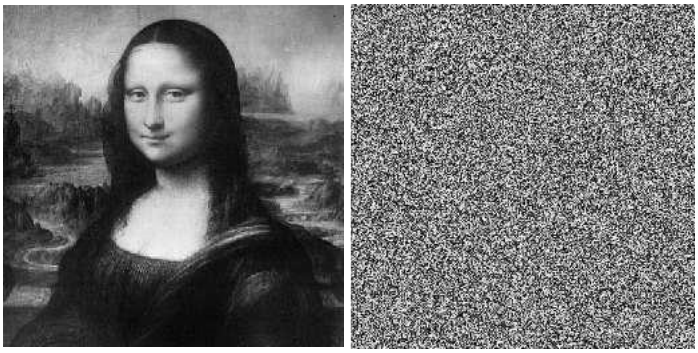
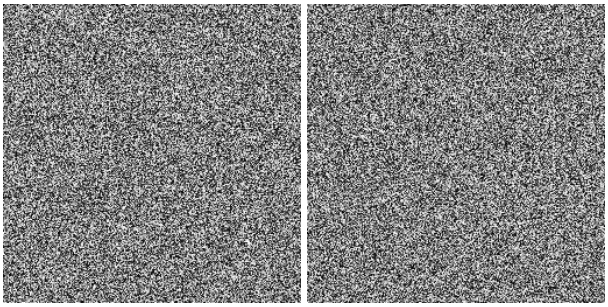


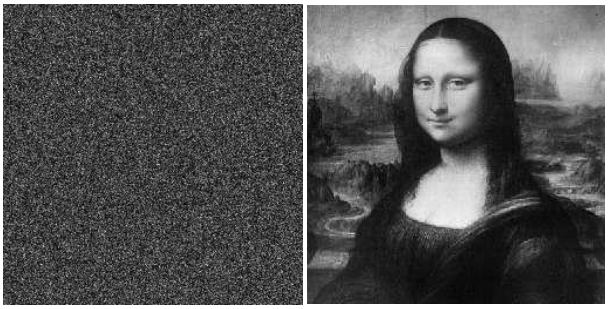
Figura 1: (a) Imagen original $g(x)$ a ser cifrada, y (b) código aleatorio $m(x)$ de la RPM $r(x)$

Las imágenes seudo-aleatorias $\phi_{p_1}(u)$ y $\theta_{p_2}(v)$ obtenidas en el método de cifrado se presentan en la Figura 2(a) y Figura 2(b), respectivamente. Las imágenes $\phi_{p_1}(u)$ y $\theta_{p_2}(v)$ presentan una apariencia ruidosa muy similar al código aleatorio $m(x)$ de la Figura 1(b), pero en realidad dichas imágenes son códigos seudo-aleatorios que dependen de la

imagen original codificada en fase $g_{ph}(x)$, las RPMs $r(x)$ y $t(u)$ y las FrHTs de órdenes fraccionarios p_1 y p_2 calculadas en las ecuaciones (11) y (13), respectivamente. Las imágenes seudo-aleatorias $\phi_{p_1}(u)$ y $\theta_{p_2}(v)$ son dos llaves de seguridad que son diferentes al cambiar la imagen a cifrar, las RPMs $r(x)$ y $t(u)$ y los órdenes fraccionarios p_1 y p_2 de las FrHTs.



(a) $\phi_{p_1}(u)$ (b) $\theta_{p_2}(v)$



(c) $e(v)$ (d) $d(x)$

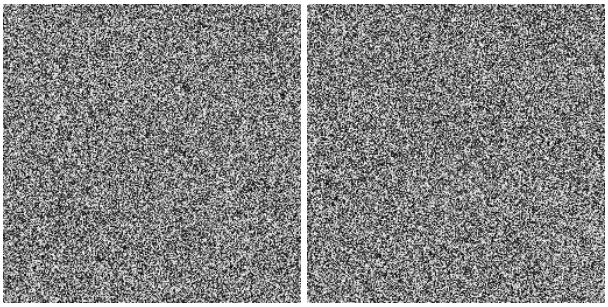
Figura 2: Imágenes seudo-aleatorias: (a) $\phi_{p_1}(u)$, y (b) $\theta_{p_2}(v)$. (c) Imagen cifrada $e(v)$ para los órdenes fraccionarios $p_1 = 1.77$ y $p_2 = 0.43$, y (d) Imagen descifrada $d(x)$ obtenida al usar los valores correctos de las seis llaves de seguridad

La imagen cifrada $e(v)$ para los órdenes fraccionarios $p_1 = 1.77$ y $p_2 = 0.43$ se muestra en la Figura 2(c). La imagen cifrada $e(v)$ es una función de valores reales positivos y un patrón ruidoso que no desvela ninguna información de la imagen original $g(x)$. La Figura 2(d) presenta la imagen descifrada $d(x)$, cuando se emplean todos los valores correctos de las llaves de seguridad en el proceso de descifrado. La imagen descifrada de la Figura 2(d) es una replica de la imagen original $g(x)$ sin ninguna pérdida visible al ojo humano. Con el fin de evaluar cuantitativamente la calidad de la imagen descifrada se emplea el error cuadrático medio (mean square error, MSE), definido de la siguiente manera [12]:

$$\mathrm{MSE} = \frac{1}{M^2}\sum_{x=1}^{M}\sum_{y=1}^{M}[g(x) - d(x)]^2 \qquad (18)$$

Donde $M = 512$. El MSE entre la imagen original $g(x)$ de la Figura 1(a) y la imagen descifrada $d(x)$ de la Figura 2(d) es 7×10^{-15}. Si las llaves de seguridad empleadas en el proceso de descifrado son diferentes a las usadas en el proceso de cifrado, la imagen original $g(x)$ no se recuperará. Las imágenes descifradas son patrones ruidosos que no se corresponden con ninguna información de la imagen original $g(x)$, tal como se muestra en la Figura 3. En esta figura aparecen dos imágenes descifradas ruidosas obtenidas al introducir errores sobre las llaves de seguridad $\theta_{p_2}(v)$ y p_1 en el proceso de descifrado. Si alguna de las restantes llaves de seguridad (las RPMs $r(x)$ y $t(u)$, la imagen

pseudoaleatoria $\phi_{p_1}(u)$ y el orden fraccionario p_1) son introducidas con valores erróneos en el proceso de descifrado, las imágenes descifradas obtenidas son patrones ruidosos muy similares a las imágenes mostradas en la Figura 3. Por lo tanto, todos los valores de las llaves de seguridad empleadas en el proceso de descifrado deben ser iguales a los utilizados en el proceso de cifrado, con el fin de obtener una replica de la imagen original $g(x)$ en la imagen descifrada $d(x)$.



(a) (b)

Figura 3: Imágenes descifradas obtenidas en el sistema de descifrado para la siguiente llave de seguridad errónea (las demás llaves de seguridad presentan sus valores correctos): (a) $\theta_{p_2}(v)$, y (b) $p_1 = 1.775$

La sensibilidad de los órdenes fraccionarios de las FrHTs se evalúa introduciendo errores (de manera individual) y dejando fijas las demás llaves de seguridad. El MSE se utiliza para medir el nivel de protección sobre la imagen cifrada. En esta prueba de desviación de los órdenes fraccionarios de las FrHTs sobre los valores correctos del proceso de descifrado se introduce un error que varía entre -0.9 y 0.9, luego para cada variación se calcula el MSE, los resultados se observan en la Figura 4.

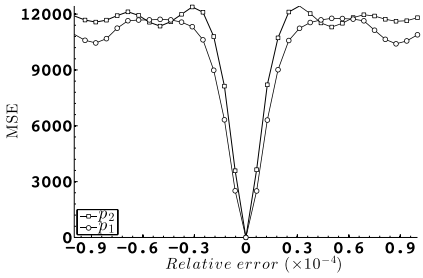


Figura 4: Prueba de desviación de los órdenes fraccionarios de las FrHTs sobre los valores correctos del proceso de descifrado

De experimentos computacionales (Figura 4) se encontró que los órdenes fraccionarios son sensible a una variación de 1×10^{-5}. Por lo tanto, el espacio de clave para los dos ordenes fraccionarios de las FrHTs es de 4×10^{-10}, ya que la FrHT tiene un periodo de 2 con respecto al orden fraccionario.

Las dos RPMs $r(x)$ y $t(u)$ tienen una resolución de 512×512 píxeles ($M = 512$) y cada píxel tiene L posibles valores. El número de intentos requeridos para recuperar ambas RPMs es del orden de L^{2M^2} para $L = 256$ y $M = 512$ y el número de RPMs para probar serían 256^{524288}. Por lo tanto, los ataques de fuerza bruta son intratables al considerar cada posibilidad de las dos RPMs. El espacio de clave con respecto a las dos imágenes seudo-aleatorias $\phi_{p_1}(u)$ y $\theta_{p_2}(v)$ es muy similar al espacio de clave de las dos RPMs. Finalmente, al combinar los espacios de clave de las seis llaves de seguridad del método de cifrado-descifrado, se obtiene un espacio de clave total robusto y grande.

5 Conclusiones

En este trabajo se presenta un nuevo sistema de cifrado de imágenes basado en la codificación en fase, la FrHT, las RPMs y las operaciones de truncamiento de módulo y fase. Se ha introducido una simplificación en el cálculo de la FrHT que permitió que el cálculo computacional de la FrHT fuera más eficiente y rápido. La imagen cifrada es una distribución aleatoria de valores reales positivos y se encuentra muy bien protegida frente a ataques de seguridad. La alta seguridad del método de cifrado radica en la gran sensibilidad en cualquier cambio que se haga sobre las seis llaves de seguridad (el espacio de clave total del método de cifrado-descifrado es robusto y grande) y en el empleo de diferentes operaciones no lineales, tales como la codificación en fase y las operaciones de truncamiento de módulo y fase. Las dos llaves de seguridad dadas por las dos imágenes seudo-aleatorias, son claves dependientes del contenido de la imagen codificada en fase, las RPMs y las FrHTs. Por lo tanto, estas dos llaves de seguridad son únicas para cada imagen que se desee a cifrar. Finalmente, con el fin de recuperar la imagen original con una alta calidad y libre de ruido en la salida del sistema de descifrado, los valores de las llaves de seguridad empleados en el sistema de descifrado deben ser iguales a los valores utilizados en el sistema de cifrado.

Referencias

[1] Millán, M. & Pérez, E. (2011). Optical data encryption. In Cristóbal, G.; Schelkens, P. & Thienpont, H. (Eds.). In Optical and digital image processing: Fundamentals and applications. New York: Editorial John Wiley & Sons Ltd.

[2] Vilardy, J.; Torres, C. & Mattos, L. (2008). Fractional Fourier transform applied to digital images encryption. Proceedings AIP Conference, pp. 1067-1072.

[3] Vilardy, J.; Torres, C. & Mattos, L. (2008). Encriptación en fase aplicado a imágenes digitales a color. Revista Colombiana de Física 40(1), pp. 143-146.

[4] Vilardy, J.; Torres, C. & Jimenez, C. (2013). Double image encryption method using the Arnold transform in the fractional Hartley domain. Proceddings SPIE 8785, pp. 1-5.

[5] Jimenez, C.; Torres, C. & Mattos, L. (2011). Fractional Hartley transform applied to optical image encryption. Journal of Physics: Conference Series 27(1), pp. 1-7.

[6] Ozaktas, H.; Zalevsky, Z. & Kutay, M. (2001). The Fractional Fourier Transform: with Applications in Optics and Signal Processing. New York: Editorial John Wiley & Sons Ltd.

[7] Pei, S. & Ding, J. (2002). Fractional Cosine, Sine, and Hartley transforms. IEEE Transactions on Signal Processing 50(7), pp. 1661-1680.

[8] Towghi, N.; Javidi B. & Luo Z. (1999). Fully phase encrypted image processor. Journal of the Optical Society of America A16(8), pp. 1915-1927.

[9] Chen, L. & Zhao, D. (2006). Optical image encryption with hartley transforms. Optics Letters 31(23), pp. 3438-3440.

[10] Zhao, D.; Li, X. & Chen, L. (2008). Optical image encryption with redefined fractional hartley transform. Optics Communications 281(21), pp. 5326–5329.

[11] Qin W. & Peng X. (2010). Asymmetric cryptosystem based on phase-truncated Fourier transforms. Optics Letters 35(2), pp. 118-120.

[12] Gonzalez, R.; Woods, R. & Eddins, S. (2009). Digital Image Processing Using Matlab. New York: Gatesmark Publishing.

Metodología para el desarrollo de software integrando SOA, BPM y TOGAF

Lourdes De Avila Gutiérrez
Instituto Tecnológico de Soledad Atlántico – Colombia

1 Introducción

Actualmente, la tendencia en las organizaciones es hacia un paradigma orientado a procesos, donde las aplicaciones abarcan la actividad global de la empresa y las herramientas son los BPMS [11]. Pero los modelos resultan insuficientes porque no se integran adecuadamente y se orientan a describir datos y transacciones. El cambio de enfoque en el modo de diseñar aplicaciones e implementar soluciones radica en:

- Explicitar el conocimiento de un proceso de negocio ayudando a documentarlo, definirlo e implementarlo
- Proveer interoperabilidad de las soluciones
- Resolver la dinámica de los problemas en términos declarativos y cubriendo todas las etapas del ciclo de vida del software

Por su parte, la comunidad del software está interesada en proveer sistemas robustos y escalables. Además, debido a que los procesos de negocios se realizan en espacios de información tecnológica compleja, la integración de los Sistemas de Información existentes se convierte en una base importante para la implementación técnica de los procesos de negocio.

Según Weske Mathias [4], la herramienta BPM se basa en la observación de cada producto que la compañía provee al mercado, lo que genera como resultado un número de actividades ejecutadas. Los procesos de negocio son la clave para organizar esas actividades y mejorar el entendimiento de sus interrelaciones. Otra de estas herramientas es SOA, que según Delgado [10] permite diseñar, construir, desplegar e integrar los servicios independientes de los lenguajes en los que estén codificados y de las plataformas en las que se ejecutan. Estos servicios están vinculados entre sí y se definen a través de procesos de negocio formando servicios compuestos que llevan a cabo las funciones empresariales. Los servicios pueden compartirse y reutilizarse en varios procesos de negocio. El resultado es un entorno altamente adaptable, con menores costos para el desarrollo de aplicaciones, mejoras en la integración y despliegue rápido.

El desarrollo de software orientado a servicios y a gestión de procesos está vinculado con el manejo de flujo de trabajo, según lo demuestra Bazán [17] en su tesis. Un sistema de manejo de *workflow* permite definir, crear y manejar la ejecución de flujos de trabajo a través del uso de software, puede correr en uno o más motores y es capaz de interpretar la definición del proceso, interactuar con los participantes del *workflow,* y donde sea requerido invocar el uso de herramientas y aplicaciones de tecnologías de la información. Lo que buscan alcanzar estos flujos de trabajo es la representación de las estructuras de los procesos a través de modelos, y el control de estos basándose en los modelos SOA y BPM. La aproximación a través del manejo de modelos facilita un mayor grado de flexibilidad, debido a que estos pueden ser adaptados a los requerimientos, y las modificaciones resultantes se pueden utilizar inmediatamente para representar los procesos de negocio.

De acuerdo con Robledo [9], para obtener verdaderos beneficios de los enfoques SOA y BPM se deben reformular roles y responsabilidades en la definición, especificación e implementación de los proyectos dentro de la organización. El equipo técnico debe reorientarse a resolver el trabajo en forma no-monolítica, identificando componentes verticales. Aparece la figura de arquitecto, que ensambla cada una de las piezas e interactúa con el analista de negocios, otro actor preponderante en este escenario que debe participar activamente en todas las etapas de la concreción del proyecto.

2 Propuesta integrando SOA y BPM utilizando *framework* TOGAF

Esta propuesta metodológica se define como un conjunto de etapas, la determinación de su alcance, la forma de interacción de las mismas y su aplicación a la resolución de problemas en términos de procesos y servicios. Está dirigida a proyectos, trascendiendo los aspectos tecnológicos y aportando una visión diferente para resolver los problemas, a la luz de la identificación transversal de procesos de negocio. Estos procesos se transforman en consumidores de servicios que pueden tener o no una implementación tecnológica. El objetivo es comprender a la organización y establecer claramente una serie de fases que ordenen las actividades a llevar a cabo de manera permanente, para alcanzar el propósito de contar con un ciclo de mejora continua de procesos, capaces de absorber los cambios que propone la realidad. Se describen a continuación las etapas propuestas, las actividades y el alcance de las mismas para el desarrollo de proyectos con enfoque en los procesos de negocio, como se muestra en la Figura 1.

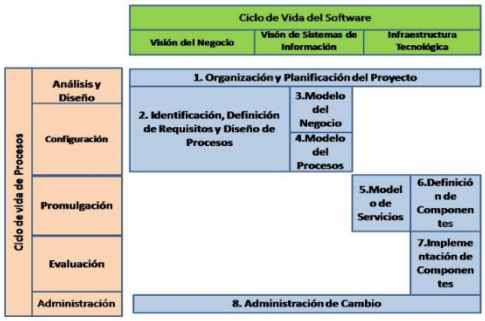


Figura 1: Metodología Propuesta: Etapa superpuesta en cada ciclo

En la Figura 2 se muestra el modo de interacción de estas etapas. Se grafican de manera independiente para facilitar su visualización, pero su descripción se considera de manera conjunta modeladas en un BPMS.

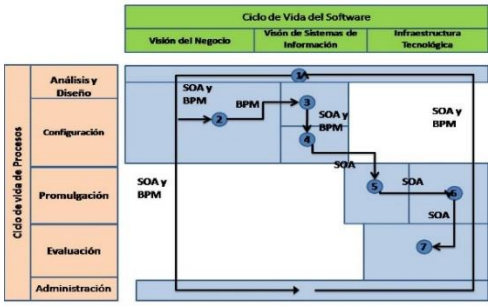


Figura 2: Metodología Propuesta: Etapa superpuesta en cada ciclo

2.1 Organización y plan estratégico

Esta etapa cubre toda la vida del proyecto y al igual que la última de las etapas, siempre está presente en cada una de las etapas sucesivas. En esta etapa la empresa se

prepara para iniciar el desarrollo del proyecto definiendo los equipos de trabajo, los principios que regirán la arquitectura y los marcos y herramientas de trabajos requeridos, además se establece el alcance, las restricciones y expectativas sobre el desarrollo del proyecto.

La organización del proyecto debe incluir los objetivos, el alcance, una consideración de alternativas de solución, el análisis financiero, patrocinadores y participantes, acuerdos de entendimiento, aprobaciones necesarias, un plan de proyecto y una revisión de los riesgos asumidos. Para llegar a esta organización se tiene en cuenta lo trabajado por Bazán [8].

2.2 Identificación, especificación de requisitos y diseño de proceso de negocio

Esta etapa incluye el análisis de requisitos como una actividad importante en el desarrollo de todo el ciclo de vida del software, que las metodologías tradicionales solamente consideran parcialmente, para lo cual es importante aclarar el término requisitos que desde el punto de vista de los usuarios simbolizan los problemas que deben ser solucionados por medio de la construcción de una aplicación [10]. Se incluye en esta propuesta una metodología basada en la idea del diseño participativo de procesos. Asimismo se inicia una iteración del proceso de arquitectura, debido a que se afianza el alcance, las limitaciones y las expectativas de la organización: también se inicia la creación de la visión de la arquitectura y se válida el contexto del negocio, tal como se desarrolla en la fase de visión del negocio en el *framework* TOGAF [7], donde se integra el desarrollo de procesos con los servicios prestados en la organización, de tal forma que tengan una alta relación con los procesos del negocio. En consecuencia, se desarrollan las siguientes actividades:

- *Diseño del proceso de flujos de trabajo de negocio*. Para esta actividad se debe realizar un cronograma de trabajo para recolectar la información sobre la temática, clientes, posibles usuarios de la organización o sector productivo, donde se desarrollará el nuevo sistema. Además se describe la estructura organizacional de los procesos de negocios, los sistemas de planeación y control, los mecanismos de gobierno y la administración de políticas y procedimientos [2]. También se debe obtener la información acerca del contexto de la organización a nivel de la arquitectura tecnológica, y de esta forma crear la visión de la arquitectura actual de la organización. Para esto se puede utilizar fuentes primarias, tales como entrevistas, encuestas, lluvia de ideas, entre otras, además de fuentes secundarias, tales como registros o documentos importantes de la organización para el sistema a desarrollar. Una vez se culmine estas actividades se debe realizar un diagrama de flujo de trabajo de negocio, donde se observe claramente el alcance, las limitaciones, los objetivos del negocio y su entorno. Para esto se utiliza el diagrama de flujo de trabajo de UML. Conjuntamente se desarrolla un documento donde se construye una declaración de la arquitectura. En esta actividad se observa el uso de la etapa de análisis y diseño del ciclo de vida de procesos.
- *Diseño del proceso de elicitación de requisitos*. En esta actividad se debe seleccionar las personas que participarán en el proceso de elicitación de requisitos. Además se seleccionan las técnicas de recolección de información, entre las que se encuentran: entrevistas y cuestionarios, *braingstorming* y reducción de ideas, *workshop*, JAD-JRD, puntos de vistas, casos de uso, prototipos, *storyboards*, reuniones, entre otras. Se debe diseñar los cuestionarios definiendo sus objetivos, población dirigida, tiempo de desarrollo y tipo de pregunta (abiertas y cerradas). También se elige un modelo de ciclo

de vida del proceso de elicitación (los más usuales son el modelo de PoHL y en espiral). El modelo de PoHL está compuesto por las actividades de elicitación, negociación, validación y verificación, especificación y documentación [29]. Finalmente, se elige una herramienta informática para dar soporte a la gestión, seguimiento, registro y documentación del proceso de elicitación. Entre las herramientas más utilizadas se encuentran REM y Enterprise Architect.

- *Elicitación de requisitos*. Para iniciar el proceso se debe realizar una priorización de los requisitos, para lo cual una de las técnicas más utilizada es la de comparación *pair wise*. Esta priorización de requisitos brinda un instrumento claro para seleccionar un conjunto optimizado de requisitos para la implementación, además, ayuda a los administradores del proyecto a predecir la satisfacción de los clientes antes que el sistema sea puesto en producción.

Como producto de esta etapa se genera un documento donde se especifican los requisitos del cliente, los usuarios y los desarrolladores. Estos requisitos deben estar clasificados en funcionales y no-funcionales, es decir, el modelo de requisitos, y teniendo en cuenta lo expresado por Fortune [28].

2.3 Modelado del negocio

En esta etapa se identifican los procesos de negocio y sus principales restricciones. Los procesos de negocio se describen como un conjunto de tareas en las que los actores participan según un flujo de trabajo determinado [25]. Estos procesos pueden organizarse en forma jerárquica, hasta con dos o tres niveles de anidamiento. Para modelar el negocio es necesario que la organización identifique los componentes de negocio, en cuanto a la estructura de la organización, los objetivos y las metas del negocio. El modelado de procesos mediante identificación de componentes requiere detallar los mismos componentes, tales como el conjunto de tareas o actividades o funciones que se llevan a cabo, realizadas en determinadas condiciones y bajo un determinado rol. Dentro del conjunto de actividades existen tareas automáticas, manuales o mixtas y son unidades indivisibles en términos de modelado, además de los servicios que ofrece el negocio, los roles que desarrolla cada servicio, actividad o tarea y la correlación entre la organización y sus funciones. Los subprocesos agrupan tareas interrelacionadas que pueden ser realizadas por varios roles y constituyen un producto intermedio para el proceso.

Cada proceso de negocio se representa como un caso de uso del negocio, que inicialmente se describe en forma textual. Para tener una visión general de los diferentes procesos de negocio de la organización puede construirse un diagrama de casos de uso del negocio, en el cual aparece cada proceso como un caso de uso UML [13]. Este diagrama permite mostrar los límites y el entorno de la organización bajo estudio y obtener un mapa de proceso.

El resultado de esta etapa es un mapa de procesos y la descripción de los casos de uso con su respectivo diagrama, para dar cuenta de la funcionalidad, es decir, un diagrama de casos de uso para el modelado del negocio, desarrollado en cualquier herramienta que soporte UML, o bien valerse de formularios prediseñados para volcar dicha información. Además, se debe definir la arquitectura base y la arquitectura objetivo, para poder realizar la documentación con el visto bueno de la arquitectura final de la aplicación. En el proceso se debe tener en cuenta dos aspectos importantes del proyecto: 1) el mapa de procesos, y 2) los requisitos definidos en el sistema, teniendo como base la fase de arquitectura del negocio implementada en el *framework* TOGAF [7] y la fase B referente a la arquitectura del negocio. Esto hace parte de la fase visión del sistema de información dentro de la etapa

de análisis, diseño y configuración para ciclo de vida de procesos, en el desarrollo de procesos del software, del cual se obtiene un documento avalado por la organización [22].

2.4 Modelado de procesos

En esta etapa se modela cada uno de los procesos identificados y detallados en los casos de uso del negocio, obteniendo un diagrama del proceso de negocio que se representa en el diagrama de clase de UML. Este diagrama describirá y estructurará los objetos de información que fluyen entre las actividades de un proceso de negocio y que representan datos del dominio. Estos datos del domino constituyen una base para crear el modelo de datos conceptual inicial. Este modelo incluirá las entidades y las relaciones entre cada uno y el ambiente, lo cual se describirá mediante un diagrama de clases UML, en el que las entidades se representan mediante clases (clases del dominio).

Acompañando a este modelo de procesos se aplican restricciones y objetos de información, que quedan en el diagrama de clases y por lo tanto resulta conveniente que se expliciten mediante un formulario genérico que acompañe el diagrama [22]. Este formulario incluirá por menos los siguientes ítems:

- Formulario para los objetos de información: nombre del objeto de información, atributos, restricciones y clase del dominio que lo representa.
- Formulario para las actividades: nombre de la actividad, origen (actividad que la origina), agente involucrado, pre-condiciones, post-condiciones, caso de uso del sistema que lo representa.

Este diccionario será el insumo principal en el desarrollo del diagrama de clases y permitirá identificar los atributos más importantes de cada dato, además de garantizar su unicidad. Esto responde a la fase denominada arquitectura del sistema de información que hace parte de la fase visión del sistema de información del *framework* TOGAF, dentro de la etapa de configuración del ciclo de vida del proceso [7, 31]

2.5 Modelado de servicios

Aquí se definen los servicios y su especificación y categoría. En esta etapa se incluye la investigación de servicios existentes y es el siguiente paso de refinamiento en el modelado de procesos. Los elementos de diseño que la guían están dados por la utilización de la metodología basada en la notación de círculos con cordones [26], junto con la disciplina de diseño propuesta por Delgado [10], que presenta una lista de actividades a realizar en este proceso de diseño y explicitando el objetivo de actividad, su entrada, su salida y los roles involucrado en la misma. Además, los servicios de negocios son los servicios de tareas con límites funcionales directamente asociados con una tarea o proceso de negocio. Es un servicio que tiene menos reuso potencial y generalmente es responsable de controlar la composición de otros servicios, por lo tanto, entre sus capacidades encapsula la lógica de negocio. Para mantener su independencia los servicios deben encapsular la lógica dentro de un contexto, que puede ser una tarea, una entidad de negocio o algún otro agrupamiento [9].

Lo más importante de esta etapa está dado por la definición del grado de granularidad de la pieza diseñada, como para que se pueda considerar un servicio con autonomía y atomicidad. Además, se obtiene el diagrama de casos de uso de servicios, el diagrama de secuencia de los servicios definidos y el diagrama de actividades a realizar en cada servicio.

2.6 Definición de los componentes

En esta etapa se definen componentes de software en términos de los servicios identificados y su modo de interacción (orquestación). Un componente es una pieza de software lo suficientemente pequeña para crear y mantener, y lo suficientemente grande para distribución y soporte con interfaces estándar para interoperabilidad. Claramente los web *services* y su modo de orquestación en una arquitectura SOA puede ser una solución tecnológicamente válida en esta etapa [30, 33].

También permite la definición de componentes y utiliza los conceptos para identificar y componer servicios con el objetivo de dar respuesta a procesos de negocio con alta reusabilidad. Además de evaluar y seleccionar las opciones de implementación de los componentes más importantes de los modelos identificados. Para esta etapa existen estándares como el SCA, que marca una fuerte tendencia absorbida por la mayoría de los productos comerciales existentes, pero también otorga un marco conceptual para el ensamblado y composición de servicios a nivel de diseño. UML también posee un diagrama que describe los componentes necesarios para el desarrollo de los servicios, identificando la relación entre ellos y los elementos necesarios en cada componente según las necesidades de la arquitectura tecnológica del proyecto, el cual responde a la fase de oportunidades y soluciones del *framework* TOGAF, y que hace parte de la fase de infraestructura tecnológica en la metodología propuesta, dentro de la etapa de promulgación del ciclo de vida de procesos. Por lo tanto, como resultado de esta etapa se obtiene el diagrama de componentes de UML desarrollado por el arquitecto y analista.

2.7 Implementación de los componentes

En esta etapa se lleva a cabo el despliegue en la plataforma elegida del resultado de las etapas anteriores. En ella no pueden faltar el desarrollo de prototipos, la retroalimentación con los actores de las etapas anteriores, la integración con los sistemas existentes y los aspectos no funcionales, tales como robustez y rendimiento. Además, describe cada uno de los componentes existentes en la organización y que están disponibles para el desarrollo del servicio dentro de los procesos escogidos en la empresa. También establece procedimientos para generar el cambio en la nueva arquitectura, si es necesario, y proporciona monitoreo continuo para asegurar la arquitectura.

2.8 Administración y seguimiento

Esta etapa se vincula con las de promulgación y administración del ciclo de vida de los procesos de negocio. Los productos obtenidos serán más fáciles de construir cuanto más apropiada sea la solución tecnológica adoptada. Los BPMS o los sistemas de gestión de *workflow* proporcionan las salidas para realizar una adecuada lectura de indicadores.

3 Conclusiones

La construcción de soluciones con enfoque de procesos y de servicios constituye una nueva manera de abordar los problemas, fomentando ampliamente la reusabilidad, la agilidad y flexibilidad para absorber los cambios y la idea de mejora continua de los procesos como consecuencia de lo anterior.

La descripción de las etapas del marco metodológico propuesto, si bien conlleva una lectura secuencial, fomenta una permanente realimentación entre ellas e incluso desde las etapas más tempranas, como la identificación de requisitos, que también se propone abordar desde un enfoque basado en el diseño de procesos.

Otro objetivo del enfoque es el acortamiento de la brecha entre el área de negocios y el de tecnología. Las etapas descritas fomentan la conformación de roles que son capaces de intercambiar información en un lenguaje común, y que además los compromete con el proyecto como una solución unificada e integrada a la organización.

La propuesta metodológica presentada beneficia la presentación de servicios, tanto nuevos como generados a partir de activos de software, para ser consumidos por procesos de negocios corporativos. Este modelo ayuda a las organizaciones a integrarse con sus pares, con sus clientes y con sus proveedores, más allá de las tecnologías e infraestructuras subyacentes, pero haciendo uso del valor que aportan las nuevas tendencias en computación.

Los trabajos futuros que se vislumbran son: 1) la creación de una herramienta CASE que utilice el modelo de integrabilidad propuesto, debido a que en la presente investigación fue necesario utilizar diferentes herramientas que solamente desarrollaron parte de la documentación del sistema.

La metodología se puede mejorar teniendo en cuenta los roles y actividades que se realizan utilizando la integración con otros *frameworks* que le dé mayores ventajas, beneficios y madurez en BPM a la organización. Además de socializarla en las empresas de desarrollo de software con el fin de que se masifique si utilización.

Referencias

[1] Bennett, N. & Balaba, N. (2007). Transformation to SOA: Part 1. From business process to service model architecture using IBM WebSphere Business Modeler and IBM Rational Software. IBM

[2] Oracle. BPEL Tutorial. Online [Dec. 2013].

[3] Fiorano Software (2012). ESB Best Practices. Online [Dec. 2013].

[4] Weske, M. (2012). Business process management: Concepts, languages, architectures. USA: Springer.

[5] OMG (2012). Service oriented architecture Modeling Language (SoaML) Specification. Object Management Group.

[6] The Open Group. Open SOA.

[7] The Open Group (2013). TOGAF 9 Manual. USA: The Open Group.

[8] Bazán, P.; Giandini, R. & Diaz, F. (2010). Tecnologías para implementar un marco integrador de SOA y BPM. Informe Técnico. UNLP, Buenos Aires.

[9] Robledo, P. (2010). SOA y BPM: La combinación perfecta para alinear TI y negocio. Club-BPM.

[10] Delgado, A.; García, I. & Ruiz, F. (2009). Desarrollo de software orientado a servicios basado en procesos de negocio. Memorias XII Conferencia Iberoamericana de Ingeniería de Requisitos y Ambientes de Software, pp. 3-14. Medellín, Colombia.

[11] De la Vara, J. et al. (2009). Modelado de requisitos de datos para sistemas de información basados en procesos de negocio. Memorias XII Conferencia Iberoamericana de Ingeniería de Requisitos y Ambientes de Software, pp. 1-14. Medellín, Colombia.

[12] Diaz J. et tal. (2009). Entornos para usar BPM en aplicaciones JAVA: un análisis comparativo. Memorias XI Workshop de Investigadores en Ciencias de la Computación, pp. 1-5. Universidad Nacional de San Juan, Argentina.

[13] OMG. Unified Modeling Language, version 2.2. The OMG.

[14] Errecalde, G. & Marcos, C. (2009). Una ontología de aspectos para la ingeniería de requisitos. Memorias XII Conferencia Iberoamericana de Ingeniería de Requisitos y Ambientes de Software, pp. 239-252. Medellín, Colombia.

[15] Bocanegra J.; Peña J. & Ruiz, A. (2009). Modelado de negocio interorganizacional: Una aproximación para la trazabilidad entre objetivos, modelos organizacionales y procesos de negocio. Memorias XII Conferencia Iberoamericana de Ingeniería de Requisitos y Ambientes de Software, pp. 1-14. Medellín, Colombia.

[16] Kerrigan, M. et al. (2009). SemanticWeb Service engineering for semantic business process management. Proceedings 4th International Workshop on Semantic Business Process Management, pp. 25-30. Crete, Greece.

[17] Bazan, P. (2009). Un modelo de integrabilidad con SOA y BPM. Tesis Maestría. Universidad Nacional La Plata.

[18] Arsanjani, A. et al. (2008). SOMA: A method for developing service-oriented solutions. IBM Systems Journal 47(3), pp. 377-396.
[19] Drake, J. (2008). Proceso de desarrollo de aplicaciones de software. Computadores y tiempo real.
[20] Garimella, K.; Lees, M. & Williams, B. (2008). Introducción a BPM. USA: Wiley.
[21] Oracle (2008). Business process management, service-oriented architecture, and web 2.0: Business transformation or train wreck? An Oracle White Paper.
[22] OMG (2009). Service oriented architecture modeling language (SoaML) - Specification for the UML el for Services (UPMS). OMG Adopted Specification.
[23] Sheina, D. (2008). Realising the promise of SOA and BPM. Ovum.
[24] Weske, M. (2008). Business process management: Concepts, languages, architectures. USA: Springer.
[25] Delgado, A. (2007). Desarrollo de software con enfoque en el negocio. Memorias XII Jornadas de Ingeniería del Software y Bases de Datos, pp. 1-8. Zaragoza, España.
[26] Erl, T. (2007). SOA Principles of Service Design. New York: Prentice Hall.
[27] Juric, M. et al. (2007). SOA Approach to Integration XML, Web services, ESB, and BPEL in real-world SOA projects. Boston: Packt Publishing.
[28] Fortuna M. (2008). Un modelo integrado de requisitos con casos de uso. PhD Tesis. Universidade Federal do Rio de Janeiro.
[29] Hurtado, D. (2007). Tesis metodología para el desarrollo de sistemas basados en objetos de aprendizaje. Tesis Maestría. Universidad del Norte.
[30] McKorkle, S. (2007). Model-Driven SOA - achieve higher productivity gains throughout the design process. USA: Telelogic.
[31] Oracle (2007). SOA Governance: Framework and Best Practices. An Oracle Whitepapaer.
[32] Brunswick, J. (2008). Extending the business value SOA through business process management. Oracle.
[33] NewComer, E. & Lomow, G. (2005). Understanding SOA with web services. USA: Addison-Wesley.
[34] Lau, D. & Mylopoulos, J. (2004). Designing web services with tropos. Proceedings IEEE International Conference on Web Services, pp. 306-313. San Diego, USA.
[35] Aiello, M. & Giorgini, P. (2004). Applying the tropos methodology for analysing web services requeriments and reasoning about quality of services. Upgrade V(4), pp. 20-26.
[36] Guo, L. et al. (2004). Mapping a business process model to a semantic web service model. Proceedings IEEE International Conference on Web Services, pp. 746-749. San Diego, USA.
[37] Giorgetti, G. (2003). Transformando - Prácticas de cambio en empresas argentinas. Buenos Aires: Editorial Eudeba.
[38] SCAMPI (2011). Appraisal Requirements for CMMI v1.1. Technical Report CMU/SEI-2011-TR-006. Software Engineering Institute.
[39] Morgenthal, J. (2001). Enterprise application integration with XML and Java. USA: Prentice Hall.

Segmentación de contornos ventriculares izquierdos, mediante el uso de la transformada generalizada de Hough

Hernando Velandia V.
Eduard García C.
Luis Mendoza
Universidad de Pamplona – Colombia

1 Introducción

La valoración morfológica y funcional del corazón es importante debido a que el daño cardiovascular (DCV) es una de las principales causas de mortalidad a nivel mundial [1]. De las cuatro cámaras cardiacas, el ventrículo izquierdo (VI) es la más susceptible a sufrir daño severo, en enfermedades tales como la hipertensión arterial, la diabetes mellitus o la arterosclerosis. En la aparición de una disfunción ventricular el VI debe hacer frente a una sobrecarga sanguínea debida a una presión elevada, por lo que fisiológicamente tiende a sufrir una hipertrofia. Si persiste el estrés el ventrículo acaba por sufrir una disfunción, que con el paso del tiempo se hace crónica e irreversible. Llegado a este punto, la función del miocardio queda comprometida y la capacidad contráctil del corazón acaba perdiéndose para siempre [2].

A nivel clínico, el análisis a través de imágenes generalmente es realizado de forma visual, lo cual genera una valoración que resulta imprecisa acerca de la función ventricular. La introducción de técnicas asistidas por computador ha incrementado la necesidad de desarrollar algoritmos eficientes para el tratamiento de imágenes cardiovasculares, con la finalidad de obtener resultados más precisos [3]. El propósito de este trabajo es diseñar e implementar una plataforma software que permita la extracción del contorno ventricular, posibilitando a futuro la cuantificación de indicadores de la funcionalidad cardiaca, tales como la fracción de eyección, el gasto cardiaco, etc., parámetros fundamentales de la funcionalidad cardiaca.

En primer lugar se presenta una breve descripción de antecedentes, donde se muestra la aplicación de la transformada generalizada de Hough en diversos campos del saber; la sección 3 relaciona aspectos importantes del sistema angiográfico de Rayos X y el equipo empleado en el proceso de adquisición de las imágenes; por su parte, la TGH se describe en la sección 4, resaltando principalmente su estructura matemática; en la 5 se relaciona la forma como se seleccionaron los estudios imagenológicos. Hasta éste punto se expone en el trabajo la teoría y la relevancia de emplear éste tipo de técnicas a nivel médico. Partiendo de éstas bases es posible especificar detalladamente todo el proceso de construcción de la TGH, descrita como metodología y relacionada en la sección 6, donde se explican aspectos como la construcción de la tabla R y las etapas de pre-procesamiento y procesamiento respectivamente. Los resultados obtenidos en el proceso de validación son presentados en la sección 7, donde se exponen los enfoques cualitativos y cuantitativos, obteniendo un buen desempeño del algoritmo implementado. Al final se presenta las conclusiones obtenidas.

2 Antecedentes

La TGH ha sido empleada en diferentes campos de la ciencia en la detección de curvas analíticas (líneas, círculos, etc.) donde se denomina transformada de Hough, también en

las denominadas formas no-exactas (contornos irregulares), la cual es el eje principal del presente trabajo. Ballard (4) mostró como los bordes o contornos de una forma arbitraria no-analítica pueden ser utilizados para construir una relación entre el espacio de imagen y el correspondiente al de la transformada de Hough. Barinova, Lempitsky y Kholi [5] la emplearon en su investigación denominada, relacionándola como una nueva estructura probabilística para la detección de objetos [5]; y Bakulina, Hlindzich y Maenner [6] relacionaron dos tipos de aplicación del método de Hough, enfocadas a la detección de objetos médicos en imágenes de resonancia magnética y ultrasonido respectivamente [6].

3 Descripción del sistema

La angiografía por rayos X es una modalidad ampliamente utilizada en la evaluación de la función cardiaca. Para visualizar el corazón con esta modalidad el paciente es sometido a una intervención denominada cateterismo cardiaco, durante la cual el VI se llena con un material de contraste radio-opaco. La adquisición de las imágenes puede ser realizada en dos planos: considerando la vista Oblicua Anterior Derecha (OAD) 30° y la vista Oblicua Anterior Izquierda (OAI) 60°. El tiempo de adquisición promedio es de alrededor de 8 a 10 s, cubriendo de 7 a 9 ciclos cardiacos. La distribución del contraste inyectado se considera óptima alrededor del segundo o tercer ciclo cardiaco [7]. El equipo utilizado para la adquisición y visualización del estudio de angiografía es el INNOVA 2000, construido por la General Electric y que presenta las siguientes características: posee un detector digital de estado sólido revolution, las imágenes completamente digitales presentan un campo de vista de 20.5 cm. X 20.5 cm., las secuencias de imágenes se registran a 30 cuadros por segundo con una resolución de 512 x 512 pixeles, y cada uno de ellos se representa con 256 niveles de gris [8, 9].

4 Transformada generalizada de Hough (TGH)

TGH se define como una transformada que permite detectar curvas no analíticas, es decir, que detecta el contorno de formas irregulares en objetos. En la Figura 1 se puede observar que los datos necesarios para obtener un contorno irregular basado son las angulaciones (α, Φ) y la distancia r.

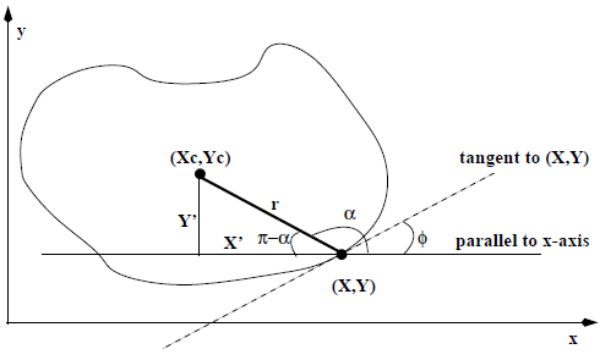


Figura 1: Representación gráfica de TGH

Donde:

Xc, *Yc* definen el centro de masa de la forma
r = radio desde el centro de masa hasta un punto (*X*, *Y*)
α = ángulo del radio con respecto a una horizontal paralela al eje x
Φ = ángulo de la recta tangente con respecto a una horizontal paralela al eje x

A nivel matemático es posible representar el centro de masas de la siguiente manera:

$$X_C = X + rcos(\alpha) \qquad (1)$$

$$Y_C = Y + rsen(\alpha) \qquad (2)$$

El arreglo conformado por r_i, α_i, Φ_i, se le denomina Tabla *R*. Esta tabla es una arreglo que contienes valores de radios y ángulos de diferentes imágenes que están relacionados con un punto del objeto que se procesa, y es la base para construir la TGH [4, 10].

5 Selección de las imágenes

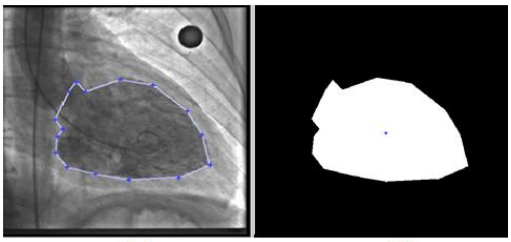
Para el desarrollo del presente proyecto se trabajó con tres estudios de angiografía, para los cuales se seleccionaron cuatro cuadros de imagen (*frames*), de la siguiente manera: una correspondiente a la fase diastólica final, otra a la sistólica final y las dos restantes se escogieron en transiciones de las dos primeras.

6 Metodología

La metodología empleada en la presente investigación se divide en dos bloques, el primero consiste en la generación del modelo de ajuste o patrón (tabla *R*), y el segundo en la detección del contorno ventricular a través del modelo elaborado.

6.1 Construcción de la tabla *R*

Inicialmente, a través de una herramienta interactiva se procede a demarcar por parte del especialista el contorno ventricular. Estas coordenadas son almacenadas con la finalidad de generar una imagen binaria en la que solamente estén presentes el objeto de interés (VI) y el fondo. El resultado se presenta en la Figura 2.



(a) (b)

Figura 2: (a) Selección del contorno ventricular por parte del especialista. (b) Máscara binaria generada y detección del centroide o centro de masas

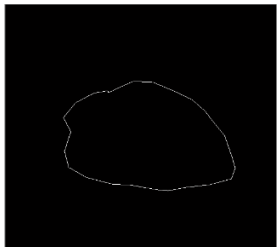


Figura 3: Contorno final del prototipo

Seguidamente se implementó un algoritmo que permitiera la detección del borde ventricular, junto a la particularidad de que los puntos o coordenadas que lo conforman se encuentren ordenadas, ya sea en el sentido de las manecillas del reloj o viceversa. El resultado es presentado en la Figura 3. Como se explicó antes, para generar la tabla R es necesario calcular los valores correspondientes para α, Φ y r respectivamente. El procedimiento empleado sobre el contorno prototipo (Figura 3) para la búsqueda de dichos parámetros es el siguiente:

Cálculo de r. Se realiza graficando rectas desde el centro de masa del prototipo (centroide) a cada uno de los puntos del contorno del prototipo (Figura 5). Posteriormente se procede a estimar su distancia algebraica y finalmente se almacena su valor en la tabla *R.*

Cálculo de α. Como se ilustra en la Figura 1, para obtener este ángulo es necesario generar rectas horizontales a cada punto del contorno ventricular (Figura 6). Realizando este procedimiento, solamente falta calcular la pendiente de cada uno de los segmentos y emplearlos en la ecuación 3.

$$\alpha = \tan^{-1}\left({(m1 - m2)}/{(1 + m2 \cdot m1)}\right) \qquad (3)$$

Donde *m1* representa la pendiente del radio *r* y *m2* la correspondiente al segmento horizontal. Finalmente se almacena su valor en la tabla *R.*

Cálculo de Φ. Aquí es necesario generar segmentos tangentes a cada uno de los puntos del contorno. Dicho procedimiento se compone de dos partes: los puntos del borde deben estar ordenados en sentido horario, y cada uno debe referenciar el vecino anterior (P_{i-1}) y posterior (P_{i+1}) a cada punto que se esté analizando (P_i). La metodología anterior se ilustra en la Figura 4.

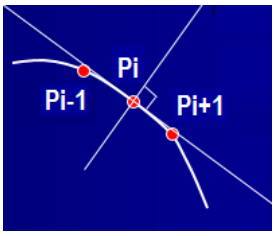


Figura 4: Selección del punto de referencia y sus vecinos más cercanos

Por ser una imagen, estos puntos están compuestos de manera espacial por un par de coordenadas *X* y *Y*, representados de la siguiente manera: P_i (X_i, Y_i), P_{i+1} (X_{i+1}, Y_{i+1}) y P_{i-1} (X_{i-1}, Y_{i-1}). Las componentes tangenciales a éste punto de referencia P_i, es posibles calcularlos a través de la ecuación 4.

$$(t_X\ , t_Y) \approx \frac{(d_X\ , d_Y)}{\sqrt{d_X^2 + d_Y^2}} \qquad (4)$$

Donde d_X y d_Y son calculados de la siguiente manera:

$$d_X = X_{i+1} - X_{i-1} \qquad (5)$$
$$d_Y = Y_{i+1} - Y_{i-1} \qquad (6)$$

Seguidamente, y con la finalidad de generar la recta tangente correspondiente (interpolación), se aplica la ecuación 7.

$$(X, Y) + i \cdot (S_n(-t_Y), S_n t_X) \qquad (7)$$

Donde (*X, Y*) son las coordenadas del punto central o de análisis, *i* =….-2, -1, 0, 1, 2,……, y S_n es una constante elegida heurísticamente. Finalmente, y teniendo construidas las rectas tangente y horizontal para cada punto del contorno, es posible obtener los valores del ángulo *Φ.* La metodología empleada para esto es la misma aplicada para el *Cálculo de α* y su valor se almacena en la tabla *R.* En la Figura 5 se evidencia la forma como se trazó

la recta *r.* Por su parte, la Figura 6 muestra un segmento ampliado del contorno y el trazo de los radios *r* y las correspondientes rectas tangenciales (líneas color violeta) y horizontales (líneas color verde) respectivamente.

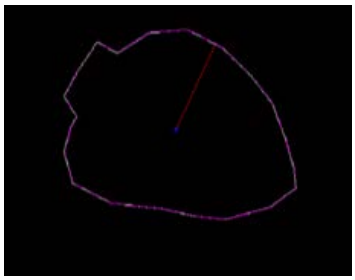


Figura 5: Demarcación de la recta *r*, para un punto del contorno ventricular

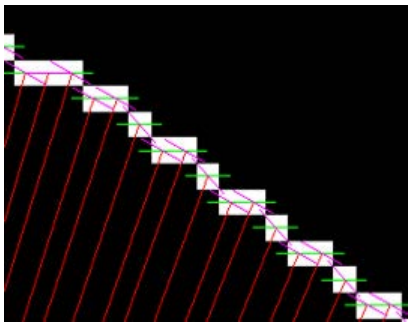
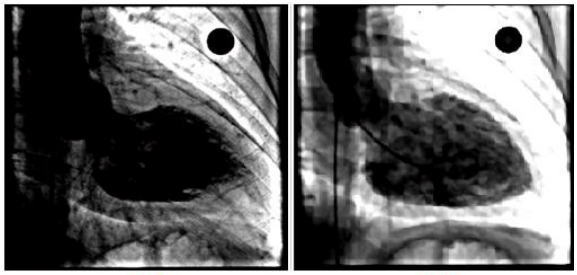


Figura 6: Trazo de las componentes tangencial y horizontal

6.2 Detección del contorno ventricular a través del modelo elaborado (tabla *R*)

Para alcanzar éste objetivo se busca convertir la imagen de trabajo en un mapa de contornos, donde sea posible comparar la estructura de la tabla *R* con las diferentes formas presentes en l imagen (costillas, catéter, abdomen, ventrículo, etc.) y de esa manera seleccionar de todas ellas la correspondiente al borde ventricular (finalidad de la TGH). El procedimiento es el siguiente:

Pre-procesamiento. Como se aprecia en la Figura 2(a), una imagen de angiografía presenta una considerable cantidad de ruido, y por lo tanto es importante realizar un proceso de acondicionamiento o mejora de la información. Para tal fin, en primer lugar se implementó un realce de contraste por ventana y nivel, el cual buscaba opacificar de mejor manera la zona de interés (VI), seguido de un filtro suavizador [11]. El resultado se ilustra en la Figura 7.



(a) (b)

Figura 7: (a) Imagen realzada, (b) Imagen suavizada

Con una mejor visualización de los datos, el paso a seguir es generar una imagen de contornos o sólo bordes, que permita emplear la información recopilada en la tabla *R*. Para

esta tarea se emplea el algoritmo de Canny [12], el cual esta basado en el cálculo de la derivada. El resultado se aprecia en la Figura 8.

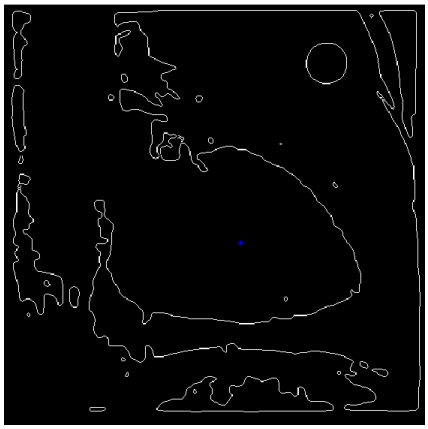


Figura 8: Resultado obtenido al emplear el detector de contornos

Finalmente, el propósito de esta fase es la de extraer el contorno ventricular con la menor cantidad de ruido, es decir, con la mínima cantidad de estructuras ajenas a éste. Para este caso se aplicó la morfología matemática [13], empleando el filtrado por áreas, la detección de continuidad, etc. La imagen obtenida se presenta en la Figura 9.

Figura 9: Empleo de morfología matemática a la imagen de trabajo

Aplicación de la TGH. Teniendo la mejor versión posible en el mapa de contornos, el objetivo se centra en determinar a través de la TGH cuál de las estructuras restantes es el contorno ventricular. Con esto en mente, lo primero es distinguir esos dos conjuntos de puntos, lo que es posible a través de un proceso morfológico denominado etiquetaje, en el cual cada integrante de la escena (basado principalmente en la continuidad de sus puntos) se rotula con un nivel de gris diferente o etiqueta. Para el caso de la Figura 9 solamente se presentarían dos conjuntos de puntos o dos etiquetas.

Seguidamente, a cada uno de los puntos que conforman los contornos presentes en la imagen se les calcula el ángulo Φ y se proceden a almacenar por etiquetas. A continuación, se compara cada uno estos ángulos con los almacenados en la tabla R, y si son iguales de ésta última se extraen los valores correspondientes a r y α respectivamente. Con dichos valores es posible calcular el centroide correspondiente (X_C, Y_C), a través de las ecuaciones 1 y 2. Finalmente, cada par de puntos (X_C, Y_C) es almacenado en una matriz de votos (acumulador), de donde se buscará el valor que más se repita junto a su etiqueta y así se decide el ganador y que será el correspondiente al borde buscado (ventrículo). El resultado se muestra en la Figura 10.

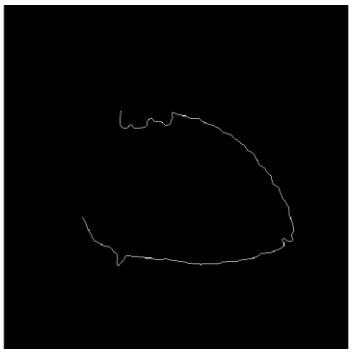


Figura 10: Resultado final del empleo de la *TGH* en la detección del contorno ventricular

En función de realizar un proceso comparativo o de validación de los resultados obtenidos, respecto a los alcanzados por el especialista de manera manual, se procede a implementar un algoritmo de cerramiento y llenado del contorno ventricular de la Figura 10. El resultado se observa en la Figura 11.

Figura 11: Cerramiento y llenado del contorno ventricular

7 Resultados y validación

Para el análisis de resultados se seleccionaron los estudios y las imágenes como se explicó previamente. Es importante resaltar que para cada estudio (paciente diferente), es necesario crear una tabla R. A nivel de demostración de la segmentación alcanzada, se presentarán 3 de los 12 *frames* empleados en éste proyecto. Se escogió uno de cada estudio para ejemplificar la variedad en los datos y contornos. En la Figura 12 se ilustran las imágenes originales donde se pretende detectar el contorno ventricular. Se superpone de igual manera el trazo hecho por el especialista, con la finalidad de realizar el proceso de validación.

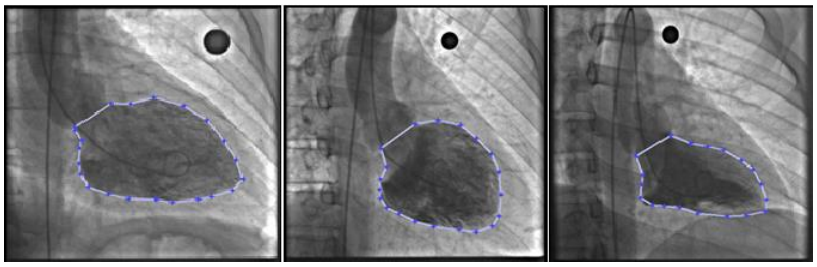


Figura 12: Estudios iniciales y en donde además se superpone el trazo del contorno ventricular realizado por el especialista

En la Figura 13 se presenta la detección de contornos obtenidos a través de la TGH. Procedimiento detallado a lo largo de este trabajo y que presenta buenos resultados.

Figura 13: Contornos detectados a través de la TGH

En la Figura 14 se realiza el llenado y cerramiento de los contornos obtenidos, con la finalidad de cuantificar el proceso de detección.

Figura 14: Cerramiento y llenado de contornos

Finalmente, en la Figura 15 se realiza la superposición de los bordes detectados sobre las imágenes originales, para visualizar de mejor forma los resultados obtenidos.

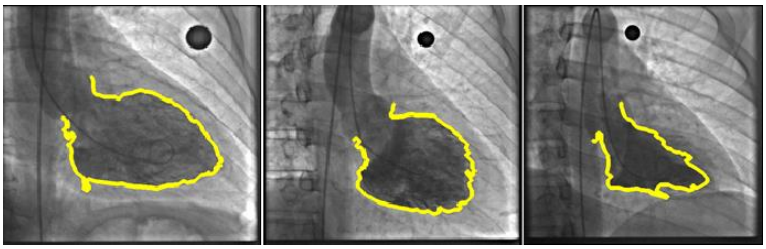


Figura 15: Superposición de los contornos detectados, sobre la imagen original

A nivel cuantitativo, en la Figura 16 se presentan los porcentajes de error obtenidos del proceso de validación. Dicha estimación fue realizada con base en la ecuación 8, y donde se compara la segmentación manual (especialista) y la obtenida con la TGH.

$$\%Error = \frac{(Seg.Manual - Seg.TGH)}{Seg.Manual} \cdot 100 \qquad (8)$$

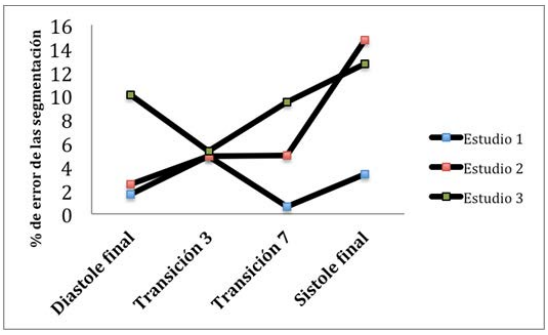


Figura 16: Análisis cuantitativo

Dichos porcentajes se dividieron en diástole final, sístole final y transición 3 y 7. Estas transiciones son imágenes que se seleccionan entre diástole y sístole final. Para esta

investigación las áreas fueron medidas en pixeles. Es importante resaltar que el mejor resultado se obtuvo en la transición 7 del estudio 1, donde el valor estimado fue del 0,65%. El porcentaje de error total del sistema fue del 4,56%.

8 Conclusiones

La segmentación de imágenes ventrículográficas basadas en TGH permite explorar nuevas resultados, los cuales soportan el desarrollo de herramientas de apoyo para los especialistas. La metodología planteada permitió vencer los retos que presentan las imágenes de éste tipo, tales como bajo contraste, ruido en la imagen, bajos detalles del contorno, etc. Esto se logro gracias a las técnicas de pre procesamiento utilizadas: filtro de mediana, realce de contraste, detectores de contornos, binarización y ordenamiento de puntos.

Por otro lado, se demostró que por medio de la TGH es posible segmentar imágenes del corazón, logrando un error promedio de 4,56%. La tabla *R*, como estructura de análisis, generaliza gran parte de los contornos que se presentan en el procesamiento de imágenes ventriculográficas.

Finalmente, es importante resaltar que la selección de las imágenes para la selección de patrones es fundamental para el buen funcionamiento de la TGH, ya que si existen errores en la selección de datos, cuando se quiera realizar un nuevo análisis de contornos los resultados no serán los más relevantes, y por ende el error del sistema aumentará.

Referencias

[1] Mackay, J. & Mensah, G. (2004). Atlas of heart disease and stroke. USA: World Health Organization.

[2] http://www.consumer.es/salud/problemas-de-salud/ [Feb. 2014].

[3] Bravo, A. (2006). Simulación y reconstrucción en 4-D del ventrículo izquierdo en imagenología cardiaca. Tesis Doctoral, USB, Venezuela.

[4] Ballard, D. (1980). Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognition 13(2), pp. 111–122.

[5] Barinova, O., Lempitsky, V. & Kholi, P. (2012). On detection of multiple object instances using Hough transforms. IEEE Transactions on Pattern Analisys and Machine Intelligence 34(9), pp. 1773-1784.

[6] Bakulina, A., Hlindzich, D. & Maenner, R. (2010). Fast dual-point Hough transform for medical object recognition. Proceedings 26th IEEE International Symposium on Computer–Based Medical Systems, pp. 297-302. Perth, USA.

[7] Oost, E., Koning, G. & Sonka, M. (2006). Automated contour detection in X ray left ventricular angiograms using multiview active appearance models and dynamic programming. IEEE Transactions on Medical Imaging 25(9), pp. 1158-1171.

[8] http://www.gehealthcare.com [Mar. 2014].

[9] Sagardi, J. (2002). El detector digital en un sistema de imagen cardiovascular. Revista de Física Médica 3(1), pp. 35-38.

[10] Hough, P. (1962). Method and means for recognizing complex patterns. U.S. Patent 306954.

[11] Gonzalez, R. & Woods, R. (2007). Digital image processing. USA: Prentice Hall.

[12] Canny, J. (1986). A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8(6), pp. 679-698.

[13] Cuevas, E., Zaldivar, D. & Pérez, M. (2010). Procesamiento digital de imágenes con Matlab y Simulink. España: Alfaomega Grupo.

Caracterización del sensor EEG BR8 Plus: Verificación del estado base y los ritmos beta y theta utilizando la prueba Stroop de atención visual

Estefanía Millán D.
Carolina Jiménez M.
Daniela Ospina T.
Oscar Henao G.
Universidad Tecnológica de Pereira – Colombia

1 Introducción

El sensor EEG BR8 Plus es un dispositivo utilizado para el registro de señales bioeléctricas en el ámbito de interfaz cerebro-computador. En Colombia no se tiene muchos electroencefalógrafos al alcance de los investigadores en esta área, por eso es importante introducir este sensor en la región, puesto que es de bajo costo y ofrece alta resolución comparado con otros dispositivos. Además, es difícil encontrar aplicaciones o desarrollos, por lo que en este trabajo se expone una verificación de que las señales EEG se corresponden con las citadas en la literatura. Inicialmente se prueba el estado base del sensor, identificando cambios en el EEG frente a diferentes movimientos faciales, tales como abrir y cerrar los ojos, pestañear, o plisar los dientes, que son referentes usados por el fabricante para la prueba del funcionamiento y para categorizarlas como datos de calibración.

El dispositivo cuenta con la interfaz Brain Viewer, que está en fase de prueba y posee problemas al registrar los filtros dados para cada ritmo. Se utiliza una prueba de atención visual, el test de Stroop, en voluntarios sanos; se espera que las señales de los ritmos *theta* y *beta* se destaquen en los canales Fp1, Fp2, Pz, O1 y O2 con una mayor amplitud en un pico diferenciador de frecuencia acorde con lo citado en la literatura.

2 Sensor EEG BR8 Plus

Es un sistema EEG inalámbrico de la compañía Brain Rhythm Incorporation, que se encuentra a la vanguardia de esta tecnología de dispositivo portátil que le permite acceder a las señales EEG en tiempo real. Funciona mediante la recopilación de los ritmos cerebrales a través de electrodos secos en el dispositivo (tanto de pines como de espuma). Este sensor, como se observa en la Figura 1, consta de dos referencias REF/GND (representadas por las pinzas azules), posee 8 canales para la ubicación de los electrodos y cuenta con aislamiento auditivo simbolizado por los auriculares. Sus características técnicas son:

- Frecuencia de muestreo de 1000 Hz.
- Resolución de 24 bit.
- Ganancia de 1365.
- Ancho de banda de 0.12 Hz – 125 Hz.
- Vía de transmisión: Bluetooth 2.1.
- Tiempo de uso de la batería cerca de 11 horas.
- Dimensiones 20 cm x 20 cm x 17 cm.
- Protocolo de obtención de la señal RS232.
- Archivos de salidas edf / vdf / txt / cnt /csv.

Figura 1: Sensor electroencefalográfico BR8 + [1]

El dispositivo transmite los datos a través de una interfaz cerebro-computador (BCI) para realizar investigaciones cognitivas y aplicaciones en el área. Los electrodos se distribuyen con base al Sistema Internacional 10-20, localizados en Fp1, Fp2, Fz, C3, C4, Pz, O1 y O2, tal como se muestra en la Figura 2 [1].

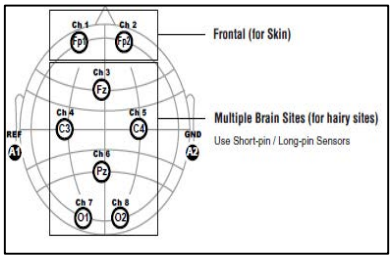


Figura 2: Distribución de los electrodos [1]

Estos electrodos tienen la característica de ser secos, es decir, no necesitan solución salina, se encuentran formados por una base de espuma con impedancia de 10 ~ 50k y los electrodos de pines cortos y largos de metal conductor poseen un fuelle de resorte y impedancia 5 ~30k [2, 3].

3 Ritmos cerebrales

La actividad de la corteza cerebral está sustentada en gran medida por la variedad de oscilaciones que genera el cerebro, y que abarcan un número de frecuencias y sitios anatómicos así como correlatos conductuales relacionados [4]. En adultos sanos, las amplitudes y frecuencias de estas señales presentan variabilidad inter-sujeto, además también cambian con la edad. Hay cinco grandes ritmos cerebrales que se distinguen por sus diferentes rangos de frecuencia bajas y altas respectivamente: *delta* (δ), *theta* (θ), *alpha* (α), *beta* (β) y *gamma* (γ). La Figura 3 muestra los ritmos cerebrales.

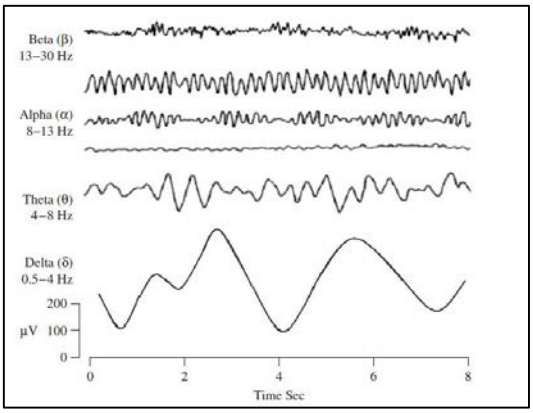


Figura 3: Cuatro típicos y dominantes ritmos del cerebro, de altas a bajas frecuencias [5]

Los ritmos *delta* se encuentran en el rango de 0,5 a 4 Hz, se asocian principalmente con el sueño profundo y pueden estar presentes en estados de somnolencia. Los ritmos *theta* se encuentran en el rango de 4 a 7,5 Hz y poseen baja amplitud. Estos ritmos se han asociado con el acceso al inconsciente, la inspiración creativa y la meditación profunda. Una onda *theta* suele ir acompañada de otras frecuencias y parece estar relacionada con el nivel de excitación. Esta onda juega un papel importante en la infancia y la niñez [6].

Los ritmos *alpha* aparecen en la mitad posterior de la cabeza, generalmente sobre la región occipital del cerebro, y pueden ser detectados en todas las partes de lóbulos posteriores del mismo. Para estos ritmos la frecuencia se encuentra en el rango de 8 a 13 Hz y comúnmente aparecen como una señal de forma sinusoidal. Los ritmos *alpha* se han definido para indicar una conciencia relajada, sin prestar ninguna atención o concentración. En la mayoría de los sujetos se producen con los ojos cerrados, por lo que se ha afirmado que no es más que un patrón de espera o de exploración producido por las regiones visuales del cerebro. Se reduce o se elimina con la apertura de los ojos, al oír sonidos desconocidos, o por ansiedad, concentración mental o atención. Un ritmo *alpha* tiene una amplitud normalmente menor de 50 mV y es mayor sobre las zonas occipitales [7].

Un ritmo *beta* es la actividad eléctrica del cerebro que varía en el intervalo de 13 a 30 Hz. Es el ritmo de vigilia normal del cerebro que se detecta en los adultos sanos, está asociado con el pensamiento activo, la atención activa y la resolución de problemas concretos. Un ritmo *beta* de alto nivel puede ser adquirido cuando un ser humano está en un estado de pánico. La actividad *beta* rítmica se encuentra principalmente sobre las regiones frontales y centrales. La amplitud del ritmo normalmente es menor a 30 mV. Las frecuencias superiores (principalmente hasta 90 Hz) corresponden al rango *gamma*. Aunque las amplitudes de estos ritmos son muy bajas y su ocurrencia es rara, su detección puede ser utilizada para la confirmación de ciertas enfermedades cerebrales [8].

4 Prueba de escogencia aleatoria de color (Stroop)

Esta prueba fue desarrollada por John Stroop [9], su objetivo es evaluar la capacidad del sujeto para inhibir una respuesta automática y para seleccionar una respuesta con base en un criterio arbitrario. Consta de tres procesos: 1) el usuario debe leer nombres de colores puestos al azar impresos en letras negras, 2) se le presentan óvalos de colores, donde debe decir el color que hay en ellos, y 3) el usuario debe leer los nombres de los colores escritos en tinta de otro color (la mayoría de veces el color de la letra no corresponde al nombre del color), como se observa en la Figura 4. El mayor interés de esta prueba es la conducta del sujeto cuando se le presentan palabras coloreadas en tintas que no corresponden, lo que permite evaluar en detalle los procesos de atención y memoria: una tarea automática favorecería la lectura de la palabra, mientras que un proceso controlado no-automático sería denominar el color de la tinta [9].

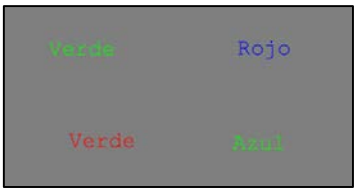


Figura 4: Tercera etapa de la prueba de Stroop [10][10]

El efecto Stroop se refiere a la interferencia que se produce en el sujeto cuando realiza el último proceso. Esta prueba precisa la ejecución en un proceso controlado de una tarea

novedosa, mientras se debe impedir la intrusión de un proceso automático. Es decir, es como se llama a una clase de interferencia semántica producida como consecuencia de la automaticidad en la lectura. Esto ocurre cuando el significado de la palabra interfiere en la tarea de nombrar, por ejemplo, el color de la tinta en que está escrita [11].

5 Metodología

5.1 Estado Base del Sensor EEG BR8 Plus

Para la identificación de las señales bioeléctricas del cerebro se implementó el uso del sensor EEG BR8 Plus, ubicándolo de manera correcta en la cabeza del usuario y acoplando los electrodos de referencia en los lóbulos de las orejas para que las medidas sean mono-polares. Se inicializó el software del sensor Brain Viewer y posteriormente se le indicó al individuo que se mantuviera en un estado de relajación en el transcurso de la toma de datos, con el fin de evitar ruido debido a movimientos involuntarios. Esto se debe a que los músculos representan artefactos de movimiento en el electroencefalograma de órdenes de magnitud más grandes que la señal EEG. En esta toma de datos no se utilizó ningún tipo de estímulos, como se visualiza en la Figura 5.

Figura 5: Registro de EEG por medio del sensor EEG BR8 Plus

En los registros brindados por la compañía Brain Rhythm Incorporation se encuentran pruebas EEG con ojos abiertos, pestañeo, ojos cerrados y apretando los dientes, que se corroboraron como estado mínimo viable de uso para la electroencefalografía de los ritmos cerebrales.

5.2 Ritmos cerebrales Theta y Beta en la señal EEG con base al test de Stroop

El sensor tiene en su interfaz las opciones de filtro para visualizar los ritmos cerebrales *delta*, *theta*, *alpha*, *beta* y *gamma*, pero presenta inconvenientes al momento de registrar la señal de cada ritmo, porque por su reciente aparición la interfaz está en una fase de prueba. Es por esto que se aplicó una prueba de atención visual, el test de Stroop, con el objetivo de verificar el correcto funcionamiento del sensor, comprobando la manifestación de los ritmos cerebrales, especialmente *theta* y *beta*, propios de la atención visual en los registros EEG.

Se aplicó una parte de la prueba de Stroop utilizando la herramienta Psychtoolbox de MATLAB. Este demo hace referencia a la tercera etapa de la prueba y consta de leer palabras coloreadas en tintas que no corresponden a su nombre, como se explicó anteriormente [10]. Está diseñado con cierto número de ensayos por condición, siendo el número 1 para un total de nueve ensayos. Luego se sigue con el protocolo de la batería Neuropsicológica NEUROPSI, referente al área de atención/concentración relacionada con la prueba de Stroop, que consiste en la presentación de una matriz 9x4. Esto indica que en la prueba se muestra un total de 36 palabras; se modificó en el código anterior el número de ensayos por condición cambiando el número 1 por el 4 para tener los 36 ensayos, logrando una matriz de respuesta de la prueba con dimensiones de 4x36 [12].

Debido a que se trabajó con una prueba asociada con eventos, el procesamiento de las señales EEG de cada uno de los canales debió ser segmentado por épocas. Una época del EEG se toma como el tiempo asociado con la presentación del estímulo al usuario mientras se registra el EEG, es decir, las épocas que se han definido corresponden a cada evento en la prueba de Stroop mostrada al usuario (palabra coloreada). Estas hacen referencia al intervalo de tiempo en el que se presenta la palabra y es allí donde se buscan los ritmos asociados a la prueba [13].

Para segmentar el primer canal se tomó el tiempo de respuesta del usuario de la primera palabra presentada y se le sumó un segundo, el cual hace referencia al lapso que hay entre la respuesta que da el usuario en ese momento y la aparición de la segunda palabra, con todo esto se compone la primera época. La segunda época parte del tiempo con que finalizó la primera, se le suma el tiempo de respuesta y el segundo del lapso, creando así un consecutivo con las demás épocas y llegando al total de 36. A continuación se procesó la señal y se verificó si el sensor EEG BR8 registró bien los ritmos cerebrales.

Las señales obtenidas por el sensor se guardaron en formato *.edf*, compatible con el programa MATLAB a través de EEGLAB, un toolbox interactivo que sirve para el procesamiento continuo y eventos relacionados con EEG, MEG y otros datos electrofisiológicos [14]. Cuando EEGLAB lee las señales del sensor crea una estructura EEG, donde se encuentra la matriz *data* compuesta por ocho filas (correspondientes a los ocho canales del sensor) y con columnas de igual longitud de la señal (en milisegundos). Posteriormente, se crean variables en las que se guardan las épocas de los canales 1, 2, 6, 7 y 8, obteniendo 36 épocas para cada canal. Se utilizó la herramienta FDATool para elaborar los filtros pasa-banda IIR *Butterworth* de orden 26. Uno para encontrar el ritmo *beta* con un rango de frecuencias de 13 a 30 Hz y el otro para encontrar el ritmo *theta* con un rango de frecuencias de 4 a 8 Hz. Se utilizó este filtro con estas características porque la herramienta necesita un mayor orden para los mismos requerimientos, en comparación con otros como los Chebyshev o el elíptico, siendo muy satisfactorio en los resultados. Después de filtrar cada época se esperaba encontrar el ritmo *theta* en los canales 1, 2 y 6, y el ritmo *beta* en 1, 2, 7 y 8, observando en cada espectro de frecuencia un pico en el rango indicado.

6 Resultados

6.1 Estado base del sensor EEG BR8 Plus

- *Ojos abiertos*. Para las señales registradas por el sensor, mostradas en la Figura 6, se observa que no hay variación importante obtenida en cada uno de los canales, debido a que el usuario se encontraba en un estado de relajación y con los ojos abiertos. Esto indica que la actividad cerebral en ese instante de tiempo estaba en reposo.

Figura 6: Señales bioeléctricas con ojos abiertos

- *Pestañeo*. La Figura 7 muestra la actividad eléctrica del cerebro en los instantes en los que el usuario pestañea sus ojos. Se observan varios picos positivos que representan el pestañeo del usuario. Estos cambios se dan en los canales 1, 2 y 3 y hacen referencia a Fp1, Fp2 y Fz, respectivamente. Por lo que se concluye que el cambio muscular que se da en los ojos representa la actividad eléctrica en la zona frontal.

Figura 7: Señales bioeléctricas con pestañeo

- *Ojos cerrados*. La actividad eléctrica cerebral que se da cuando el usuario cierra los ojos presenta una curvatura inicial evidente en los canales 1 y 2 de la Figura 8. Consecutivamente se observa que en el transcurso de la toma de datos no hay ninguna otra deformación positiva en las señales de los canales.

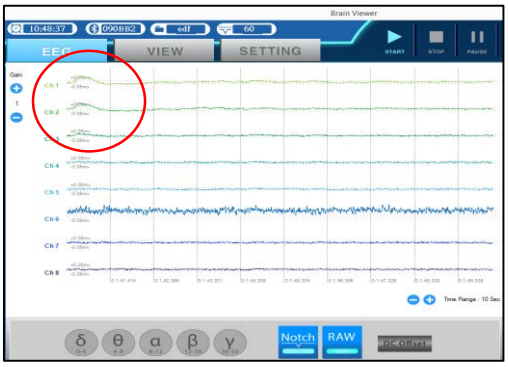


Figura 8: Señales bioeléctricas con ojos cerrados

- *Plisar los dientes*. Cada vez que el usuario aprieta los dientes se genera una contracción muscular, lo que provoca que el electroencefalograma registre el movimiento. En la Figura 9 se puede diferenciar el momento en el que el usuario plisa los dientes y se evidencia que la señal presenta más vibraciones, mientras que cuando no se da este movimiento la señal es más estable.

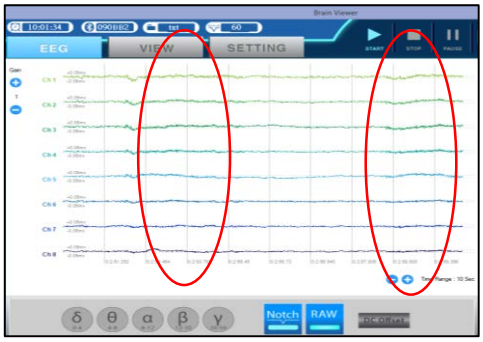


Figura 9: Señales bioeléctricas apretando los dientes

6.2 Ritmos cerebrales *theta* y *beta* en la señal EEG con base al test de Stroop

Para verificar el funcionamiento del sensor EEG BR8 Plus se realizaron dos *test* de Stroop a una persona sana, y que no presenta problemas de atención para la ejecución de la tarea. Se registraron sus señales cerebrales en dos sesiones: una a las 11:24 a.m. y otra a las 16:04 p.m. En la Figura 10 se observa el arreglo instrumental con el que se llevó a cabo la prueba de Stroop, una pantalla donde se observa la tarea visual y un computador en el que se registraron las señales en la interfaz de usuario del sensor.

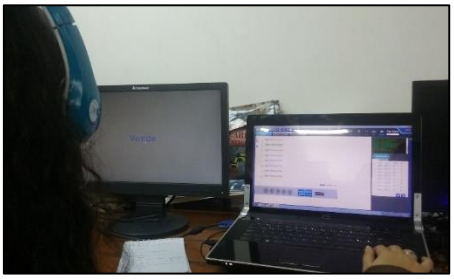


Figura 10: Usuario realizando la prueba de Stroop mientras se registra sus señales EEG

La primera prueba tuvo una duración de 70 segundos aproximadamente, donde solamente cometió un error. La prueba entrega una matriz 4x9, en la que la primera fila es la palabra presentada, la segunda es el color de la palabra, la tercera es la respuesta del usuario y la última fila es el tiempo de respuesta que se da entre cada palabra medido en segundos. Esta prueba se evalúa con las flechas del teclado, donde la izquierda hace referencia al color rojo, la hacia abajo al color verde y la derecha al color azul.

Tabla 1: Matriz respuesta de la Prueba de Stroop #1. Las casillas verdes hacen referencia a que la respuesta del usuario fue correcta y la casilla rosa a que su respuesta fue incorrecta

3	2	1	1	3	2	3	2	2
2	2	2	3	3	2	1	1	3
2	2	2	3	3	2	1	1	3
3,4811	0,6600	0,6688	1,3719	0,6689	0,7193	0,6189	0,5688	1,1878
3	1	1	1	3	3	3	1	2
1	2	3	3	1	3	3	1	1
1	2	3	3	1	3	3	1	1
0,7192	0,6691	0,9703	0,8030	0,6693	1,3214	0,7194	0,5354	0,6190
1	1	2	2	3	1	2	2	3
2	1	2	3	2	3	2	1	3
2	1	2	3	2	3	2	1	2
0,6190	0,6357	0,8364	0,7360	0,9702	1,3884	0,7695	0,8197	1,0705
2	2	1	1	1	3	3	3	2
3	1	1	2	1	2	1	2	3
3	1	1	2	1	2	1	2	3
1,255	0,803	0,485	0,954	0,602	0,819	0,569	1,255	0,719

Los resultados son números del uno al tres, siendo uno el color rojo, dos el verde y tres el azul, tal como se observa en la Tabla 1. Es decir, si la palabra escrita es Azul, en la primera fila aparecerá un 3, en la segunda fila un 2, en la tercera fila se mostrará la respuesta del usuario y en la última se presentará el tiempo que tarde en escoger el color.

Luego de filtrar la señal de cada época de cada canal, se buscó que en el espectro de frecuencias de la señal filtrada apareciera un pico (el más alto), dentro del rango de frecuencias de cada ritmo (*theta* y *beta*). También se observó la amplitud dada en mili voltios de la señal filtrada. La Tabla 2 hace referencia a un compendio de todos los datos arrojados por la prueba, tanto para la amplitud como para las frecuencias, con su respectiva desviación estándar, demostrando que todos los picos se encontraron en el rango de frecuencias que se especifica en la literatura.

Tabla 2: Promedios de amplitud y frecuencia de los ritmos encontrados de la prueba de Stroop #1

	Promedios			
	Theta		Beta	
	Amplitud [mV]	Frecuencia [Hz]	Amplitud [mV]	Frecuencia [Hz]
Ch1	0,04 ± 0,013	5,49 ± 0,691	0,039 ± 0,007	16,798 ± 3,181
Ch2	0,042 ± 0,015	5,666 ± 0,736	0,041 ± 0,007	16,606 ± 2,922
Ch6	0,107 ± 0,034	5,506 ± 0,838		
Ch7			0,088 ± 0,024	17,442 ± 2,948
Ch8			0,09 ± 0,017	15,231 ± 4,440

La segunda prueba tuvo una duración de 76 segundos aproximadamente y no se cometió ningún error, como se evidencia en la Tabla 3.

Tabla 3: Matriz respuesta 4x36 de la prueba de Stroop #2

3	3	1	2	2	1	3	1	1
2	2	3	2	2	2	1	3	1
2	2	3	2	2	2	1	3	1
1,3566	0,5642	1,1037	0,6856	0,5015	0,6523	0,9533	1,1371	0,6190
1	1	1	1	1	2	3	1	3
3	2	3	2	1	1	2	2	3
3	2	3	2	1	1	2	2	3
0,9867	0,8363	1,0374	0,8026	0,9532	0,8191	1,1204	0,9030	1,1877
1	1	2	3	3	2	2	3	3
1	1	1	1	3	2	3	1	1
1	1	1	1	3	2	3	1	1
0,753	0,736	0,636	0,569	1,220	0,836	1,472	0,769	1,271
2	2	2	3	3	2	2	3	2
2	3	3	3	2	3	1	3	1
2	3	3	3	2	3	1	3	1
1,422	0,870	0,853	0,886	0,568	0,869	0,719	1,823	1,806

Como se observa en la Tabla 4, se encontraron todos los picos justo en el rango de frecuencias indicado, lo que demuestra que se encontraron los ritmos.

Tabla 4: Amplitud y frecuencia de los ritmos encontrados de la prueba de Stroop #2

	Promedios			
	Theta		Beta	
	Amplitud [mV]	Frecuencia [Hz]	Amplitud [mV]	Frecuencia [Hz]
Ch1	0,023 ± 0,015	5,542 ± 0,810	0,021 ± 0,006	16,911 ± 3,537
Ch2	0,024 ± 0,014	5,384 ± 0,861	0,026 ± 0,010	16,694 ± 2,575
Ch6	0,145 ± 0,016	5,494 ± 0,920		
Ch7			0,074 ± 0,007	15,238 ± 1,856
Ch8			0,065 ± 0,009	15,606 ± 2,254

Posteriormente, se calcularon los ritmos para las dos pruebas utilizando un muestreo del 10% de las épocas para cada canal involucrado, por lo tanto, se procesaron 3 épocas por cada canal: 10, 20 y 30, con el fin de mostrar que siempre se debe dar la aparición de todos los ritmos cerebrales. En la Figura 11 se presenta los cinco ritmos con sus respectivas amplitudes y sus componentes frecuenciales para los electrodos frontales de la prueba número 1. En la Figura 11A los ritmos para prueba 1, canal 1, época 10 y en la Figura 11B los ritmos para la prueba 1, canal 2, época 10. En ellas se puede notar que las amplitudes de los ritmos *beta* y *theta* para cada canal son más prominentes y aparecen de forma cuasi-simétrica. La descripción en frecuencia de los ritmos obtenidos en las dos pruebas está dentro del rango reportado en la literatura. Se muestran en la parte derecha de las figuras la señal de la época analizada y su respectiva densidad espectral de potencia con todos sus componentes frecuenciales. Finalmente, se evidencia la parte oscilatoria de los ritmos procesados, notándose el corrimiento frecuencial dentro del rango de definición de las diferentes ondas analizadas.

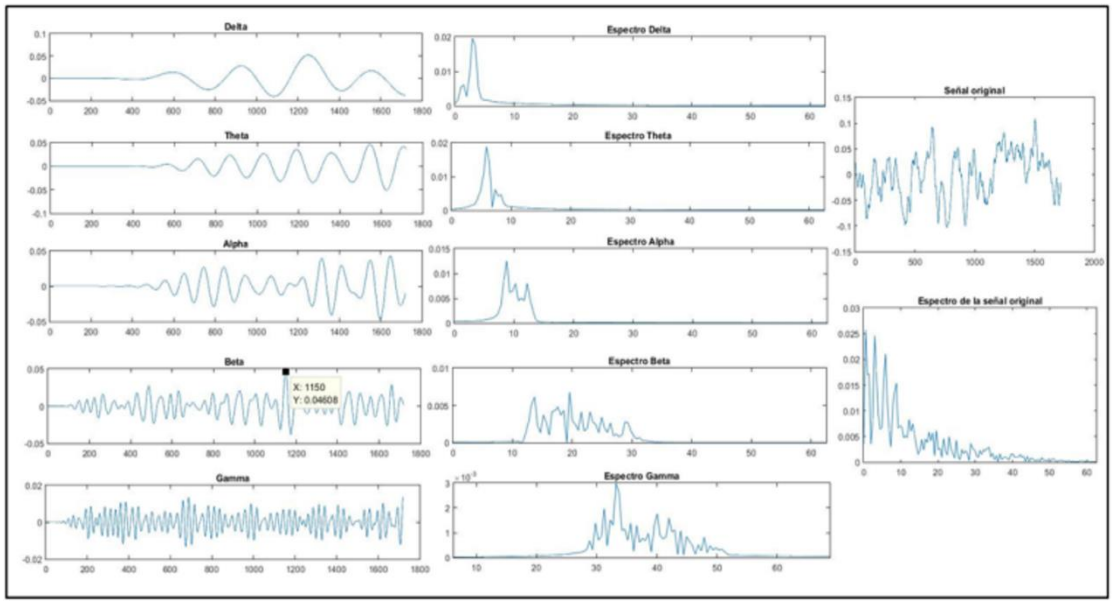


Figura 11A: Señales de prueba 1, canal 1, época 10

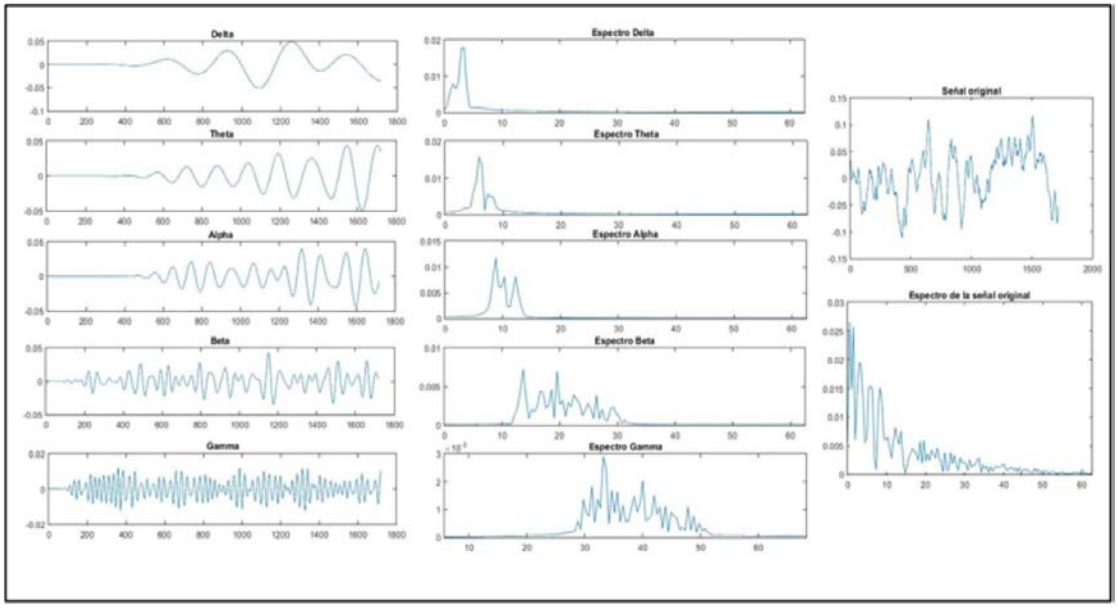


Figura 11B: Señales de prueba 1, canal 2, época 10

7 Discusión y Conclusiones

Desde la época de Hubel y Wisel [15], quienes realizaron los primeros experimentos con estímulos visuales, se evidenció que los estímulos con diversas orientaciones podrían activar de manera diferenciada los centros visuales occipitales. Más tarde, hacia finales del siglo XX, se encontró que las ondas lentas tipo *beta* podrían modular el transporte de señal visual desde los ojos a través del tálamo y hacia la corteza occipital [16]. Se ha observado que en estado consiente el ritmo *beta* desplaza el sistema visual a estados atencionales, que posibilitan la sincronización de los estados de alta frecuencia o ritmos *gamma* y por ende a un estado de percepción enfocada. El profesor Llinás [17, 18] ha demostrado que los procesos de sincronización están mediados por aprestamientos motores (ritmos *theta*) que enfocan la atención en la corteza prefrontal (ritmos *beta*),

mediando un estado atencional de sincronía y coherencia altas. Finalmente, en la Figura 11 se muestra un patrón de ritmo *gamma* altamente sincronizado en los canales prefrontales de la primera prueba de Stroop, realizada por uno de los voluntarios.

Como se ha expuesto a lo largo de este trabajo, el sensor EEG BR8 Plus es una herramienta en desarrollo altamente novedosa, con la que se ha logrado probar la captura de señales electroencefalográficas con tareas muy básicas, específicamente tareas de ojos abiertos, pestañeo, ojos cerrados y ceñir de dientes. Por lo que se pudo comprobar que estos resultados tuvieron un correcto funcionamiento de acuerdo con los obtenidos por la compañía Brain Rhythm Incorporation. Además, se implementó una mini prueba Stroop de atención visual. Con ello se pudo encontrar los ritmos *theta* y *beta* en las señales registradas en dos sesiones satisfactorias, lo que comprueba que el sensor capta bien los ritmos cerebrales.

Aparte de esto, no se pudo evidenciar ningún cambio en el estímulo 27 (Ver columna rosa, Tabla 1), cuando el usuario cometió el error, puesto que en la mayor parte de la prueba las respuestas fueron correctas. Pero se obtuvo el patrón de calibración diseñado por la compañía y los ritmos altamente importantes para la prueba de atención visual.

Referencias

[1] Brain Rhythm Inc. BR8 PLUS. http://www.bri.com.tw/product.html.

[2] Lin, C. et al. (2011). Novel dry polymer foam electrodes for long-term EEG measurement. Transactions on Biomedical Engineering 58(5), pp. 1200-1207.

[3] Liao, L. et al. (2012). Biosensor technologies for augmented brain–computer interfaces in the next decades. Proceedings of the IEEE 100 - Special Centennial Issue, pp. 1553-1566.

[4] Buzsáki, G. & Watson B. (2012). Brain rhythms and neural syntax: implications for efficient coding of cognitive content and neuropsychiatric disease. Dialogues Clin Neurosic 14(2), pp. 345-367.

[5] Sanei, S. & Chambers, J. (2007). EEG Signal Processing. UK: John Wiley & Sons.

[6] Ashwal, S. & Rust, R. (2003). Child neurology in the 20th century. Pedia. Res. 53(2), pp. 345–361.

[7] Niedermeyer, E. (2005). Electroencephalography, basic principles, clinical applications, and related fields. Philadelphia: Lippincott Williams&Wilki.

[8] Pfurtscheller, G., Flotzinger, D. & Neuper, C. (1994). Differentiation between finger, toe and tongue movement in man based on 40 Hz EEG. Electroencephalogr. Clin. Neurophysiol 90, pp. 456–460.

[9] Ardila, A. & Ostrosky, F. (2012). Guía para el diagnóstico neuropsicológico. Ardila & Ostrosky.

[10] Stroop Experiment Demo. Online [Jul. 2014].

[11] Qué es el test de Stroop y su utilidad en la evaluación del TDAH. Online [Jul. 2014].

[12] Ardila, A. & Ostrosky, F. (2007). Neuropsi attention and memory: a neuropsychological test battery in Spanish with norms by age and educational level. Applied Neuropsychology 14(3), pp. 156–170.

[13] Silva, J. (2011) Métodos en neurociencias cognoscitivas. México: Manual Moderno.

[14] Delorme, A. & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. Journal of Neuroscience Methods 134, pp. 9-21.

[15] Hubel, D. & Wiesel, T. (2005). Brain and visual perception: The story of a 25-year collaboration. Oxford: Oxford University Press.

[16] Wróbel, A. (2000). Beta activity: A carrier for visual attention. Acta neurobiologiae experimentalis 60(2), pp. 247-260.

[17] Llinás, R. & Steriade, M. (2006). Bursting of thalamic neurons and states of vigilance. Journal of neurophysiology 95(6), pp. 3297-3308.

[18] Llinás, R. (2002). I of the vortex: From neurons to self. USA: MIT press.

Uso de tecnologías emergentes para el monitoreo de tráfico vehicular

Daniel Silva
Johan Henao
César Pedraza
Félix Vega
Universidad Nacional de Colombia – Colombia

1 Introducción

Debido al rápido crecimiento en la demanda en el transporte que se presenta en las ciudades, se hace necesario implementar diferentes métodos y sistemas de adquisición y procesamiento de información, para desarrollar soluciones que permitan mitigar los efectos negativos que conlleva el alto número de vehículos en circulación. Se ha propuesto diferentes sistemas de medición y monitoreo para realizar esta tarea, de forma que se adapte a las necesidades de cada escenario. Actualmente, una de las problemáticas es la medición de las variables que se presentan en el tráfico urbano, ya que permiten establecer las conductas que se exhiben en los escenarios que componen el tráfico vehicular, y por tanto, esta toma de información debe ser confiable, precisa y que esté siempre disponible para su posterior análisis.

Para obtener esta información se emplean diferentes tecnologías, con las cuales se recogen datos tales como velocidades, flujos vehiculares, conteo de vehículos, etc., que pueden proporcionar información acerca de las conductas en el tráfico. Gracias a esto se establecen patrones y predicciones para ofrecer soluciones a los retos que se presentan en el transporte. El empleo de la tecnología de Identificación por Radio Frecuencia (RFID) y de dispositivos Bluetooth es una alternativa para la adquisición de esta información en diferentes campos. Estas tecnologías ofrecen soluciones tales como el control de inventario, las comunicaciones, la logística, entre otras, en los diferentes campos económicos. De esta forma es posible definir que los sistemas RFID y Bluetooth son una opción aceptable para tener en cuenta en el diseño de sistemas de monitoreo vehicular inteligente [1, 2].

El monitoreo vehicular inteligente hace parte de una estrategia para mejorar el flujo de vehículos a partir de la adquisición de información relevante para el proceso y toma de decisiones en tiempo real. Sin embargo, existe el problema de la adquisición de la misma y de la definición de las condiciones y requerimientos de un sistema que integre tecnologías para el monitoreo de tráfico de vehículos en una ciudad, así como el reconocimiento de los factores que posee la ciudad en la que va a operar. El proyecto de Plataforma de Control Vehicular Inteligente (PCIV), desarrollado en la Universidad Nacional de Colombia, es parte de un esfuerzo por mejorar los sistemas de monitoreo para el control de tráfico, mediante la integración de las tecnologías RFID y Bluetooth. Se pretende brindar una solución al problema en la identificación de requerimientos para este tipo de aplicaciones.

El presente trabajo expone dos soluciones diseñadas para la aplicación de estas tecnologías en sistemas de monitoreo de tráfico vehicular, teniendo en cuenta requisitos tales como soporte para intercambio de datos, filtrado de información, interpretación de datos, escalabilidad, confiabilidad, entre otros. De este modo se presenta la forma en que deben ser validadas las tecnologías empleadas a fin de dar solución efectiva al problema. Para la validación del experimento propuesto se implementó y probó una plataforma para

monitoreo vehicular, que cual realiza conteo de vehículos, mediciones de velocidades medias, ocupación de parqueaderos, entre otras. Este experimento se realizó en el campus de la Universidad Nacional de Colombia Sede Bogotá.

2 Tecnologías emergentes en los Sistemas Inteligentes de Transporte

Los Sistemas Inteligentes de Transporte (ITS) parten del crecimiento de la demanda de transporte que se presentan en la sociedad [3], y es la base para el desarrollo de ciudades inteligentes en este sector. Proveen herramientas tecnológicas para el desarrollo de soluciones que permitan el mejoramiento en los diferentes campos que conforman esta área en las ciudades. Los ITS se emplean en diferentes áreas del transporte: información al viajero, administración del tráfico, sistemas de seguridad vehicular, peajes electrónicos, entre otras. La principal innovación es que integran las tecnologías existentes por medio de la creación de nuevos servicios [4], que permiten obtener información de las diferentes variables implicadas en el tráfico urbano. De acuerdo con Garbolino et al. [4], los retos que afrontan estos sistemas son:

- Reducción de la congestión que se presenta en las vías, mejorando la administración demanda-capacidad del sistema de transporte
- Mejoramiento de la seguridad en el transporte, tanto en los vehículos como en las vías, adecuando sistemas de rápida respuesta en caso de una emergencia
- Reducción de las emisiones de CO_2, realizando una mejor administración en los servicios de transporte (flota de vehículos, vías, etc.)

La implementación de los ITS ha permitido desarrollar nuevos procedimientos de administración y supervisión, gracias a las diferentes tecnologías que pueden proporcionar información de las variables en el tráfico urbano y rural, tales como estado de las vías, rutas, velocidades medias, etc. Además, emplean tecnologías de corto alcance (1-100 m) para la recopilación de información, entre las que se encuentra Bluetooth [1, 5], que se emplea para realizar mediciones de flujo vehicular [6] y tiempos aproximados de desplazamiento. Bhaskar y Chung [7] exponen cómo determinar el tiempo de viaje de un vehículo que pasa por dos puntos geográficos, donde se ubican dos lectores tipo Bluetooth que interrogan la dirección MAC (Medium Access Control) de los dispositivos que lleve el usuario en el vehículo (Tabletas, PC, Teléfonos etc.), o cómo los sistemas de red de comunicaciones permite compartir información con vehículos que estén alrededor [8]. Otra de las tecnologías utilizadas son los sensores inductivos ubicados en las vías para realizar el conteo de vehículos que pasan por unidad de tiempo [6], y determinar el flujo y el tiempo de viaje de recorrido para evaluar el comportamiento dinámico del sistema de transporte. También se realiza detección vehicular a través de los teléfonos inteligentes [9], recopilación de información que se realiza a través del hardware integrado en el dispositivo, tales como Global Positioning Satélite (GPS), acelerómetros, Bluetooth, etc. En este caso, el dispositivo móvil envía la información recopilada a los servicios para su posterior procesamiento.

En la literatura se propone arquitecturas con diferentes tipos de tecnologías para obtener información del tráfico, es el caso de los Wireless Sensors Networks (WSN) [10, 11], y se proponen los requerimientos que debe llevar para una aplicación ITS, tales como seguridad, codificación, sistemas de validación, entre otras. Los sistemas RFID (Radio Frequency IDentification) [12] permiten realizar la identificación, seguimiento e inventario de los vehículos en zonas urbanas, ofreciendo diferentes *frameworks* y arquitecturas [12, 13] para las especificaciones que exigen las diferentes aplicaciones de los ITS. Autores como Zongwei [2], Zhiyang [14] y Bhaskar [6] han experimentado con

diferentes tecnologías para realizar la recolección de información relacionada con el tráfico en la ciudades. Hu [15], Jing [16] y otros han propuesto diferentes *frameworks* para definir el diseño y la implementación de sistemas aplicados al monitoreo del tráfico, con cualquiera de estas tecnologías. La definición de plataformas basadas en tecnología RFID se aplica en diferentes áreas de la economía, tales como la industria, la educación, la logística, la salud, entre otras. En general, las arquitecturas plateadas para los sistemas RFID pueden ser divididas en tres partes [17].

- Adquisición de la información, que se realiza a través de un conjunto de etiqueta pasiva o activa (Tag) y un lector RFID. Por lo general, para el área del transporte vehicular los lectores son ubicados al lado de las vías, donde capturan el Código de Producto Electrónico (EPC) o la información contenida en la memoria del Tag.
- Administración de la información, que realiza un procesador (por lo general por un sistema embebido), que se encarga de filtrar, organizar y enviar la información a los diferentes servicios o aplicaciones que requieren dicha información.
- Administración organizacional, que corresponde a los servicios que realizan las diferentes tareas en la manipulación de la información. Estos servicios se encuentran alojados en un servidor (Nube) y están en constante comunicación con los lectores para el envío o recepción de información.

Por su parte, Uckelmann [18] ha investigado acerca del impacto de la tecnología RFID en el área de la logística, desde una perspectiva económica. Este autor concluye que los sistemas RFID son eficientes y son una adecuada inversión para las compañías de logística. Otros trabajos se enfocan en la definición de *frameworks* y arquitecturas basadas en tecnologías RFID en diferentes áreas de aplicación. Amaral [19] define uno proponiendo algunas condiciones que debe cumplir una plataforma RFID para ser aplicada en cualquier área. Estas restricciones involucran seguridad, interoperabilidad, escalabilidad,…

Aunque se encuentra una cantidad representativa de investigaciones en tecnologías RFID, solamente algunos trabajos exponen claramente requerimientos y restricciones para su empleo en el monitoreo del tráfico vehicular. Chattaraj [20] presenta a RFID como una opción para el monitoreo en las intersecciones de las vías, incrementando la seguridad y previniendo posibles accidentes de tránsito. Aunque no hace una descripción detallada del *framework* propuesto, presenta las condiciones técnicas y de operación para este tipo de aplicaciones. Kim [21] presenta los requerimientos en red para el uso de la tecnología RFID para el monitoreo del tráfico. Este trabajo muestra algunas recomendaciones con respecto a la red y el RFID, tales como el empleo de Tags activos a una frecuencia de 2.45 GHZ. De la misma forma, realiza una diferenciación en tres principales subsistemas para las aplicaciones en el monitoreo del tráfico: un subsistema de hardware y software que se encarga de la lectura de los Tags y el procesamiento de la información, un subsistema de transmisión y almacenamiento en un servidor remoto y subsistema de aplicación que se encarga de tomar las decisiones. Estos subsistemas son semejantes a la arquitectura que propone Fosso [17].

3 Arquitecturas propuestas

3.1 Arquitecturas RFID y Bluetooth

Para la definición de la arquitectura RFID se propone la instalación de un tag pasivo en los vehículos, posteriormente el vehículo pasa cerca de un lector RFID y este obtiene la información desde el tag. Al validar que el tag pertenece al sistema, la información es enviada al servidor (nube) donde se hace el procesamiento de la misma (Figura 1).

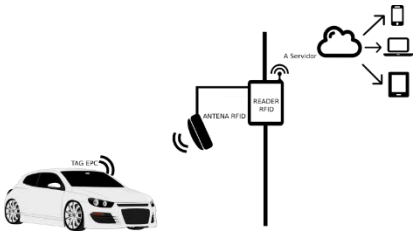


Figura 1: Esquema empleado con la tecnología RFID para el monitoreo vehicular

Algunos requerimientos para la arquitectura del sistema son:

- Filtro de Tags. Se emplea filtrado de la información porque el sistema puede detectar tags que no pertenecen al sistema y generar mediciones erróneas.
- Escalabilidad. La arquitectura diseñada debe permitir el crecimiento del sistema, por ejemplo, más lectores y más tags, permitiendo el procesamiento de grandes cantidades de información (Big Data) sin perder sus características básicas.
- Configuración del equipo. Los parámetros del sistema se pueden cambiar en cualquier momento, tales como cantidad de antenas, posición geográfica, id del equipo, etc.
- Interpretación de la información. El sistema debe ser capaz de leer cualquier tipo de información que está contenida en el tag (Código de Producto Electrónico (EPC), información contenida en la memoria, etc.)
- Soporte para compartir la información. El sistema debe ser capaz de compartir la información con otros lectores o con diferentes plataformas, con el objetivo de que los diferentes usuarios o administradores del sistema puedan tener acceso a la información.

De la misma forma se plantea una opción empleado tecnología Bluetooth pero partiendo de la solución anterior. En este caso se reemplaza la tecnología de Radio Frecuencia por tecnología Bluetooth para la adquisición de la información del tráfico (Figura 2). Se plantea el uso de esta tecnología como una solución a los diferentes problemas que se presentan en los sistemas de transporte.

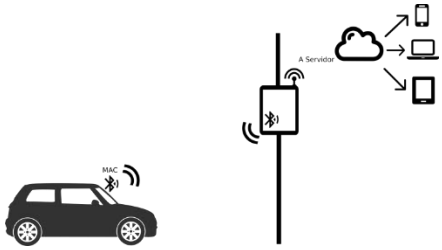


Figura 2: Esquema de lectura por medio de tecnología Bluetooth

Características como seguridad, escalabilidad, confiabilidad de la información y adaptabilidad hacen parte de la solución propuesta con esta tecnología. Para realizar la estructura del dato se cambia el parámetro EPC y Dato del tag por la dirección MAC (Medium Access Control) del dispositivo Bluetooth; también se realiza un filtrado de la información porque cualquier dispositivo Bluetooth puede ser leído por los lectores, lo que se realiza haciendo un filtrado por clase del dispositivo.

Con estos requisitos se plantea una arquitectura para el empleo de tecnología RFID, como se representa en la Figura 3. Se observan dos partes principales: el subsistema RFID

y subsistema del Software. Este último se encarga de procesar la información y almacenarla para que los usuarios finales la tengan a disposición.

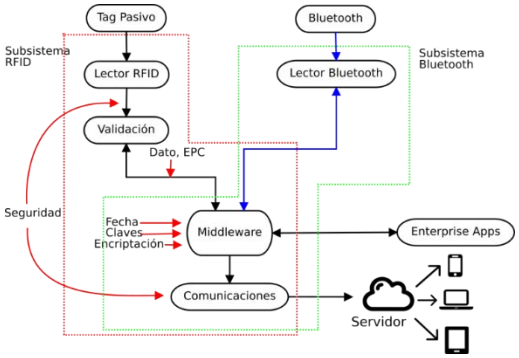


Figura 3: Arquitectura RFID y Bluetooth propuesta

Subsistema RFID. El subsistema RFID incluye un lector RFID con sus respectivas antenas, un sistema embebido que realiza la validación y encriptación de la información y un sistema de comunicaciones que se encarga de enviar la información a la plataforma. Una de las tareas más relevantes es preparar la información que será enviada al servidor remoto (Subsistema Software). Esta tarea se encarga de la preparación y encriptación del dato para lograr un estándar de seguridad y que pueda ser enviado al servidor. Además, se agregan otros campos de información que son relevantes para el procesamiento y que se relaciona con el trafico (Figura 4). Finalmente, para lograr la comunicación de los lectores con la plataforma se emplea la red GSM, que ofrecen los distintos operadores de telefonía móvil.

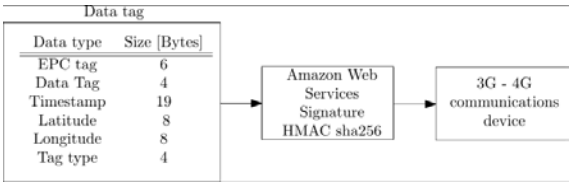


Figura 4: Estructura del dato del tag, encriptación y comunicación al servidor

Subsistema Bluetooth. Este subsistema emplea un dispositivo Bluetooth para realizar la interrogación de las direcciones MAC de los dispositivos instalados o que se encuentren dentro del vehículo. De igual manera cuenta con un sistema embebido que se encarga de realizar la preparación y encriptación de la información. En esta etapa de preparación se agregan los diferentes campos de información (posición geográfica, hora de lectura, hora del sistema, etc.) para el envío a la plataforma. En la etapa de encriptación se agregan las firmas y las claves para evitar el envío de información por parte de impostores. Finalmente, cuenta con un módulo de comunicaciones que se encarga de enviar la información a la plataforma.

Subsistema Software. El sistema de información fue desarrollado como una aplicación web usando la arquitectura cliente servidor, fue pensada de forma que pueda ser escalable, flexible, de gran capacidad de procesamiento de información y que permita integración con otros servicios que ofrecen las diferentes plataformas en la nube. Este subsistema se utilizó con los dos tipos de tecnologías y consta de tres módulos:

- Módulo de estadísticas. Calcula los volúmenes de vehículos, matrices origen-destino, histogramas, zona de ocupación, velocidades, etc. La información se encuentra disponible de forma que el usuario pueda tomar las acciones pertinentes con respecto a la movilidad.

- Módulo de Administración. Controla el número de lectores, las claves de acceso, los permisos de usuario, etc. Adicionalmente, monitorea el estado de los lectores en tiempo real y permite el reseteo de los lectores de forma remota.
- Base de datos .Recibe, valida y guarda toda la información y estado de los lectores. Además, provee un servicio web que interroga las lecturas con los parámetros que debe contener la trama de la información (Figura 4).

Seguridad. Para proteger la información enviada desde los lectores y evitar intrusiones a la plataforma se emplea un protocolo HTTPS en todos los sistemas de comunicaciones, tanto en los lectores como en los sistemas externos. Para garantizar la integridad durante el cambio de información, los lectores son autenticados por la aplicación usando un protocolo basado con una firma de Amazon Web Service Version 4. Con él se garantiza que la información enviada pertenece a los lectores registrados a la plataforma y no de un impostor.

4 Metodología

4.1 Tecnología RFID

El experimento fue realizado en el campus de la Universidad Nacional de Colombia-Sede Bogotá. En cada uno de los accesos vehiculares se instaló un lector con tecnología RFID (Figura 5) que realiza la toma de información desde los diferentes tipos de vehículos (Motos, carros, camionetas, etc.) que tenían la etiqueta instalada (Figura 6). Se contó con la participación de los miembros de la comunidad universitaria que ingresaban frecuentemente al campus en un vehículo. Para la verificación del sistema RFID se instalaron aproximadamente 700 etiquetas en los diferentes vehículos, esta muestra es representativa porque al campus ingresan y salen aproximadamente 8.000 vehículos.

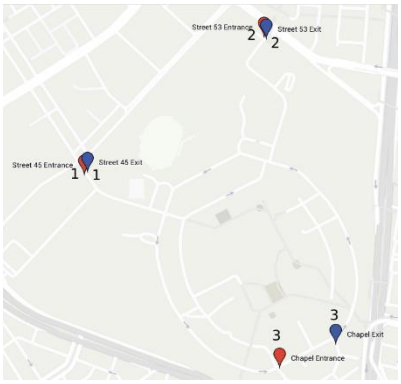


Figura 5: Lectores RFID instalados en los accesos vehiculares

Figura 6: Instalación del tag en la moto de un participante del proyecto

En la Figura 7 se observa un lector instalado en uno de los accesos vehiculares del campus de la universidad. Esta unidad cuenta con una configuración de 2 antenas y realiza la interrogación de los Tags a una distancia aproximada de 4 metros.

Figura 7: Equipo instalado en un acceso vehicular

4.2 Tecnología Bluetooth

En esta parte del experimento se instalaron dos lectores con tecnología Bluetooth sobre la avenida 26 (Figura 8). Aquí los dos lectores están con una separación de 835 m aproximadamente, se realiza la interrogación de la dirección MAC de los dispositivos Bluetooth que están instalados en los vehículos y después de obtener la dirección MAC se procede a hacer el cálculo de la velocidad promedio del vehículo al transitar este tramo de avenida.

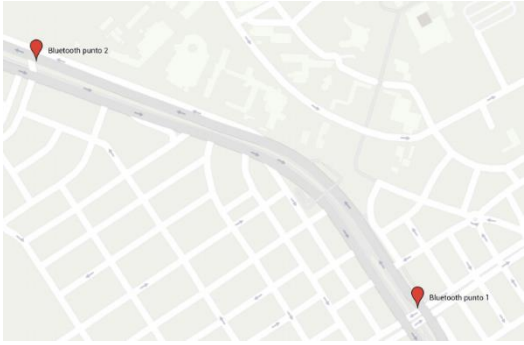


Figura 8: Posición geográfica de los lectores Bluetooth en inmediaciones del campus

Al tomar la dirección MAC del dispositivo se garantiza que la información, tanto del usuario como del vehículo, sea transparente a la plataforma, porque esta dirección no tiene ninguna relación con el conductor o el vehículo. Por tanto, se garantiza que la obtención de la información no vulnera la privacidad de la información del conductor ni del vehículo. Para la validación del experimento se realizó un conteo manual de los vehículos que circulaban por el corredor vial, solamente se efectuó el conteo de vehículos particulares. Finalmente, las lecturas se realizaron durante una hora diaria en el lapso de tres días.

5 Resultados

Para el lector RFID se exponen los resultados que se obtuvieron de los equipos instalados en el campus (Figura 5). El conteo manual se realizó el 8/10/2014 y la información de la plataforma corresponde a la misma fecha. En la Figura 9 se observan los dos tipos de lecturas que se realizaron ese día, conteo manual y lecturas realizadas por los lectores. En la Tabla 1 se describe el porcentaje de error que hubo durante las lecturas realizadas en esa fecha. También, y gracias a la plataforma, se obtuvo mediciones de volúmenes vehiculares (Figura 10) y de la ocupación de los parqueaderos del campus (Figura 11), y para realizar las mediciones de ocupación de parqueaderos se desarrolló un lector portable con un sistema GPS, que permitió obtener la posición geográfica aproximada en donde estaba parqueado el vehículo que tenía instalado el tag.

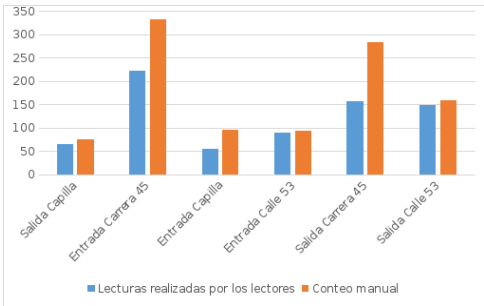


Figura 9: Comparación de las lecturas realizadas

Tabla 1: Tabla de errores de lecturas de los accesos vehiculares del campus

Accesos Vehiculares	Error porcentual (%)
Salida Capilla	13.33
Entrada Carrera 45	33.13
Entrada Capilla	42.10
Entrada Calle 53	5.37
Salida Carrera 45	44.52
Salida Calle 53	5.69

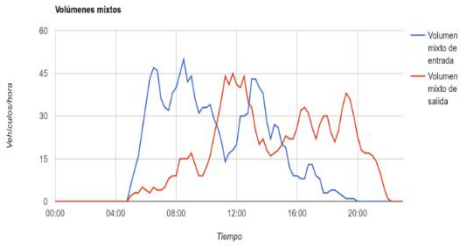


Figura 10: Mediciones de volúmenes mixtos en el campus

Figura 11: Ocupación de parqueaderos campus

Los resultados que se obtuvieron en las mediciones de velocidad con la tecnología Bluetooth se muestran en la Tabla 2.

Tabla 2: Tabla de mediciones de velocidad con el lector Bluetooth

Promedio de vehículos por día	Número de vehículos escaneados	Velocidad promedio (km/h)
1567	33	49.01
1536	17	59.75
1668	35	42.58

6 Análisis de resultados

En algunos de los accesos vehiculares del campus se presentan más lecturas manuales que las realizadas por los lectores RFID (Figura 9), esto se debe a factores como la ubicación de las antenas con respecto al tag instalado en los vehículos. Debido a la variedad de vehículos que ingresan y salen no hay un patrón para la orientación de las antenas en los accesos vehiculares (Figura 7). En la ubicación de los tags en los vehículos (Figura 6) se presenta otro factor de error, debido al que debe estar ubicado en una zona del vehículo donde pueda recibir la mayor cantidad de energía irradiada por la antena del lector. Otro factor a tener en cuenta es el tipo de polarización que tienen las antenas del lector y las del tag, porque es un factor determinante a la hora de enviar la señal de interrogación y de lectura de la información por parte del lector.

En la Tabla 2 se observa que la muestra es baja con respecto a la cantidad de vehículos contados. Esto se debe a los diferentes factores que influyen en la medición, por ejemplo, la cantidad de vehículos que tienen instalado el dispositivo Bluetooth. Hay que observar que no todos los vehículos que tienen instalado el dispositivo lo mantienen encendido.

7 Conclusiones

- Mediante la implementación de dos sistemas de identificación de vehículos, basados en RFID y Bluetooth, se comprobó que las arquitecturas propuestas permiten la adquisición de información referente al tráfico vehicular de manera automática y fiable. Gracias a parámetros como escalabilidad y seguridad, es posible implementar estas opciones para el desarrollo de soluciones en el monitoreo del tráfico vehicular.
- Determinar la posición de un vehículo en un lugar como un parqueadero es un factor relevante en términos de planeación y distribución urbanística, generando nuevos elementos para formular planes de ordenamiento territorial.
- La implementación de tecnologías de bajo costo para los sistemas inteligentes de transporte abre nuevas posibilidades en la adquisición y levantamiento de información de manera autónoma, controlada y organizada.
- Para un óptimo funcionamiento de cualquier tecnología es necesario evaluar las posibles configuraciones de trabajo y ubicación de los equipos, porque pueden influir en las estadísticas finales.
- Cabe resaltar que el escaneo que efectúa el lector Bluetooth se realiza en aproximadamente 8s, por tanto, es posible que algunos vehículos no puedan ser escaneados debido a que en el corredor vial circulan a altas velocidades.

Agradecimientos

Este trabajo fue desarrollado por la Universidad Nacional de Colombia – Sede Bogotá y Datatools S.A., dentro del proyecto de investigación "Plataforma para el Control Inteligente de Vehículos PCIV" (HERMES 20097), el cual es financiado por el Ministerio de Transporte a través del contrato 06554-2013 y otorgado mediante la convocatoria 622-2013 de Colciencias.

Referencias

[1] Friesen, M. & McLeod, R. (2015). Bluetooth in intelligent transportation system: A survey. International Journal of Intelligent Transportation System 13(3), pp. 143-153.

[2] Luo, Z.; Zhang, T. & Wang, C. (2011). RFID enabled vehicular wireless query for Travel Information in Intelligent Transportation System. Proceedings IEEE International Conference on In RFID-Technologies and Applications, pp. 415-420.

[3] An, S.; Lee, H. & Shin, R. (2011). A survey of intelligent transportation systems. Proceedings Third International Conference on Computational Intelligence, Communication Systems and Networks, pp. 332-337.
[4] Garbolino, E. et al. (2012). Transport of dangerous goods. Canada: The NATO Science for Peace and Security Programme.
[5] Guadarrama, R.; Rodríguez, M. & Salcedo, A. (2010). A Bluetooth development platform for wireless sensors. Proceedings Conference Electronics, Robotics and Automotive Mechanics, pp. 332-337.
[6] Bhaskar, A; Qu, M. & Chung, M. (2014) Bluetooth vehicle trajectory by fusing Bluetooth and loops. IEEE Transactions on Intelligent Transportation Systems 16(1), pp. 113-122.
[7] Bhaskar, A. & Chung, M. (2013). Fundamental understanding on the use of Bluetooth scanner as a complementary transport data. Transportation research Part C: Emerging Technologies 37, pp. 42-72.
[8] Sawant, H. et al. (2004) Using Bluetooth and sensor network for intelligent transportation system. Proceedings 7th international IEEE conference on Intelligent Transportation Systems, pp. 767-772.
[9] Xia, H. et al. (2014). Using smart phone sensors to detect transportation modes. Sensors 14(11), pp. 20843-20865.
[10] Khanafer, M.; Guennoun, M. & Mouftah, H. (2009). WSN architectures for intelligent transportation systems. Proceedings 3rd International conference on Mobility and Security, pp. 1-8.
[11] Katiyar, V.; Kumar, P. & Chand, N. (2011). An intelligent transportation systems architecture using Wireless sensor networks. International Journal of Computer Application 14(2), pp. 22-26.
[12] Jrjung, L. et al. (2014). A study of rfid readiness framework and its application. International journal of Electronic Business Management 12(2), pp. 112-121.
[13] Xiang, S. et al. (2014). A hybrid positioning strategy for vehicles in a tunnel base don RFID and in vehicle sensors. Sensors 14(12), pp. 23095-23118
[14] Zhiyang, X. & Pengfei, S. (2003). Wireless location determination for mobile objects based on GSM in intelligent transportation systems. Journal of Systems Engineering and Electronics 14(2), pp. 8-13.
[15] Hu, J. & Li, G. (2006, June). Design of city-bus intelligent control system framework. Proceedings 2006 IEEE International Conference on Mechatronics and Automation, pp. 2307-2311.
[16] Jing, G. et al. (2006). Design of an intelligent transportation system based on GPS and GPRS. Proceedings International Conference on Wireless, Mobile and Multimedia Networks, pp. 1-3.
[17] Wamba, S.; Anand, A. & Carter, L. (2013). A literature review of RFID-enabled healthcare applications and issues. International Journal of Information Management 33(5), pp. 875-891.
[18] Uckelmann, D. (2012). Quantifying the value of RFID and the EPCglobal Architecture Framework in Logistics. USA: Springer Science & Business Media.
[19] Amaral, L. et al. (2011). eCloudRFID–A mobile software framework architecture for pervasive RFID-based applications. Journal of Network and Computer Applications 34(3), pp. 972-979.
[20] Chattaraj, A.; Bansal, S. & Chandra, A. (2009). An intelligent traffic control system using RFID. Potentials 28(3), pp. 40-43.
[21] Kim, M. et al. (2008). Study on network architecture for traffic information collection systems based on RFID technology. Proceedings Conference Asia-Pacific Services Computing, pp. 63-68.

Desarrollo de un plataforma para el control inteligente de vehículos basada en identificación por radio frecuencia

Ernesto Neira
John Jairo Pantoja
Natalie Figueroa
Gabriel Mañana
Cesar Pedraza
Félix Vega
Paola Ordoñez
Universidad Nacional de Colombia – Colombia

1 Introducción

En 2013, por interés del Ministerio de Transporte Nacional y el Departamento Administrativo de Ciencia, Tecnología e Innovación (Colciencias), y en el marco de la convocatoria 622-Investigación Aplicada e Innovación para el Desarrollo de Sistemas Inteligente de Transporte, del Programa Nacional de Desarrollo Tecnológico Industrial y Calidad, le fue aprobado a la Universidad Nacional de Colombia, en conjunto con la empresa privada Datatools, el proyecto "Plataforma para el control Inteligente de Vehículos (PCIV)". El objetivo del proyecto fue desarrollar una plataforma tecnológica escalable que sirviera de soporte para la implementación de un sistema inteligente de gestión de vehículos.

Desde su inicio, la plataforma estaba concebida para ser integrada con diferentes sistemas de control y monitoreo de vehículos, en especial con los sistemas de identificación por radiofrecuencia (RFID). Esta tecnología se seleccionó debido a la iniciativa del Ministerio de Transporte para la implementación de un Sistema Nacional de Identificación Electrónica Vehicular (SINIEV), en el que se determinó el uso de ella bajo la norma ISO 18000-6C [1].

La tecnología RFID permite almacenar y extraer datos de dispositivos de manera remota, denominados Etiquetas o Tags, que se realiza a través de otros de lectura y escritura, llamados Lectores o Readers. Los Tags RFID pueden ser de dos tipos: activos o pasivos. Los activos presentan una fuente de alimentación eléctrica, mientras que los pasivos no la tienen. Generalmente se han usado dispositivos FRID pasivos en tres frecuencias: 125 kHz, en este caso se habla de etiquetas de proximidad para distancias de lecturas entre 5 cm y 10 cm; 13.56 MHz, logra distancias de lectura de hasta 1.5 metros y son conocidas como etiquetas de vecindad; y 915MHz, para distancias de lectura sobre los 10 metros. Además, los sistemas RFID han sido estandarizados por los organismos ISO y EPCglobal, encargados de la generación de normas para las diferentes aplicaciones de esta tecnología. Las normas particulares para sistemas vehiculares pueden revisarse en [1-4].

RFID ha sido propuesto para implementaciones de Sistemas Inteligentes de Transporte (ITS) [5, 6], sistemas que hacen referencia a un conjunto de soluciones tecnológicas para mejorar la operación de los sistemas de transporte. En ITS pueden encontrarse técnicas de monitoreo y control de tráfico basadas en procesamiento de imágenes, detección magnética, sistemas de infrarrojos, sistema de posicionamiento global (GPS), Bluetooth y RFID. Con el uso de RFID en ITS se puede obtener información en tiempo real de una operación de transporte para toma de decisiones en etapas de planeamiento o intervención.

En este trabajo se presenta los resultados obtenidos en el desarrollo del proyecto PCIV hasta la terminación del primer piloto (octubre 2014). En la Sección 2 se detalla la metodología adoptada; en la sección 3 la arquitectura del sistema; en las secciones 4 y 5 se describen los componentes hardware y software desarrollados; en la sección 6 las aplicaciones software desarrolladas para procesar y presentar la información durante el primer piloto. En la sección 7 se presenta el estudio del canal de comunicaciones; en la 8 se explica la metodología utilizada para la prueba piloto y en la sección 9 los resultados del piloto ejecutado en el campus de la Universidad Nacional de Colombia, sede Bogotá. Finalmente, en la sección 10 se presenta las conclusiones.

2 Metodología

La metodología propuesta está definida en el ciclo de vida para un proyecto propuesta por NASA en los Estados Unidos, específicamente acerca del flujo de procesos y de la documentación requerida con base en el documento NASA Procedural Requeriments 7120.5E (NPR 7210), de agosto de 2012. Esta metodología se aplica al ciclo de vida de proyectos que impliquen el desarrollo o apropiación de una tecnología para el diseño y fabricación de un sistema, que desempeñe una función previamente especificada o proporcione una capacidad operativa necesaria para el desarrollo de funciones misionales de una administración. El documento NPR 7210 define las siguientes fases:

1. Estudio de viabilidad de la misión
2. Definición de la misión
3. Definición del sistema
4. Diseño preliminar
5. Diseño final
6. Fabricación e integración
7. Preparación para el despliegue
8. Despliegue y verificación operacional
9. Operación
10. Retirada

Adicionalmente, para el desarrollo del proyecto se definieron 8 paquetes de trabajo o WP (Work Package) correspondientes al desarrollo de los módulos que comprenden PCIV, su integración y la implementación de un demostrador. Los WP utilizados fueron:

- WP1. Coordinación y gestión del proyecto
- WP2. Diseño de la red de sensores
- WP3. Diseño de interface entre lectores RFID y la red de comunicaciones
- WP4. Diseño de la interface entre red de comunicaciones y la plataforma de computación en la nube
- WP5. Diseño e implementación de la aplicación para almacenado de datos y la gestión de indicadores
- WP6. Integración del sistema
- WP7. Diseño e implementación del demostrador
- WP8. Diseño de un modelo de negocios

3 Arquitectura

En la Figura 1 se muestra la arquitectura general del sistema PCIV con los cuatro módulos principales. El primer módulo es el vehículo en el que se instala una etiqueta RFID pasiva y que interactúa a través del medio de propagación con el lector RFID. El segundo corresponde al componente de lectura RFID, denominado unidad terminal remota RTU y que realiza la lectura de los códigos contenidos en las etiquetas RFID,

procesarlos y enviarlos al Sistema de Información. Este módulo contempla soluciones para evitar la pérdida de información por fallas de comunicación entre RTU y el sistema. El tercer módulo corresponde al Sistema de Información, que contempla un sub-módulo de base de datos, un conjunto de aplicaciones para procesar la información y presentarla y un sub-módulo de administración o gestión. Este módulo está basado en una plataforma de computación en la nube y fue construido con capacidades de auto-escalamiento. Por otro lado, la arquitectura del sistema permite la interacción con bases de datos externas, aplicaciones corriendo en servidores externos y usuarios que consulten el sistema de información.

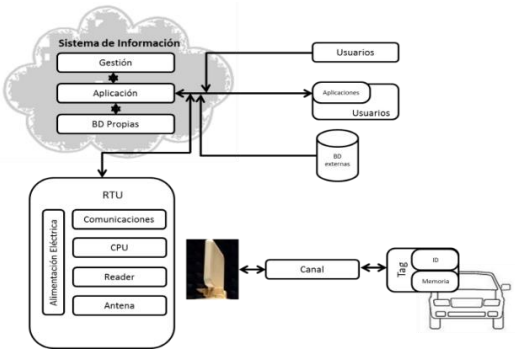


Figura 1: Arquitectura General de PCIV

La arquitectura definida por el equipo de trabajo es similar a otros sistemas propuestos para el control inteligente de vehículos [7]. Pero se diferencian en la capacidad para integrarse con Bases de Datos externas y la forma en que la información fluye hacia y desde aplicaciones que corren de manera local en equipos o servidores remotos. Adicionalmente, incluye una unidad de procesamiento basada en Linux instalada en RTU. Esta unidad realiza tareas de procesamiento en sitio. Algunas programadas durante la prueba piloto fueron: 1) el filtrado de etiquetas, que se realiza localmente para evitar el flujo de información en caso de detección de etiquetas que no pertenezcan a las usadas, 2) el almacenamiento en un Buffer, para evitar pérdida de información por fallas en la conexión con el servidor, y 3) tareas de diagnóstico y recuperación, con la finalidad de hacer una primera recuperación en caso de presentarse fallos en hardware.

Cabe resaltar que al mometo de definición de la arquitectura el equipo de trabajo de PCIV no contaba con conocimiento del sistema presentado en [7]. De hecho, esta patente fue publicada el 23 de abril de 2014, momento en el que la arquitectura ya estaba definida y en proceso de desarrollo.

4 Hardware

En la parte inferior izquierda de la Figura 1 se detalla el módulo RTU con sus cinco sub-módulos. La función de RTU es leer de manera remota los códigos contenidos en los Tags y luego, mediante un sub-módulo de comunicaciones, enviar los datos al software. En una primera etapa de diseño del módulo de hardware, la definición de los requerimientos arrojó que el sistema tendría que realizar lecturas a vehículos con velocidades superiores a 150 km/h, con un mínimo de lecturas de cinco por segundo. Debido a los requerimientos en la distancia de lectura, la frecuencia de trabajo fue determinada en 915 MHz bajo el estándar EPCGlobal Gen 2 (ISO 18000-6C), y de esta manera se identificaron los circuitos integrados para lectura RFID AS399X de Austria Microsystems [8] e Indy R2000 de Impinj [9]. Los circuitos fueron evaluados y se seleccionó el indy R2000 por rendimiento y especificaciones. Dentro de los sistemas

embebidos disponibles fueron evaluadas las placas de distribución comercial Raspberry PI y BeagleBone Black. Adicionalmente, se evaluó la placa miniStamp de desarrollo propietario de la Universidad Nacional de Colombia. Finalmente fue escogida la placa BeagleBone Black por las características ofrecidas y por la relación costo/beneficio.

Para el desarrollo del sub-módulo de antena se evaluaron tres diseños tipo parche con elementos parásitos. Para cada uno se construyó un recubrimiento usando una impresora 3D en ABS. Los tres modelos de antenas fueron simulados computacionalmente y caracterizados en la cámara anecoica de la Universidad de los Andes. Los resultados de la caracterización experimental se presentan en la Tabla 1 y en la Figura 2.

Tabla 1: Medición de la ganancia para las antenas diseñadas y construidas

Antena	Ganancia
Antena 1. Tipo parche	6.74 dBi
Antena 2. Tipo Parche	7.21 dBi
Antena 3. Tipo Parche	7.95 dBi

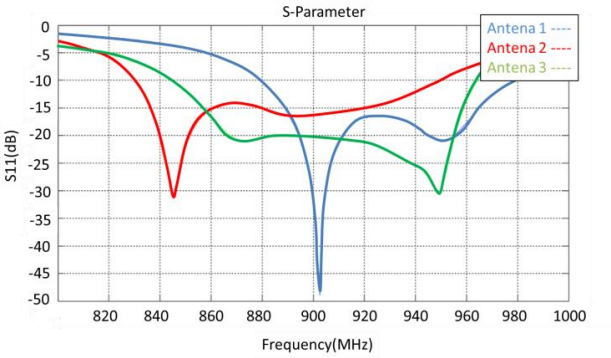


Figura 2: Coeficiente de reflexión de las antenas diseñadas para la RTU

La selección de la antena se realizó con base en la relación beneficio/costo, capacidad de reproducibilidad, facilidad de fabricación y fabricación nacional. La etiquetada como 3, que es una tipo parche con dieléctrico de aire y un direccionador, mostrada en la Figura 3, fue la escogida para ser usada en PCIV. Como se aprecia en los resultados experimentales esta antena presenta un desempeño adecuado con un coeficiente de reflexión menor a -15 dB y una ganancia cercana a 8 dBi en la banda de frecuencias de interés (915 MHz).

Figura 3: Antena tipo parche

En el sub-módulo de comunicaciones y debido a la disponibilidad de puertos de la placa BeagleBone Black, fueron probados los *modem* comerciales de conexión USB, de 4G de conexión cableada o *wireless* y los industriales. El mejor resultado en términos de disponibilidad fue logrado con los industriales y en definitiva se seleccionó el *modem* Sierra Wireless.

5 Software

Desde el punto de vista del software se definieron 37 casos de uso básicos de la plataforma software y se implementaron en una arquitectura Plataforma como Servicio (PaaS). En particular se usó el *cloud* público ofrecido por Google (Google Cloud Computing). A continuación se listan los casos de uso definidos:

1. Registrar una lectura.
2. Consultar estadísticas.
3. Autenticación de usuarios.
4. Recuperación de nombre de usuario.
5. Recuperación de contraseña.
6. Cambiar contraseña.
7. Actualización de datos personales.
8. Solicitar una cuenta de usuario nueva.
9. Confirmar solicitud de cuenta de usuario.
10. Agregar una secretaría de movilidad en el sistema.
11. Eliminar una secretaría de movilidad en el sistema.
12. Asignar un administrador a una secretaría de movilidad.
13. Asignar un auditor a una secretaría de movilidad.
14. Actualizar datos de una secretaría de movilidad.
15. Administración de todo el sistema SaaS.
16. Tarea de auditoría: Visualizar la bitácora de eventos y actividades de los demás usuarios.
17. Tarea del administrador: Visualizar solicitudes de creación de cuentas de usuario para una secretaría de movilidad.
18. Tarea del administrador: Activar/Desactivar cuentas de usuario para una secretaría de movilidad.
19. OCC: Visualizar la información general de los AT y los pórticos asociados a la secretaría de movilidad.
20. OCC: Visualizar la información detallada de los AT y los pórticos asociados a la secretaría de movilidad.
21. AT: Consultar información de uno(s) vehículo(s).
22. Control de una motocicleta en un retén de la Policía de Tránsito y Transporte.
23. Control de motocicletas cerca de un AT.
24. Lectura fallida de una etiqueta RFID por parte de un AT.
25. Lectura fallida de una etiqueta RFID por parte de un pórtico.
26. Registro de lecturas de etiquetas RFID.
27. Pórtico: Consultar información de una(s) moto(s).
28. Registra la anomalía y evidencia fidedigna por parte de los pórticos.
29. Pórticos alertan a los AT cercanos.
30. Cobro automático de peajes.
31. Registro de un propietario de una motocicleta.
32. Denuncia de robo de motocicleta.
33. Registro de accidentes por parte de un AT.
34. Registro de accidentes por parte de un propietario de una motocicleta.
35. Calibración de rutas de buses, flotas, y demás transportadores públicos.
36. Comparendos automáticos de anomalías.
37. Controles de velocidad.

El sistema usa el patrón cliente-servidor, donde el servidor está representado por un único componente (aplicación web) llamado PCIV, y que se despliega sobre la plataforma de Google App Engine. Esta plataforma se encarga de garantizar la escalabilidad a través de la gestión automática de la cantidad de instancias de la aplicación web, que deben estar disponibles en cada momento para atender todas las peticiones recibidas. Existe un componente cliente que se despliega en cada lector y se encarga de acceder a los servicios

proporcionados por la aplicación web. Los servicios ofrecidos ella están representados por una API que usa tecnologías web estándar soportadas por la plataforma de Google App Engine. La aplicación web usa dos servicios de la plataforma Google App Engine: Datastore y Logs para el almacenamiento y consulta de datos y registros, respectivamente. La Figura 4 muestra el diagrama de componentes del sistema.

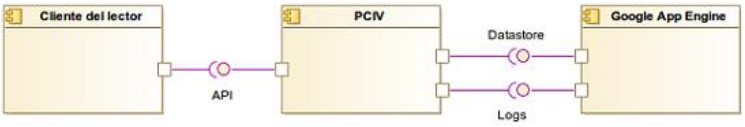


Figura 4: Diagrama de componentes del software

La finalidad del uso de la plataforma Google Cloud Computing es solucionar el requisito de escalabilidad desde el comienzo. Para probar esta característica de auto-escalabilidad desde el punto de vista de la validación de los parámetros, se definió un escenario de pruebas para la plataforma, consistente en la generación de una carga de tráfico equivalente a 2.100 lecturas, generadas por 70 lectores con 30 lecturas cada uno. Este tipo de carga correspondería a la hora pico de una ciudad del tamaño de Bogotá. Los resultados de las pruebas permiten concluir que la plataforma responde de manera automática y efectiva a la carga generada en este tipo de escenarios. El rendimiento obtenido en la prueba fue de 480 lecturas por segundo.

Otro desarrollo del componente software es la definición de un protocolo que incluye el formato de intercambio de datos y el mecanismo de autenticación y cifrado de datos entre RTU y la plataforma software. Por ejemplo, la autenticación entre RTU y el software se realiza usando el servicio Personal de Mozilla basado en el estándar VerifiedEmailProtocol [10].

6 Aplicaciones desarrolladas

Para la presentación de información a usuarios se construyeron nueve aplicaciones con presentación gráfica de los resultados, que corresponden con las necesidades para análisis de la información en el primer piloto.

Volúmenes vehiculares. Esta aplicación calcula la cantidad de vehículos que entraron y/o salieron de una zona durante un determinado intervalo de tiempo. Permite realizar la consulta con la posibilidad de filtrar por la clase de RTU (definidos como de entrada o salida) y por clase de vehículo.

Matriz rigen-destino. Permite calcular la distribución de viajes de los vehículos que se movilizan por una zona dada. A partir de esta estadística se puede analizar el volumen de vehículos que se dirigen de cierto origen a cierto destino.

Permanencia. Permite calcular un histograma con los tiempos de permanencia de cada vehículo leído dentro de la zona durante el intervalo de tiempo. Se grafican dos histogramas: uno con el valor nominal vs la cantidad de horas y el otro con el porcentaje de vehículos correspondientes de la muestra vs la cantidad de horas.

Ocupación. Permite mostrar la ocupación de una zona por clase de vehículo durante un período de tiempo. La ocupación consiste en la cantidad de vehículos que se encuentran dentro de ella en un instante dado.

Parqueaderos. Permite mostrar la ocupación absoluta y relativa de los diferentes parqueaderos de una zona utilizando mapas. Los parqueaderos están representados por polígonos sobre un mapa y el color de relleno representa su ocupación. Un color verde

oscuro representa un valor bajo, mientras que un color rojo oscuro representa un valor alto. En la ocupación absoluta el color de relleno de un parqueadero utiliza una escala global para todos los parqueaderos, mientras que en la ocupación relativa el color de relleno utiliza una escala propia de cada parqueadero, donde el valor máximo viene dado por su capacidad.

Etiqueta por día. Permite realizar un conteo las etiquetas diferentes que se leyeron por día o durante un período de tiempo. Muestra los datos usando un histograma.

Etiquetas. Permite ver las etiquetas registradas en el sistema y las lecturas que se realizaron de cada una de ellas durante un período de tiempo.

Lectores. Permite ver los lectores registrados en el sistema, crear nuevos o editar los existentes. Permite ver la última dirección IP y la fecha y hora en la cual un lector se reportó. Revisa periódicamente el estado de los lectores y envía un correo electrónico para avisar sobre posibles fallos. Muestra una gráfica que indica si el lector se estuvo reportando durante un período de tiempo (herramienta para medir la disponibilidad de cada lector).

Parqueaderos. Permite ver los parqueaderos registrados en el sistema, crear nuevos o editar los existentes. Permite al usuario trazar un polígono sobre un mapa para definir los límites del parqueadero.

7 Estudio del canal de comunicación

Las etiquetas pasivas extraen la energía para funcionar de la misma señal de interrogación proveniente del lector. En este caso, la energía recibida en el lector puede ser estudiada con la ecuación de Friis para espacio libre para identificar la distancia mínima definida por Rao, Nikitin y Lam [11] como:

$$r = \frac{\lambda}{4\pi}\sqrt{\frac{P_T G_T G_t \tau}{P_{th}}} \qquad (1)$$

Donde r es el rango de lectura, λ es la longitud de onda, P_T es la potencia transmitida por el reader, G_T es la ganancia de la antena del *reader* en dirección a la etiqueta, G_t es la ganancia de la antena de la etiqueta en dirección al lector, P_{th} es la potencia mínima necesaria para el funcionamiento de la etiqueta y τ es el coeficiente de transmisión de potencia dado por:

$$\tau = \frac{4R_c R_a}{|Z_c + Z_a|^2} \qquad (2)$$

Donde $Z_c = R_c + jX_c$ es la impedancia interna el chip de la etiqueta y $Z_a = R_a + jX_a$es la impedancia de la antena de la etiqueta. La ecuación 1 no incluye las pérdidas adicionales por ubicación de la etiqueta sobre un objeto, para este caso, el vidrio del vehículo. La ecuación presentada por Griffin y Durgin [12] ofrece mejor aproximación:

$$P_t = \frac{P_T G_T G_r \lambda^2 X \tau}{(4\pi r)^2 \Theta BF} \qquad (3)$$

Donde P_t es la potencia acoplada en el interior de la etiqueta, X es el desajuste de polarización, Θ es el un factor de penalización a la ganancia de la antena de la etiqueta por instalación sobre objetos, F es el margen de desvanecimiento en una sola trayectoria y B es la pérdida por obstrucción en trayectoria. Debido a la existencia del factor Θ, a que las etiquetas disponibles para las pruebas iniciales no eran adecuadas para su instalación

sobre vidrios y después de la realización de múltiples pruebas de ubicación, se estableció el espejo retrovisor como el más adecuado para ubicar las etiquetas, porque ofrece el mayor rango de lectura.

8 Piloto Universidad Nacional

Para probar los desarrollos realizados durante el proyecto, se estableció la realización de una prueba en las instalaciones de la Universidad Nacional de Colombia Sede Bogotá. En definitiva, la prueba piloto se planteó con la finalidad de establecer la estabilidad del hardware y el software desarrollados. Así mismo, identificar requerimientos que hubiesen escapado a la etapa de diseño y que pudieran realizarse durante o después del mostrador. Por otro lado, al poder elaborar una prueba piloto se estableció la realización de algunas mediciones típicas en el campo de transportes en el campus.

Durante el diseño de la prueba se definió un requerimiento mínimo de 300 vehículos, entre motocicletas y autos, que asistieran permanentemente a las instalaciones de la universidad. Para lograr esta participación de la comunidad universitaria se definieron cuatro fases:

1. Campaña de expectativa
2. Campaña de instalación de etiquetas
3. Pruebas y recolección de información
4. Retirada

En primer lugar, se parte de reconocer la importancia de un proceso de difusión y socialización que se acercara a las realidades sociales de los actores involucrados, y de la necesidad inherente de promover el acceso oportuno y transparente a la información. El objetivo es mitigar la generación de falsas expectativas e impulsar los impactos positivos que surjan desde una población informada e interesada. En segundo lugar, se vio la necesidad de proponer un nombre para el proyecto, diferente a PCIV, con el objetivo de lograr mayor aceptación y recordación en la comunidad universitaria. Luego de varias opciones se optó por SINDEMORAS, como acrónimo de Sistema de Movilidad en Red Automático y Seguro. Se establecieron contactos físicos y virtuales con las 11 facultades de la universidad y los 32 departamentos, con las oficinas de Bienestar de cada facultad y con la oficina de Bienestar de la sede. El propósito era llegar a los miembros de la comunidad abarcando escalas con grados diferentes de cercanía a cada posible participante: departamento, facultad y Universidad, porque de cada una de estas dependencias se enviaría información al correo de los posibles participantes.

Para dar inicio a la campaña de instalación de etiquetas, que tuvo una duración de 3 semanas, se convocaron estudiantes de corresponsabilidad de la Universidad Nacional. Los estudiantes de corresponsabilidad son estudiantes de pregrado que reciben apoyo socio económico de alimentación u hospedaje por parte de la institución. A cambio deben cumplir un mínimo de 20 horas de ayuda en labores de archivo o soporte de investigación, asistencia y demás que requieran las diferentes dependencias administrativas o investigativas del campus. Finalmente fueron 11 estudiantes, quienes luego de ser capacitados en los aspectos técnicos y logísticos de la tecnología e instalación de etiquetas, completaron sus horas con el proyecto SINDEMORAS. El resultado de la estrategia fue la participación de más de 700 vehículos, cerca de 400 más de la cantidad mínima que se había calculado para una prueba exitosa, asegurando así una mayor confiabilidad en los datos y confirmando un acercamiento efectivo a los participantes.

El campus cuenta con tres entradas y salidas vehiculares. En total se instalaron 6 unidades RTU en las porterías (tres para entradas y tres para salidas) (Figura 5). Se construyó otra RTU móvil con la finalidad de realizar recorridos periódicos por los diferentes parqueaderos del campus, que adicionalmente tiene un GSP para reportar a la plataforma la posición geográfica en la cual fue realizada la lectura. El flujo de la información entre RTU y la plataforma permite el intercambio de la siguiente información: fecha, hora de lectura, id de la etiqueta, tipo de vehículo y localización.

Figura 5: Ubicación de los lectores en las porterías de la Universidad

9 Resultados

La determinación de los volúmenes vehiculares forma parte de la información básica para el estudio y análisis de las condiciones del tránsito en corredores viales urbanos. Por esta razón su cuantificación constituye una de las principales medidas en cualquier estudio de tránsito y transporte. Así mismo, el estudio de origen y destino sirve para obtener datos del número y tipo de viajes en un área, incluyendo movimientos de vehículos y pasajeros o carga de varias zonas de origen a varias zonas de destino. El origen del viaje corresponde a la estación donde se registra un vehículo por primera vez y el destino es el lugar donde se le vio por última vez. En este estudio se recomienda que, de modo paralelo, se efectúen conteos vehiculares para realizar el ajuste a la muestra que finalmente se logre obtener.

Por otro lado, en el desarrollo de los estudios de estacionamiento el mayor problema que se presenta es determinar cuántos espacios de parqueo se requieren en cualquier tipo de proyecto, y sobre todo identificar dónde deben localizarse los espacios. Para esto se determina el tiempo de permanencia de los vehículos en un espacio cerrado, mediante la diferencia entre la hora de entrada y la de salida y se identifica la ocupación del estacionamiento a los largo de un periodo horario.

Estos tres ensayos fueron realizados en la prueba piloto de la universidad utilizando tecnología RFID. La prueba se realizó en un periodo de 40 días calendario, lo que permitió tener lecturas 24 horas del día de los vehículos que ingresan y salen del campus, adquiriendo resultados en línea de lo sucedido en tiempo actual o en días anteriores y posteriores (durante el periodo de prueba). A continuación se presentan resultados obtenidos en la prueba piloto, que permiten evidenciar patrones de comportamiento a nivel global de la universidad a través de la información recopilada de los Tags instalados a los usuarios voluntarios que hicieron parte del ensayo, y teniendo en cuenta que los números reflejados hacen referencia exclusivamente a esta población.

La Figura 6 muestra la representación gráfica del patrón de comportamiento, horario de entradas y salidas al campus, para un periodo de consulta entre el 6 de Octubre a las 00:00 horas y el 11 de Octubre a las 00:00 horas, lo que permite ver el comportamiento de una semana típica en el campus. Además, esta figura presenta la interfaz del software diseñado para las salidas de información.

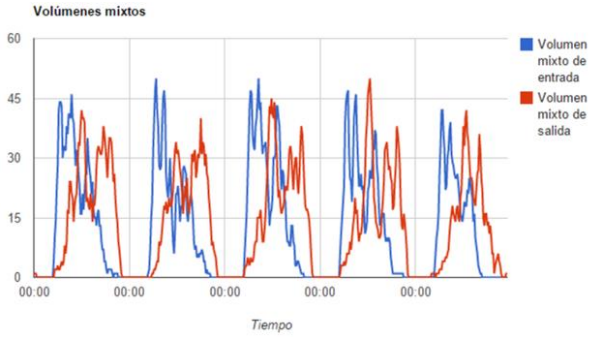


Figura 6: Volúmenes vehiculares, prueba piloto Universidad Nacional

La Figura 7 representa los resultados obtenidos en un día de consulta (09 a 10 de Octubre de 2014) de las tres porterías en donde fueron instalados los lectores, esto permite identificar el lugar de entrada, salida y en general el más utilizado por los usuarios de la prueba.

	[1] Salida Capilla	[2] Salida Carrera 45	[3] Salida Calle 53	Total general
[1] Entrada Capilla	15	8	17	40
[2] Entrada Carrera 45	37	71	38	146
[3] Entrada Calle 53	7	20	38	65
Total general	**59**	**99**	**93**	**251**

Figura 7: Matriz origen–destino de la prueba piloto

La Figura 8 lustra la ocupación durante un día en el campus para automóviles y motocicletas que poseían el Tag, esto para un periodo en entre el 07 y 08 de Octubre de 2014. Así mismo permite identificar la duración promedio de los vehículos por horas dentro de la Universidad. Dentro de los datos obtenidos mediante la prueba piloto realizada con la tecnología RFID, se identificaron los siguientes comportamientos: las horas pico de entradas obtenidas se encuentran dentro de los periodos horarios entre las 6:30 am y las 10:00 am en la mañana y entre las 13:00 y 14:00 en la tarde, que se relacionan con las horas de inicio de actividades en el campus (en la mañana) y la hora de almuerzo (en la tarde). En cuanto a las salidas, se evidencia que el primer pico se presenta en las horas de la mañana, entre las 9:00am y las 10:00am, el segundo y el más relevante del día se presenta entre las 12:00 y las 13:00 horas debido a la hora de almuerzo, y finalmente entre las 18:30 y 19:30 debido a la finalización de la jornada. De igual forma se evidencia que el ingreso más utilizado por los usuarios del campus es la entrada de la carrera 45 o Uriel Gutiérrez, seguido por la calle 53 y por último la capilla. Este mismo comportamiento se presenta en la preferencia de acceso a la salida de la Universidad. En cuanto a la duración de los vehículos dentro del campus la mayoría de la población vehicular permanece al interior de la universidad entre 6 y 12 horas, seguidos por un periodo de entre 1 y 3 horas. Haciendo referencia a la ocupación, el periodo pico se encuentra entre las 11:00am y 12:00 del mediodía, presentándose en este periodo el mayor número de vehículos dentro del campus.

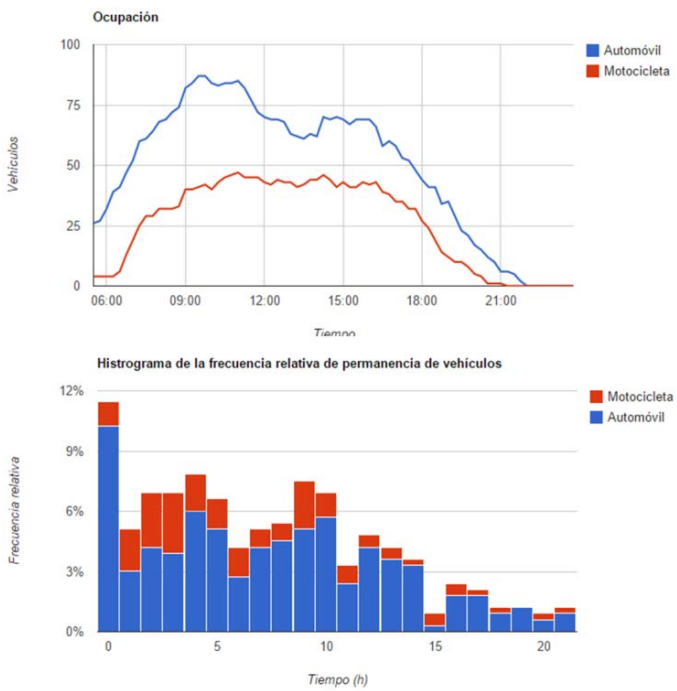


Figura 8: Permanencia en estacionamiento en la prueba piloto

10 Conclusiones

El presente trabajo describe de manera detallada las etapas llevadas a cabo para el desarrollo de la plataforma de control inteligente de vehículos PCIV. Las pruebas a nivel de componentes y a nivel de sistema realizadas muestran el cumplimiento de manera satisfactoria con los requisitos de auto-escalabilidad, estabilidad e integrabilidad planteados para la plataforma. Adicionalmente, los resultados de las pruebas muestran que la tecnología RFID basada en etiquetas pasivas permite la adquisición de variables para el análisis de sistemas de transporte de manera automatizada y confiable, mediante la identificación individualizada de vehículos. Análisis de volúmenes vehiculares, matrices origen destino y permanencia en estacionamientos, como los efectuados en la prueba piloto, son algunos estudios que posibilita el sistema PCIV y que permite el paso a implementaciones de un ITS.

Actualmente, se está realizando una segunda prueba a nivel de sistema en el municipio de Chía, Cundinamarca, Colombia. Esta prueba permitirá validar la operación del sistema en un ambiente urbano y mejorar la visualización de la información mediante pruebas de usabilidad con usuarios finales.

Agradecimientos

Este trabajo fue desarrollado por la Universidad Nacional de Colombia - Sede Bogotá y Datatools S.A., dentro del proyecto de investigación "Plataforma para el control inteligente de vehículos PCIV" (HERMES 20097), el cual es financiado por el Ministerio de Transporte a través del contrato 0654-2013 y otorgado mediante la convocatoria 622-2013 de Colciencias.

Referencias

[1] ISO/IEC (2013). ISO:IEC 18000 - Information technology -- Radio frequency identification for item management.
[2] ISO/IEC (2013). ISO/IEC 15962 - Information technology -- Radio frequency identification (RFID) for item management.
[3] ISO/IEC (2010). ISO/IEC 15693 - Identification cards -- Contactless integrated circuit cards -- Vicinity cards.
[4] EPCGlobal (2015). Radio-Frequency Identity Protocols Generation-2 UHF RFID.
[5] Togashi, H. & Yamada, S. (2010). An Information provision system based on a Multi-Hop RFID scheme for ITS (Intelligent Transportation System). Proceedingas IEEE conference on Open Systems, pp. 97–102.
[6] Ali, K. & Hassanein, H. (2009). Using Passive RFID Tags for Vehicle-Assisted Data Dissemination in Intelligent Transportation Systems. Proceedings 3rd IEEE LCN Workshop on User MObility and VEhicular Networks, pp. 688–694.
[7] Google Patents (2014). Vehicle intelligent management system based on cloud computing platform.
[8] Alldatasheet (2015). AS3990/AS3991 UHF RFID Single Chip Reader EPC Class1 Gen2 Compatible.
[9] Impinj (2014). Indy® R2000 Reader Chip (IPJ-R2000).
[10] MozillaWiki (2012). Labs/Identity/VerifiedEmailProtocol.
[11] Rao, S.; Nikitin, P. & Lam, S. (2005). Antenna Design for UHF RFID Tags: A Review and a Practical Application. IEEE Trans. Antennas Propag. 53(12), pp. 3870–3876.
Griffin, J. & Durgin, G. (2009). Complete Link Budgets for Backscatter-Radio and RFID Systems. IEEE Antennas Propag. Mag. 51(2), pp. 11–25.

Diseño y fabricación de un sistema monopulso para aplicaciones RFID

Carolina Alcaraz[1]
Juan Balbastre[2]
Félix Vega[3]
[1, 3]*Universidad Nacional de Colombia – Colombia*
[2]*Universidad Politécnica de Valencia – España*

1 Introducción

Las antenas monopulso se emplean ampliamente en aplicaciones donde el tamaño, el peso, el rendimiento y la facilidad de fabricación son importantes, por ejemplo, en los satélites, misiles y sistemas de comunicación inalámbrica. Por lo cual, recientemente se han realizado varias propuestas de sistemas monopulso para diferentes aplicaciones [1, 2], donde se proponen antenas de bajo costo con estructuras simples de fabricación alimentando bidireccionalmente antenas microstrip (parche) y lograr los patrones suma y diferencia para aplicaciones de radar y sistemas de comunicaciones.

Los sistemas RFID pasivos pueden detectar y registrar el EPC (Electronic Product Code) de una etiqueta previamente instalada en el vehículo, pero el sentido de movimiento del vehículo no puede ser fácilmente determindado, a menos que se implementen dos o más lectores RFID en una misma vía, elevando los costos del sistema. Sin embargo, con el diseño e implementación de una antena monopulso que permita identificar el vehículo a través del canal suma ($\sum$), como en un sistema RFID convencional, y determinar la polaridad de la señal recibida por el canal diferencia (Δ), se establecerá la dirección de movimiento del vehículo con respecto de la antena. Por ejemplo, el cambio de positivo a polaridad negativa podría significar que el vehículo está circulando de izquierda a derecha [4]. De esta manera, al integrar la antena monopuso al sistema RFID se añade la posibilidad de detectar la direcicción de movimiento, sin incrementar el costo del sistema.

En este trabajo se propone el diseño e implementación de antenas monopulso en aplicaciones de sistemas RFID para el control vehicular, con la finalidad de proveer la detectección del sentido de movimiento de los vehículos, sin incrementar el costo del sistema. Los sistemas RFID se utilizan cada vez más en aplicaciones de control vehicular: para la verificación automática del cumplimiento de las normas de transito y transporte, el reacudo automático de peajes y para la determinación de velocidad y tiempos de recorrido de vehículos de transporte público [3]. Además, estos sistemas también pueden proporcionar información sobre los flujos vehiculares que generan estadísticas de tráfico, para apoyar las decisiones estratégicas sobre la regulación de transito y transporte. En la sección 2 se presenta el diseño y la fabricación de las antenas parche y el acoplador híbrido; en la 3 se detallan los resultados de medición de la ganancia de la antena y las pérdidas de retorno; en la 4 se presenta un modelo determinista basado en la ecuación de Friis para evaluar el comportamiento del sistema, y las conclusiones del trabajo se destacan en la sección 5.

2 Metodología de diseño y fabricación de una antena tipo parche y el acoplador híbrido de 180°

El sistema monopulso consta de dos antenas de parche rectangular, que se utilizan comúnmente en los lectores de RFID debido a su portabilidad y fácil fabricación. Para

implementar el sistema monopulso, se conectan dos parches simétricos a un acoplador híbrido de 180º con el fin de producir los canales suma (Σ) y diferencia (Δ). Este sistema se diseñó por separado en CST Studio Suite 2014, sobre un sustrato FR4 de 1.6 mm de grosor, con permitividad relativa de 4.16, y simulando en primer lugar una antena tipo parche resonante en 915 MHz y después el híbrido de 180°. Se analizaron los resultados y se realizó una simulación final conectando las antenas parche simétricas por medio del acoplador híbrido.

La antena parche consiste de un metal situado sobre un sustrato dieléctrico con su parte inferior sobre un plano referenciado a tierra. Estas antenas tienen diferentes tipos de alimentación, y una de las más comunes es por línea de transmisión. Para obtener mejores resultados la línea se conecta a un adaptador de impedancia. El diseño en CST se muestra en la Figura 1 y los resultados de simulación en la Figura 2. Se pueden observar las pérdidas por retorno a 915 MHz de -10.72 dB.

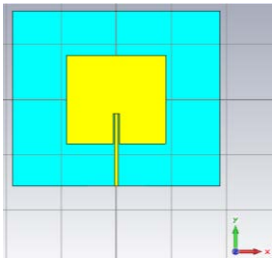


Figura 1: Diseño de antena tipo parche resonante en 915 MHz

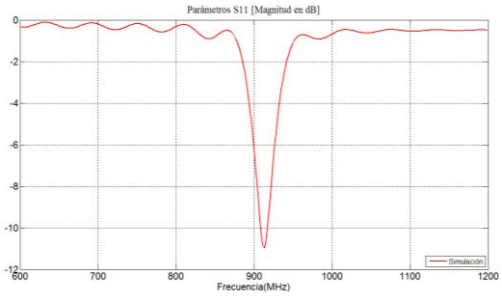


Figura 2: Perdidas por retorno de antena parche

El híbrido de 180° se obtuvo diseñando un acoplador estándar *branch-line* (híbrido de 90°) y añadiendo un desfasador tipo meandro, para sumar los 90° de desplazamiento de fase necesarios para obtener los 180° deseados. El diseño se muestra en la Figura 3. Los valores medidos de parámetros de dispersión significativos se presentan en las Figura 4 a 9. Se muestra un comportamiento razonablemente bueno para la aplicación monopulso: error de menos de 4% en desplazamiento de fase Figura 4 y menos de 0.15% Figura 6, y 0.19% Figura 8 de asimetría entre los puertos acoplados.

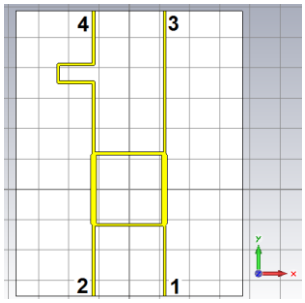


Figura 3: Diseño de acoplador híbrido de 180°

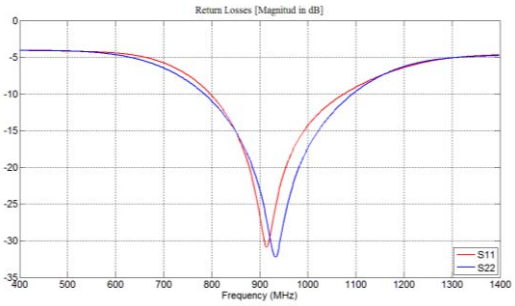


Figura 4: Pérdidas por retorno del híbrido de 180°

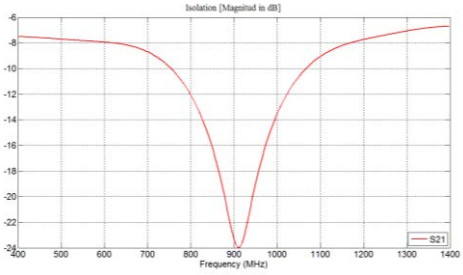


Figura 5: Isolación del híbrido de 180°

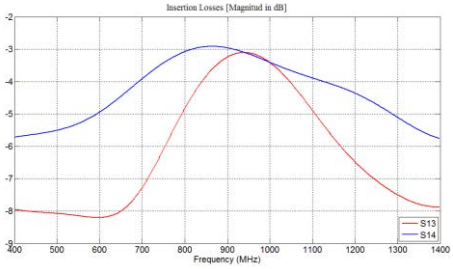


Figura 6: Pérdidas por inserción en canales 13 y 14

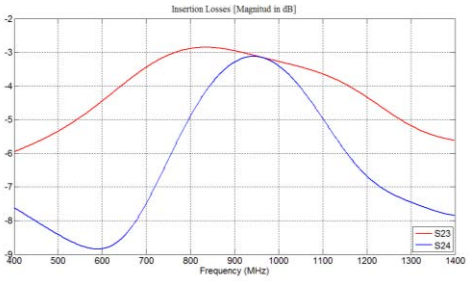


Figura 7: Fase de pérdidas por inserción en canales 13 y 14

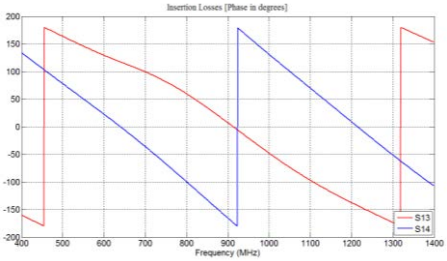


Figura 8: Pérdidas por inserción en canales 23 y 24

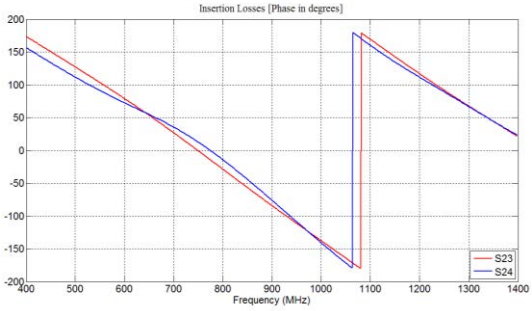


Figura 9: Pérdidas por inserción en canales 23 y 24

Después de obtener los resultados de la antena y el acoplador híbrido por separado se unieron en un único diseño, se hizo una simulación final y se fabricó el prototipo por el proceso de fotolitografía. Para realizar este proceso se sensibilizó el sustrato por medio de una película fotosensible [5], la cual se adhiere al sustrato y reacciona al contacto con la luz formando una imagen latente, luego se somete la placa sensibilizada a un proceso de calentamiento para que la película se adhiera completamente a la lámina. En un equipo Photoplotter se imprime el diseño de la antena en negativo para crear la máscara fotográfica, después de tener el circuito impreso se somete a un proceso químico de

revelado. Al transferir la máscara fotográfica al cobre se coloca la máscara sobre la superficie de la lámina sensibilizada y se ingresa en la máquina insoladora. Al retirarla de la insoladora se sumerge en una mezcla de agua con químico revelador *photoresist* y se talla para ir revelando el diseño de la antena en la lámina. El resultado de fabricación del prototipo se puede observar en la Figura 10.

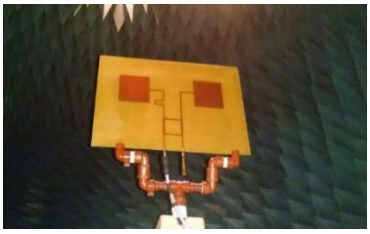


Figura 10: Sistema monopulso

3 Resultados y análisis del sistema monopulso

El sistema monopulso de la Figura 10 consta de la conexión de dos parches cuadrados resonantes ($\lambda/2 \times \lambda/2$ a 915 MHz) acoplados a los puertos del híbrido. La separación de los parches de centro a centro es de 20.35 cm. El prototipo se fabricó sobre sustrato FR4 de 1.6 mm de espesor con medidas de 40 x 30 cm; se caracterizó en la cámara anecoica de la Universidad de los Ándes, usando un Agilent Modelo HP E5062A VNA) y una antena de referencia ETS-Lindgren (LPDA) Modelo 3148B [6]. La medición se realizó colocando una carga de 50 Ω en un canal y por el otro se tomaron los datos de pérdidas por retorno y patrones de radiación.

Los resultados de medición y simulación de las pérdidas por retorno se muestran en las Figuras 11 y 12, donde se observa un ligero cambio de fase. Este cambio se debe a la incertidumbre en los datos de permitividad relativa y es común de las estructuras resonantes como la considerada en este trabajo. Aunque este efecto se puede eliminar mediante antenas de banda ancha, se prefiere esta forma en aras de la simplicidad y el costo, por lo que el rendimiento del sistema conseguido con esta antena es aceptable.

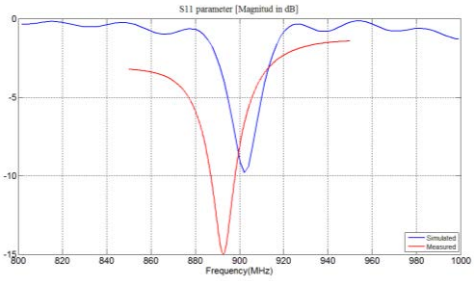


Figura 11: Pérdidas por retorno en el canal ∑

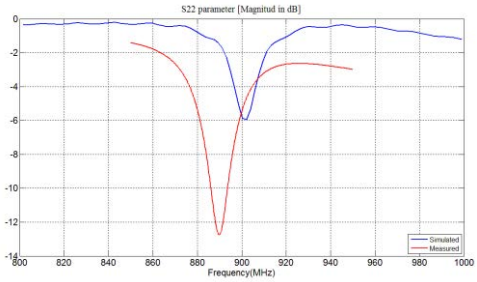


Figura 12: Pérdidas por retorno en el canal Δ

Los diagramas de radiación medidos se muestran en las Figuras 13 y 14, junto con los simulados, mostrando una buena concordancia. Por el canal ∑ se obtuvo 2.44 dB en simulación y 3.54 dB en medición. Por el canal Δ se obtuvo 2.23 dB en simulación y 2.77 dB en medición.

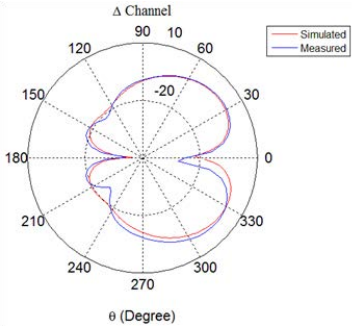


Figura 13. Patrón de radiación de ganancia por el canal ∑

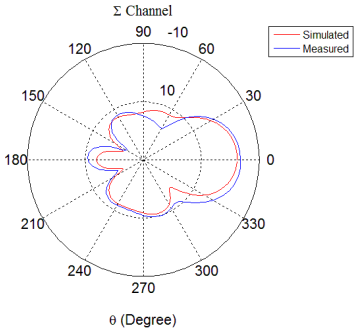


Figura 14: Patrón de radiación de ganancia por el canal Δ

4 Modelo determinista basado en la ecuación de Friis

Se diseñó un modelo con el fin de evaluar el comportamiento del sistema de recepción monopulso integrado en un sistema RFID. Este es un modelo simple basado en las respuestas de una etiqueta a la señal radiada por el lector, en función de los parámetros de funcionamiento y utilizando la ecuación de Friis de espacio libre. El modelo asume que un vehículo pasa a una velocidad constante en una trayectoria paralela al eje *x* y separado a una distancia *y* de la antena que se encuentra en el origen de coordenadas, como se indica en la Figura 15. Para un tiempo inicial t=0 el vehículo se encuentra en la posición de trayectoria x=-20 m.

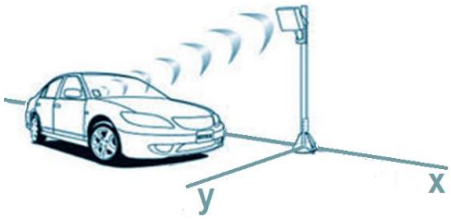


Figura 15: Escenario de simulación del modelo determinista

Usando este modelo la potencia recibida por el tag en cada punto de la trayectoria se calcula usando la ecuación (1) y la potencia recibida por el lector se calcula utilizando las ecuaciones (2) y (3), donde $G_{\Sigma}[\theta, \varphi(t)]$ es la ganancia medida correspondiente al canal suma y $G_{\Delta}[\theta, \varphi(t)]$ es la ganancia medida en el canal diferencia. G_{tag} es la ganancia del tag, que se asume como un monopolo elemental ideal con $G_{tag} = 3$.

$$P_{r,tag} = P_{t,reader}\Sigma[\theta,\varphi(t)]G_{tag}\left(\frac{\lambda}{4\pi r(t)}\right)^2 \qquad (1)$$

$$P_{r,\Sigma} = P_{t,tag}G_{\Sigma}[\theta,\varphi(t)]G_{tag}\left(\frac{\lambda}{4\pi r(t)}\right)^2 \qquad (2)$$

$$\boldsymbol{P_{r,\Delta} = P_{t,tag}G_{\Delta}[\theta,\varphi(t)]G_{tag}\left(\frac{\lambda}{4\pi r(t)}\right)^2} \qquad (3)$$

$$V_{\Delta}(t) = \sqrt{Z_{in}P_{r,\Delta}(t)}sgn\left\{\Delta\left[\frac{\pi}{2},\varphi(t)\right]\right\} \qquad (4)$$

La tensión recibida en el canal Δ se calcula utilizando (4), donde $Z_{in} = 50\ \Omega$. Aunque este valor no es crítico debido a que la magnitud relevante es la polaridad de la tensión y no su amplitud.

Por medio de la potencia recibida en el canal Σ se decodifica la respuesta, como se realiza comúnmente en un sistema RFID, mientras que la tensión recibida en el canal Δ se utiliza para identificar el signo del voltaje. Los cambios en la polaridad de voltaje indican el sentido del movimiento del vehículo (por ejemplo, un cambio de positivo a negativo indica un desplazamiento a $+\hat{x}$). Las Figuras 16 y 17 muestran la potencia recibida en el lector, por lo que se observa que el tag siempre emite una respuesta a las interrogaciones del lector. Estos resultados son para una velocidad constante y varias separaciones transversales entre los vehículos que llevan la etiqueta y el lector, y los resultados de las Figuras 18 y 19 representan el modelo para una separación transversal constante entre los vehículos y el lector a diferentes velocidades.

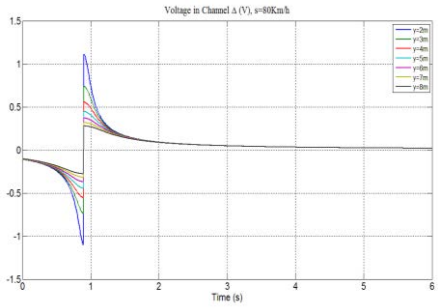


Figura 16: Resultados del modelo determinista en el canal Σ

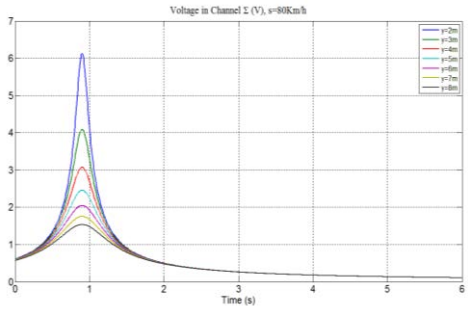


Figura 17: Resultados del modelo determinista en el canal Δ

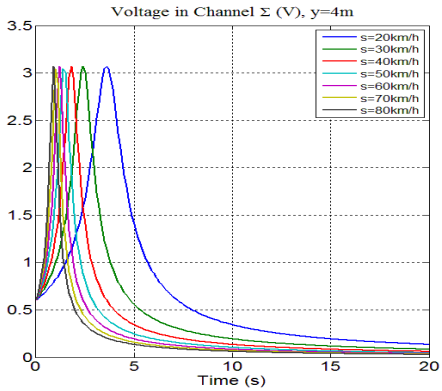


Figura 18: Resultados del modelo con separación constante y diferentes velocidades para el canal Σ

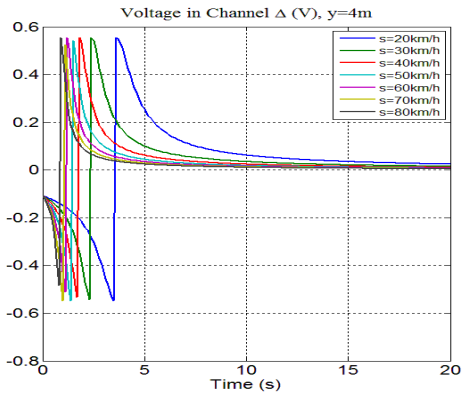


Figura 19: Resultados del modelo con separación constante y diferentes velocidades para el canal Δ

Debido al comportamiento aleatorio de las etiquetas reales debido al ruido y al desvanecimiento, la potencia recibida por la etiqueta se calcula teniendo en cuenta que esta potencia sigue una distribución exponencial con el valor medio dado por (1), y que corresponde a un desvanecimiento Rayleigh, y el modelo de ruido Gaussiano de media cero [3]. Si la potencia está por encima del umbral correspondiente se asume una respuesta válida para ser emitida. La potencia recibida por el lector en los canales Σ y Δ, se calcula utilizando (2) y (3), teniendo en cuenta el lector como un detector ideal. La probabilidad de detección para una sola respuesta se calcula utilizando el método de Monte Carlo con 100.000 ensayos. La Figura 20 muestra la probabilidad de detección para un auto que pasa 2 m de distancia del lector, a lo largo del eje de la calle a 50 km/h, y para diversos umbrales que van desde -10 dBm a 0 dBm, que corresponden a los valores de umbral reportados en la literatura [7, 8].

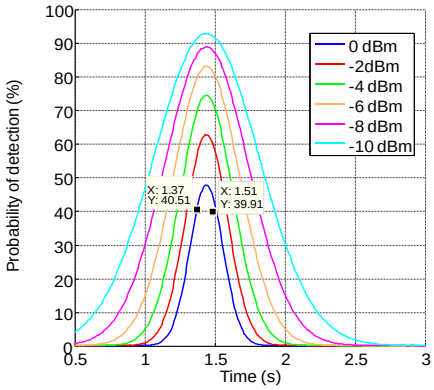


Figura 20: Probalidad de detección para un tag pasando a una distancia de separación 'y' de 2m a 50 km/h y para varios valores de umbrales de etiquetas

Como era de esperar, la fig. 20 muestra que cuanto menor sea el umbral, mayor será la probabilidad de detección. Para obtener una probabilidad de detección superior al 90% con un solo interrogatorio, se necesitan umbrales más bajos que -10 dBm. Sin embargo, teniendo en cuenta el hecho de que cuando la etiqueta está pasando delante del lector que va a recibir una serie de interrogatorios dados por:

$$n = t_{ob} N_s \qquad (5)$$

Donde t_{ob} es el tiempo de observación para el cual la probabilidad individual de detección es mayor que un valor dado (es decir, $P_{d,0}$) y N_s es la tasa de interrogación, la probabilidad acumulada de detección $P_{d,C}$ se puede calcular con:

$$P_{d,C} \geq 1 - \prod_{i=1}^{n} \left(1 - P_{d,0}\right) \qquad (6)$$

Por lo tanto, un buen rendimiento de detección se puede obtener usando umbrales moderados. Por ejemplo, usando los valores proporcionados por la Figura 20, para 0dBm un umbral $P_{d,0} = 40\%$, el tiempo de observación es $t_{ob} = 140$ ms; asumiendo una tasa de interrogatorio de $N_s = 100$ interrogaciones por segundo, se logra una probabilidad acumulada de detección de $P_{d,C} \geq 99.92\%$.

5 Conclusiones

Para mejorar los resultados del sistema monopulso se deben realizar técnicas de optimización para las antenas parche, en aras de aumentar el ancho de banda, mejorar la adaptación y disminuir su tamaño, y evitar el ligero cambio de fase que se obtuvo en la caracterización. Los valores medidos de ganancia se han utilizado para desarrollar un

sencillo simulador del sistema de procesamiento RFID con un sistema monopulso integrado, tanto la amplitud de la señal recibida (a través del canal Σ) y su polaridad (a través del canal Δ), permitiendo de este modo no solamente detectar e identificar la etiqueta sino también el sentido de movimiento del vehículo.

El rendimiento a nivel del sistema se ha evaluado mediante el método de Monte Carlo, con datos de la sensibilidad de la etiqueta obtenidos de la literatura. En trabajos futuros, se considerarán algunas mejoras en el receptor monopulso (usando antenas de banda ancha para reducir el efecto de la incertidumbre del sustrato y antenas de polarización circular para reducir el efecto de desvanecimiento). También son necesarios una caracterización de la etiqueta y el lector más precisas en entornos reales para una mejor evaluación del rendimiento del sistema.

Referencias

[1] Kim, S. & Chang, K. (2003). Low-cost monopulse antenna using bidirectionally-fed microstrip patch array. Electron Lett. 39(20), pp. 1428 -1429.

[2] Yu, Z. et al. (2011). Design of a broadband planar monopulse antenna of S band. Proceedings Cross Strait Quad-Regional Radio Science and Wireless Technology Conference, pp. 510-513. Harbin, China.

[3] Ministerio de Transporte (2012). Proyecto de decreto: Por medio del cual se adopta el Sistema Nacional de Identificación Electrónica Vehicular.

[4] Alcaraz, C.; Balbastre, J. & Vega, J. (2015). Enhancement of ITS application of RFID technology by means of radar techniques. Proceedings International Symposium on Antennas and Propagation and North American Radio Science Meeting AP-S/URSI. Vancouver, Canada.

[5] Diseños Electrónicos (2010). Fabricación de PCB - Método UV. Online [Feb. 2014].

[6] ETS-Lindgren (2014). 3148B Log periodic dipole array. Online [Mar. 2014].

[7] Orlenius, J. et al. (2011). Measurements of RFID Tag Sensitivity in Reverberation Chamber. Antennas and Wireless Propagation Letters 10, pp. 1345-1348.

[8] Nikitin, P. et al. (2009). Sensitivity and Impedance Measurements of UHF RFID Chips. Transactions on Microwave Theory and Techniques 57(5), pp. 1297-1302.

Hacia una plataforma de software libre de tutoría inteligente para B-learning: Arquitectura de referencia

Martha Giraldo G.
Armando Ordóñez
Fundación Universitaria de Popayán – Colombia

1 Introducción

Una de las plataformas más utilizadas para la enseñanza es los Sistemas para la Gestión de Aprendizaje (LMS). Sin embargo, en los LMS los estudiantes se enfrentan a contenidos y estrategias pedagógicas uniformes que desconocen los procesos individuales y pueden conllevar a la desmotivación o deserción. La desmotivación surge como consecuencia de que la mayoría de los procesos de aprendizaje se encuentran muy lejanos de la realidad del estudiante, y los procesos pedagógicos se aplican a todos uniformemente, sin reconocer las particularidades individuales [1]. Algunos estudios demuestran que la personalización de procesos de aprendizaje contribuye a la motivación de los estudiantes, mejorando sus resultados en las evaluaciones [2, 3]. En este sentido, los tutores inteligentes de aprendizaje (ITS) ofrecen una herramienta para promover la construcción activa de conocimiento, debido a que potencializan la tutoría uno a uno de forma automática [4]. Los ITS se han aplicado ampliamente en aulas de todo el mundo con buenos resultados, en comparación con los mecanismos tradicionales [5].

Los ITS son ambientes de aprendizaje computarizado que incorporan modelos computacionales de las ciencias cognitivas, ciencias del aprendizaje, lingüística, Inteligencia Artificial, matemáticas, y otros campos [6]. Para realizar la personalización del proceso de aprendizaje, modelan efectivamente algunas características de los estudiantes (habilidades, estrategias, motivación, emociones) y a los modelos pedagógicos, y guían a los primeros en los procesos de aprendizaje de acuerdo con sus necesidades particulares [6].

Aunque los ITS se enfocan en la enseñanza autónoma y son usados desde lugares remotos, también pueden ser utilizados para soportar procesos presenciales [7]. Mientras que los docentes presenciales pueden dirigir los cursos, los ITS se pueden usar como herramientas para reforzar particularidades del aprendizaje de cada estudiante. Esto contribuirá a mejorar la calidad y disminuir la deserción [8]. En los procesos de implantación de ITS, los docentes presenciales cumplen el papel de guiar, motivar y hacer seguimiento a los estudiantes. A pesar de esto, es muy poca la relevancia que los docentes presenciales le dan a los ITS, y de hecho esto se reconoce como una barrera importante a la no-implantación de los ITS [9]. Adicionalmente, muchas veces los profesores tienen muy bajo conocimiento en TIC, factor que puede influir en su motivación para participar en estos proyectos [10]. Existen otras barreras a la implantación adecuada de los ITS, tales como los costos, la falta de estandarización, la dificultad para la reutilización de contenidos y la complejidad tecnológica de las interfaces, que obliga la necesidad de personal técnico altamente cualificado [8].

En este trabajo se presenta los avances en la construcción de la plataforma AUTO4L [11] de B-Learning soportada en un ITS estándar de software libre, que reduce los costos de implantación. Igualmente, la propuesta busca reducir la complejidad de las interfaces para mejorar la participación de los profesores y estudiantes en los procesos de aprendizaje [12]. Está organizado como sigue: la sección 2 presenta los trabajos

relacionados, la sección 3 contextualiza el *framework* generalizado de tutoría inteligente que se utilizará; en la sección 4 se propone la arquitectura del ITS a partir del proceso de tutoría local de GIFT y de la inclusión de los profesores en los sistemas orientados a aprendizaje semipresencial. Finalmente, en la sección 5 se muestran las conclusiones y trabajo futuro.

2 Trabajos relacionados

Los cambios sociales y tecnológicos han llevado a la utilización de las TIC en entornos educativos, razón por la cual es importante estudiar la formación de los profesores en función de los cursos que orientar [13]. A continuación se presenta algunos trabajos relacionados con la importancia de los profesores y el uso de ITS en entornos presenciales para apoyar la instrucción impartida.

Steenbergen-Hu y Cooper [14] llevaron a cabo un meta–análisis enfocado a analizar la efectividad de los sistemas de tutoría inteligente en entornos de educación superior. El trabajo demuestra que los ITS facilitan el aprendizaje en los estudiantes universitarios, debido a su capacidad para superar varios métodos de instrucción o actividades de aprendizaje. Los autores concluyen que los ITS pueden ser utilizados como herramienta de instrucción independiente o como herramienta de aprendizaje complementario, además de evidenciar la importancia que tienen los profesores y la pedagogía en el aprendizaje mediante ITS.

Por su parte, Nye [7] realizó una revisión sistemática de literatura entre los años 2009 y 2012, orientada al análisis de las posibles barreras de la adopción de los ITS en relación con estudiantes, profesores y sistema escolar. Concluye que se ha prestado mayor atención a las barreras concernientes a los estudiantes, dejando de lado las demás. La adopción de los ITS en entornos de aprendizaje formal requiere unificar la comunicación entre profesor, administración e incluso padres. Adicionalmente, fortalecer el lazo ITS-profesores es dar un paso a la adopción y efectividad de estos entornos de aprendizaje.

Vanlehn [15] hace un estudio del estado del arte sobre tutorías guiadas, tanto por humanos como por sistemas de tutoría, y refleja que los humanos son más eficaces. Pero que estos últimos están muy cerca de la eficacia de la tutoría humana al tener una mínima diferencia. El autor define nueve teorías para explicar qué hacen los tutores humanos que los sistemas de tutoría no, y a lo que atribuye la efectividad de estos frente a esos sistemas: 1) evaluaciones y diagnósticos más detallados y acertados de los estudiantes; 2) selección de tareas individuales; 3) estrategias de tutorías más sofisticadas; 4) control de la interacción por parte del estudiante; 5) conocimiento del dominio; 6) aumento de la motivación de los estudiantes; 7) retroalimentación por parte de los estudiantes; 8) ejecución y cooperación en los procesos de tutoría por parte del estudiante y el tutor; 9) clasificar a los estudiantes de acuerdo a su comportamiento para definir y aplicar estrategias de tutoría.

Olney et al. [16] presentan un ITS basado en el diálogo, llamado Guru, para el área de biología en secundaria. Guru involucra al estudiante mediante conversaciones colaborativas, donde se muestra información relevante para la conversación. Adicionalmente, adapta cada sesión con base en los niveles de conocimientos individuales de los estudiantes, analiza sus respuestas y proporciona retroalimentación formativa. Los autores presentan los resultados de aprendizaje en la escuela mediante una comparación entre Guru, tutoría humana y enseñanza en clase. A pesar de que no hay diferencias entre el efecto de Guru y los tutores humanos, el primero cuenta con la capacidad de enseñar individualmente.

Por su parte, Dolenc y Aberšek [17] realizan el diseño y evaluación de un sistema inteligente de tutoría llamado TECH8, que se adapta de forma individual con base a las características cognitivas del estudiante. Este sistema fue diseñado de manera modular, incluyendo características que lo diferencian de otros, tales como el contenido relacionado con el enfoque cognitivo, la adaptabilidad y la recogida de metadatos y variables. Los autores realizaron pruebas con una muestra de 177 estudiantes, donde obtuvieron un mejor rendimiento en comparación con la enseñanza tradicional, pero que no supera la enseñanza personalizada. Los autores concluyen que este tipo de sistemas puede apoyar los conocimientos impartidos de manera presencial.

Las conclusiones de estos autores evidencia la importancia de los profesores en los procesos de formación mediante el uso de sistemas de tutoría inteligente, así como el de los ITS para el fortalecimiento de los temas impartidos de forma presencial. Por esta razón es importante proponer una arquitectura para un ITS que incluya al profesor, tanto en los procesos de formación de los estudiantes, como en la mejora en el uso de las herramientas de aprendizaje electrónico.

3 *Framework* generalizado para tutoría inteligente

El Framework Generalizado de Código Abierto para Tutoría Inteligente (GIFT) es un proyecto realizado por el Laboratorio de Investigación del Ejército de los Estados Unidos, con el objetivo de desarrollar herramientas, métodos y estándares para ITS [18]. GIFT provee una arquitectura modular [19] que brinda: 1) herramientas para el desarrollo de nuevos ITS; 2) gestor de instrucción que utiliza los mejores principios de tutoría, estrategias y tácticas; y 3) Plataforma experimental que permite analizar los componentes, herramientas y métodos [18].

GIFT está centrado en el estudiante, con el objetivo de mejorar iteractivamente la cadena de efectos del aprendizaje de tutoría adaptativo [19]. Esta cadena de efectos (Figura 1) parte de la comprensión del comportamiento, rasgos y preferencias del estudiante, obtenidos a partir de su desempeño, encuestas y sensores, los cuales permiten determinar de manera más precisa sus estados, dando como resultado un mejor modelo [20]. Posteriormente, se necesitan métodos que clasifiquen con precisión los estados del estudiante, para mejorar la capacidad de adaptación y seleccionar las estrategias instruccionales con base en ellos [20]. Esto genera una probabilidad más alta de influir positivamente en los avances del aprendizaje, al tener las estrategias de enseñanza acordes a las necesidades de los estudiantes [20].

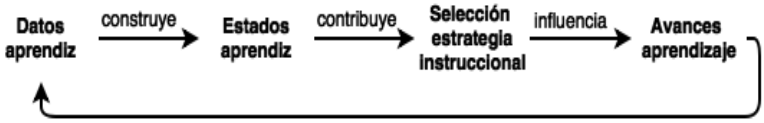


Figura 1: Cadena de efectos del aprendizaje de tutoría adaptativo [19]

Dentro de los elementos funcionales de GIFT se encuentran componentes, módulos, modelos, mensajes, bases de datos, bibliotecas e interfaces, que soportan procesos de creación, instrucción y análisis, como se observa en la Figura 2 [20]. GIFT presenta contenido genérico como documentos y contenidos multimedia, así como contenido especializado típicamente a través de un sistema software externo, llamado aplicación de entrenamiento (TA) [22].

GIFT provee un estándar para la integración de aplicaciones de formación o entrenamiento, incluyendo diversas herramientas y servicios necesarios para su

transformación dentro de un sistema de tutoría inteligente y/o adaptativo [22]. Entre los servicios y estándares se encuentran:

- Estandarización para la interconexión de aplicaciones de formación o entrenamiento
- Representación del conocimiento de dominio (incluye herramienta de creación)
- Evaluación del desempeño
- Secuencia del curso (incluye herramienta de creación)
- Modelo pedagógico, que incluye tanto la adaptación micro como la macro
- Modelado del estudiante
- Soporte de encuestas (con herramientas de creación)
- Sistema de gestión de aprendizaje
- Estandarización para la integración de sensores fisiológicos y otros

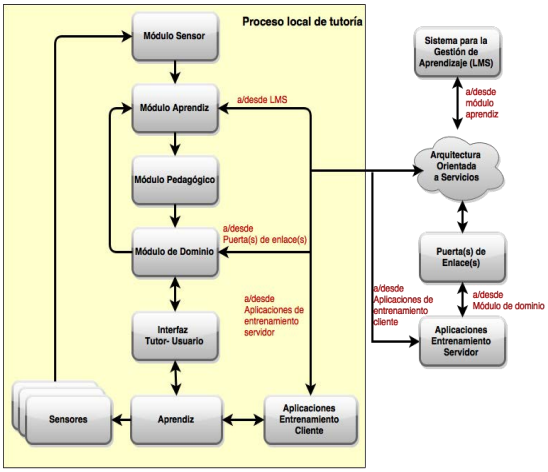


Figura 2: Diagrama funcional de GIFT [21]

4 Arquitectura propuesta

Existen diversos ITS [5, 23, 24] cuyos enfoques, arquitecturas y tecnologías pueden variar de acuerdo al tema, las necesidades y los usuarios. Sin embargo, para proponer la arquitectura se tuvo en cuenta algunos componentes que utiliza GIFT en el proceso de tutoría local, con base en el amplio uso en proyectos orientados al aprendizaje virtual. La arquitectura propuesta (Figura 3) busca apoyar los procesos de formación en las diferentes instituciones educativas presenciales, dándoles a los profesores la posibilidad de reforzar algunos temas en estudiantes que presentan dificultades en el aprendizaje. Adicionalmente, los profesores llevan a cabo una evaluación del manejo de la plataforma de aprendizaje que determina sus falencias, buscando mejorarlas a partir de capacitaciones adaptadas a cada uno. El objetivo es lograr que los docentes orienten mejor a los estudiantes en el uso de la herramienta de aprendizaje y de esta forma motivarlos en el uso de las mismas.

4.1 Módulo usuario [20, 22, 25]

Este módulo provee una interfaz para que el usuario (estudiantes o profesores) interactúe con el sistema. Es un módulo de aprendizaje apoyado por las valoraciones del estado cognitivo del usuario, teniendo como base los datos históricos y los datos de las encuestas. También predice el estado futuro del usuario a partir de una evaluación de probabilidades conmutadas de niveles de rendimiento más altos o bajos del estado actual del usuario. Los insumos utilizados para realizar el cálculo del estado provienen de fuentes tales como evaluación de desempeño de las aplicaciones de aprendizaje enviadas

desde el módulo de dominio, respuestas a la encuesta y los rasgos que se han venido almacenando a largo plazo en el LMS. Al no existir un modelo de estudiante estándar para todos los cursos, en este módulo se considera un esquema flexible que permita cambios según los requerimientos [26]. A futuro, un modelo estándar permitirá integrar con facilidad las bases de datos de usuarios externos, como los sistemas de gestión de aprendizaje.

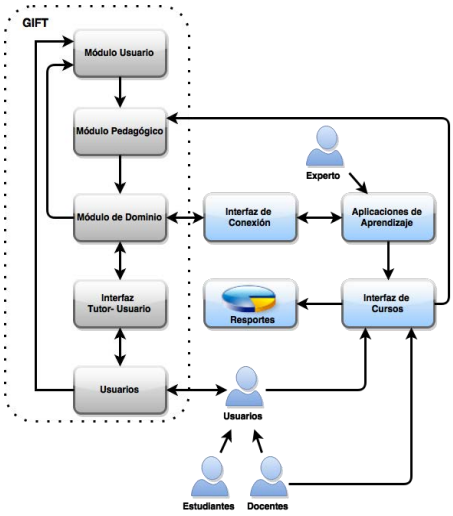


Figura 3: Arquitectura propuesta para ITS

4.2 Módulo de dominio [20, 22, 25]

Este módulo tiene como propósito definir y estructurar el conocimiento a partir de los datos obtenidos del entorno de formación y la evaluación de resultados con base en conceptos claves. La selección o desarrollo de conocimientos específicos debe estar soportado por los principios de diseño del sistema de enseñanza [25]. Se incluyen los siguientes elementos en el conocimiento de dominio específico: objetivos de aprendizaje, materiales, tareas, condiciones, normas de rendimiento, medidas, conceptos erróneos comunes, una librería de retroalimentación de contexto específico y preguntas [26]. Mediante el mejoramiento de los procesos para reutilizar el conocimiento de dominio específico existente se lograría una reducción significativa en el tiempo y costo de desarrollo.

4.3 Módulo pedagógico [20, 22, 25]

Utiliza los datos del módulo de usuario y el módulo de dominio para determinar el orden, el contenido y el flujo de la instrucción. El estímulo visual y auditivo se hace a través de las recomendaciones de este módulo. Las estrategias pedagógicas toman decisiones a la hora de actuar sobre una serie de opciones, logrando manipular los elementos del escenario de aprendizaje, es decir, se toman decisiones pedagógicas a partir del rendimiento y estado del usuario. Este módulo influye en el entorno de formación maximizando el aprendizaje individual de manera efectiva, a través de la información que se presenta al usuario por medio del sistema.

4.4 Interfaz tutor- usuario [20, 22, 25]

Presenta al estudiante y al docente las interfaces con las que debe interactuar para llevar a cabo tareas tales como: conexión del usuario, encuestas, retroalimentación, revisiones después de la acción, diálogos interactivos, presentación del material de

aprendizaje y evaluaciones tanto de desempeño como de manejo de la plataforma de aprendizaje. Para apoyar las estrategias de instrucción, el profesor puede tener acceso a los datos del estudiante, provenientes del modelo de usuario, o de datos en tiempo real a través de las interfaces profesor–usuario. Estas interfaces utilizan técnicas de Inteligencia Artificial para entender y generar respuestas en lenguaje natural.

4.5 Usuarios

Aquí se encuentran los datos históricos asociados al estado cognitivo del estudiante y del profesor, colectados a partir de los objetivos de aprendizaje propuestos. Adicionalmente, se encuentran los datos de las encuestas asociadas a cada usuario y la evaluación del manejo de la plataforma en el caso de que el usuario sea un profesor.

4.6 Aplicaciones de aprendizaje

Contine todos los objetos de aprendizaje asociados al ITS para la consecución de los objetivos de aprendizaje, los cuales son sugeridos por el ITS a partir del estilo de aprendizaje y del desempeño del estudiante en el transcurso del curso. Dentro de las aplicaciones se pueden mencionar las siguientes:

- Serius games
- SCORM
- PowerPoint
- Simuladores

El Ministerio de Educación Nacional de Colombia, con el apoyo de algunas instituciones de educación superior, propone que los objetos de aprendizaje estén constituidos principalmente por tres componentes internos [27]:

- *Contenidos.* Tipos de conocimiento con sus diferentes maneras de representación, entre los que se pueden mencionar: definiciones, explicaciones, documentos de lectura, enlace a otros recursos, vídeos, entre otros.
- *Actividades de aprendizaje.* Orientan al estudiante o profesor con el fin de alcanzar los objetivos de aprendizaje propuestos para un determinado curso.
- *Elementos de contextualización.* Componente que permite la reutilización de los objetos de aprendizaje en otros escenarios.

Estos componentes serán tenidos en cuenta para llevar a cabo las aplicaciones de aprendizaje, las cuales serán utilizadas por los diferentes cursos propuestos por los profesores.

4.7 Interfaz de cursos

Módulo donde los profesores elaboran los diferentes cursos, propuestos tanto para estudiantes como para otros profesores, y tomando como principal de insumo los objetos de aprendizaje desarrollados en el módulo de aplicación de aprendizaje. Adicionalmente, se ingresan los parámetros iniciales para los diferentes procesos de formación dentro del módulo pedagógico, entre las que se destacan:

- Expresar los objetivos de aprendizaje de manera explícita, relacionado con lo que el estudiante o profesor va a aprender.
- Relacionar cada objetivo de aprendizaje con un tipo de conocimiento específico.
- Evaluación del estudiante de acuerdo con las características predeterminadas.

4.8 Interfaz de conexión

Conecta el GIFT con las aplicaciones de aprendizaje realizadas por el especialista. Este módulo debe tener en cuenta funcionalidades adicionales que se puedan requerir en determinados casos, tales como recepción de mensajes de estado para recibir o transmitir tanto en GIFT como en la aplicación de aprendizaje, de forma que sean entendidas por cualquiera de estos.

4.9 Reportes

Se muestran los avances y las mejoras obtenidos con la utilización de la herramienta de aprendizaje por parte de los usuarios, motivándolos a continuar para lograr los objetivos propuestos en cada curso. Adicionalmente, los profesores pueden acceder a un historial de los avances de los estudiantes para compararlos con los indicadores obtenidos en los exámenes realizados en el salón de clases, alcanzando de esta forma una coherencia de resultados entre la herramienta y la clase presencial

5 Conclusiones y trabajo futuro

Las tecnologías de la Información y la Comunicación han tomado fuerza en los últimos años en los entornos de aprendizaje, debido a los diferentes mecanismos de accesibilidad que poseen las personas. Es importante que la mayoría de los profesores aprendan y manejen adecuadamente estas tecnologías, para poder motivar y guiar a sus estudiantes en todo el proceso de formación, logrando la culminación exitosa de los objetivos pedagógicos específicos propuestos.

Este trabajo presenta una arquitectura de ITS que vincula a los profesores dentro de un sistema de tutoría inteligente, buscando el fortalecimiento de sus habilidades en el manejo de herramientas tecnológicas orientadas a la educación. De esta manera se consolida como un insumo para la construcción de nuevos ITS enfocados tanto en los estudiantes como en los profesores.

Actualmente se trabaja en la implementación de un ITS soportado en esta arquitectura para los cursos de matemáticas del colegio San Agustín y la Fundación Universitaria de Popayán.

Agradecimientos

Este trabajo se enmarca dentro del Proyecto "Modelado del docente en sistemas adaptativos de aprendizaje para el mejoramiento de la calidad de la educación en matemáticas". El proyecto está financiado por el Sistema de Investigación, Desarrollo e Innovación de la Fundación Universitaria de Popayán, en la convocatoria 001 de 2014.

Referencias

[1] Brophy, J. (2013). Motivating students to learn. London: Routledge.
[2] Weld, D. et al. (2012). Personalized online education - A crowdsourcing challenge. In Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence, pp. 1-31.
[3] Hwang, G. et al. (2012). Development of a personalized educational computer game based on students learning styles. Educational Technology Research and Development 60(4), pp. 623–638.
[4] Woolf, B. (2010). Building Intelligent Interactive Tutors: Student-centered strategies for revolutionizing e-learning. USA: Morgan Kaufmann.
[5] D'Mello, S. et al. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. International Journal of Human-Computer Studies 70(5), pp. 377–398.
[6] Graesser, A.; Conley, M. & Olney, A. (2012). Intelligent tutoring systems. In Major, S. & Swanson, H. (Eds.), APA educational psychology handbook, Vol 3: Application to learning and teaching, pp. 451-473.

[7] Nye, B. (2014). Intelligent tutoring systems by and for the developing world: a review of trends and approaches for educational technology in a global context. International Journal of Artificial Intelligence in Education 25(2), pp. 177–203.
[8] Ong, J. & Ramachandran, S. (2000). Intelligent tutoring systems: The what and the how. Online Learning Circuits [May 2014].
[9] Nye, B. (2014). Barriers to ITS adoption: A systematic mapping study. Lecture Notes in Computer Science 8474, pp. 583-590.
[10] Hennessy, S.; Harrison, D. & Wamakote, L. (2010). Teacher factors influencing classroom use of ICT in Sub-Saharan Africa. Itupale Online Journal of African Studies 2(1), pp. 39–54.
[11] Ordonez, A. (2014). Including user devices and context into STEM education for minorities. Proceedings Integrated STEM Education Conference, pp. 1–5.
[12] Faghih, B. et al. (2013). User Interface Design for E-Learning Software. International Journal of Soft Computing and Software Engineering 3(3), pp. 786-794.
[13] Ramírez, D. & González, A. (2012). Modelo de acción docente con medios informáticos y telemáticos. Pixel-Bit: Revista de Medios Y Educación 40, pp. 151–170.
[14] Steenbergen, S. & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. Journal of Educational Psychology 106(2), pp. 331–347.
[15] Vanlehn, K. (2011). The relative effectiveness of human tutoring, intelligent yutoring systems, and other tutoring systems. Educational Psychologist 46(4), pp. 197–221.
[16] Olney, A. et al. (2012). Guru: a computer tutor that models expert human tutors. Lecture Notes in Computer Science 7315, pp. 256-261
[17] Dolenc, K. & Aberšek, B. (2015). TECH8 intelligent and adaptive e-learning system: Integration into Technology and Science classrooms in lower secondary schools. Computers & Education 82, pp. 354–365.
[18] Lane, H. et al. (2013). Artificial Intelligence in Education. International Conference, AIED 2013. Memphis, USA.
[19] Sottilare, R. & Holden, H. (2013). Recommendations for authoring, instructional strategies and analysis for intelligent tutoring systems (ITS): towards the development of a generalized intelligent framework for tutoring (GIFT). Proceedings AIED 2013 Workshops 7, pp. 1-150.
[20] Sottilare, R. et al. (2012). The generalized intelligent framework for tutoring (GIFT). U.S. Army Research, Development and Engineering Command, pp. 1–12.
[21] Goldberg, B. & Cannon-Bowers, J. (2013). Experimentation with the generalized intelligent framework for tutoring (GIFT): a testbed use case. Proceedings AIED 2013 Workshops 7, pp. 27–36.
[22] Ragusa, C., Hoffman, M. & Leonard, J. (2013). Unwrapping GIFT: a primer on authoring tools for the generalized intelligent framework for tutoring. In Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium, pp. 11.
[23] Maciejewski, A. & Leung, N. (2013). The nihongo tutorial system: an intelligent tutoring system for technical Japanese language instruction. CALICO Journal 9(3), pp. 5–25.
[24] Regalado, R. et al. (2015). Salinlahi III: an intelligent tutoring system for Filipino language learning. In ACL-IJCNLP 2015, pp. 87.
[25] Goldberg, B. et al. (2012). Use of evidence-based strategies to enhance the extensibility of adaptive tutoring technologies. In The Interservice/Industry Training, Simulation ¬ Education Conference.
[26] Sottilare, R. (2012). Considerations in the development of an ontology for a generalized intelligent framework for tutoring. In International Defense & Homeland Security Simulation Workshop.
[27] Colombia aprende (2014). Objetos virtuales de aprendizaje e informativos: ¿Qué es un objeto de aprendizaje? Ministerio de Educación Nacional.

Sistema para caracterización de paneles fotovoltaicos

Carlos Robayo L.
Sergio Serna G.
Instituto Tecnológico Metropolitano – Colombia

1 Introducción

Las fuentes de energía alternativa han tomado un gran auge en los últimos tiempos, fundamentalmente las pilas de combustible, los aerogeneradores, los paneles fotovoltaicos, entre otros [1-10]. Las pilas de combustible basadas en hidrógeno tienen como ventajas la ausencia de polución y la predictibilidad de la energía almacenada en el combustible, lo que las hace ideales para aplicaciones móviles, tales como automóviles o electrónica de consumo. Por el momento, sus deficiencias son el suministro de hidrógeno y la respuesta a las perturbaciones en la carga [6, 11-14]. Los aerogeneradores se usan principalmente en aplicaciones estacionarias, tales como residenciales aisladas o con conexión a la red. Estos dispositivos se encuentran en un amplio rango de potencias y son usados principalmente en grandes granjas eólicas conectadas directamente a la red, que requieren de amplias extensiones de campo abierto y presentan riesgos debido a su tamaño y peso. Otros sistemas eólicos de baja potencia se utilizan en sectores urbanos o en instalaciones aisladas de los sistemas eléctricos [1, 4, 8, 10, 15, 16].

La fuente de energía de interés para este trabajo son los paneles fotovoltaicos, que son principalmente usados en aplicaciones residenciales y de inyección a la red. Estos sistemas necesitan áreas extensas debido a su eficiencia de conversión (aproximadamente 12% - 14%) y su potencia máxima es altamente sensible al sombreado parcial de la superficie. Los principales retos que debe afrontar esta tecnología tienen que ver con la optimización de su punto de trabajo frente a perturbaciones ambientales y de carga, así como la minimización del efecto de las sombras [17-24].

Se han realizado varios trabajos para caracterizar paneles fotovoltaicos, como el de Yang y Chen [25], quienes utilizan un modelo de celda solar construido en LabVIEW [26] y se basan en las características de salida de un panel de silicio, simulando la potencia solar con diferentes puntos de radiación, resistencia interna y factor de forma del modelo. Los resultados verifican la característica no-lineal de salida de estas celdas y confirman que la resistencia interna es un factor clave para el punto de máxima potencia y para el factor de forma de las celdas solares. Ming, Zhizhang y Zhang [27] describen el efecto del viento en la eficiencia de los paneles solares y utilizan programación en LabVIEW para comprobar los resultados del cálculo teórico con el análisis comparativo.

Li y Zheng [28] investigan las características V/I y V/P de las celdas solares y los módulos fotovoltaicos cuando operan con sombreado parcial. Examinan la relación entre las propiedades semiconductoras y las exigencias eléctricas externas; además, grafican las características de extracción de potencia del sistema. Concluyen que bajo condiciones de sombra irregulares los paneles presentan múltiples picos en sus características fotovoltaicas. Magsi, Macdonald y Shterengas [29] estudian dos tipos de tecnologías de celdas solares: el estándar de la industria basada en silicio y el emergente basado en nuevos materiales de tintes sensibilizados. Diseñan experimentos y realizan mediciones utilizando LabVIEW y una tarjeta de adquisición de datos. Los resultados formulan perspectivas para mejorar la eficiencia de la nueva tecnología. Por último, Ma et al. [30]

construyen un sistema con LabVIEW para describir las curvas características V/I y calcular los parámetros característicos básicos de los paneles. Los resultados demuestran que el sistema puede medir con precisión todos los parámetros de la celda solar.

Estos autores no realizan la caracterización automática de paneles solares reales, mientras que en este trabajo se presenta un sistema para la caracterización de paneles fotovoltaicos que permita determinar de forma automática las curvas V/P, I/P y V/I de paneles fotovoltaicos de silicio, de esta manera se puede comprobar la veracidad de las hojas de datos proporcionadas por los fabricantes y cuantificar con mayor precisión los parámetros de los controladores diseñados para las pruebas de laboratorio. El trabajo se organiza de la siguiente manera: se presentan los elementos teóricos básicos que fundamentan el modelo eléctrico del panel, luego la metodología y los componentes empleados en el desarrollo del trabajo, y posteriormente los resultados y conclusiones.

2 Fundamentación de la caracterización

Dos elementos básicos se deben considerar al momento de caracterizar los paneles: la radiación solar y la temperatura. El rendimiento de los paneles se puede evaluar con el factor de forma.

2.1 Radiación solar

Es la energía electromagnética emitida por los procesos de fusión del hidrógeno contenido en el sol. La intensidad y la distribución espectral de esta radiación que llega a la superficie terrestre dependen de la composición de la atmósfera. En el suelo la energía solar se encuentra dentro del intervalo de longitud de onda de 0.2 a 2.5 micrómetros distribuidos así: de 0.2 a 0.38 micrómetros es la denominada luz ultravioleta que contiene aproximadamente el 6.4% de la energía total en el espectro; de 0.38 a 0.78 micrómetros está la luz visible que aporta aproximadamente el 48% de la energía total del espectro; de 0.78 a 10 micrómetros está el infrarrojo, que contiene el restante 45.6% de la energía. La zona útil para las celdas fotovoltaicas empieza en 0.35 micrómetros y comprende todo el infrarrojo. La intensidad máxima de la energía solar está comprendida entre 0.5 y 0.55 micrómetros.

2.2 Temperatura

Es importante tener en cuenta la temperatura en el momento de la toma de muestras, porque la generación de energía en los paneles dependa de ella. En el modelo eléctrico de un panel solar se encuentra, entre otros elementos, un diodo (Figura 1) cuya corriente depende de la temperatura, como se relaciona en las ecuaciones (1) y (2).

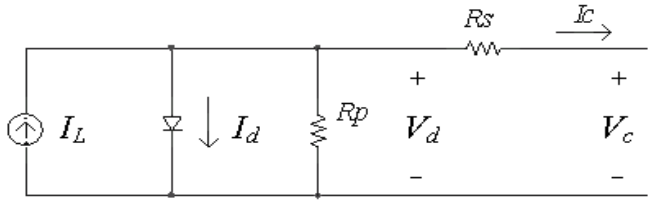


Figura 1: Modelo de panel solar

$$I_C = I_L - I_d\left(e^{\frac{qV_d}{mkT}} - 1\right) - \frac{V_d}{R_p} \qquad (1)$$

$$V_C = V_d - I_C \cdot R_S \qquad (2)$$

Donde I_L es proporcional a la intensidad de iluminación de la luz solar, m es el factor de idealidad del diodo (1 para un diodo ideal), Id es la corriente de saturación inversa que depende de la temperatura, q es la carga elemental, k es la constante de Boltzmann y T es la temperatura absoluta [28].

2.3 Factor de Forma

El Factor de Forma (FF) de una celda o panel solar es el cociente entre la potencia real (V_{pmax} x I_{pmax}) de las celdas solares frente a la salida de potencia en corto-circuito (V_{oc} x I_{sc}) (Figura 2), y es un parámetro clave para evaluar el rendimiento de las celdas solares. Las celdas solares típicas tienen un factor de forma mayor 0.70. Se deben rechazar las celdas con bajo FF (celdas de grado B) que tienen un factor de forma por lo general del 0.4 al 0.65, lo que las hace ineficientes.

$$FF = \frac{I_{pmáx} \times V_{pmáx}}{I_{SC} \times V_{OC}} \qquad (3)$$

El FF es otro término para definir la eficiencia de una celda solar (4).

$$\eta = \frac{FF \times I_{SC} \times V_{OC}}{P_m} \qquad (4)$$

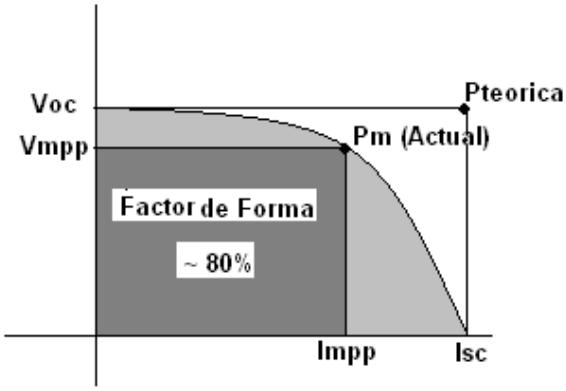


Figura 2: Área efectiva del Factor de Forma

Donde η es la eficiencia de conversión y P_m es la potencia máxima real del panel. Las celdas analizadas hasta el momento con el sistema que se describe en este artículo tienen un FF con valores entre 0.51 y 0.81 y una eficiencia entre 6% y 10.2%.

3 Metodología del proyecto

El proyecto se llevó a cabo en el Laboratorio de Fuentes de Energía Renovable del Instituto Tecnológico Metropolitano [31], que cuenta con los equipos de medida y de potencia, paneles solares y software, necesarios tanto para este proyecto como para otros enfocados en las fuentes de energía renovable y que ha sido escenario de trabajos de investigación de Ingeniería, Maestría y Doctorado de la Universidad Nacional de Colombia y el Instituto Tecnológico Metropolitano, y se proyecta como alternativa de capacitación e investigación en el tema para toda la región.

El diagrama de despliegue del sistema de caracterización de paneles fotovoltaicos puede verse en la Figura 3, donde se identifican cinco grandes bloques: los paneles fotovoltaicos, los sensores de voltaje y corriente, la carga electrónica, la tarjeta de adquisición de datos y el computador con la aplicación LabVIEW.

3.1 Dispositivos y equipos

La plataforma requiere equipos de generación, paneles fotovoltaicos, equipos de medida, sensores y tarjeta de adquisición de datos, y equipos para prueba y carga electrónica.

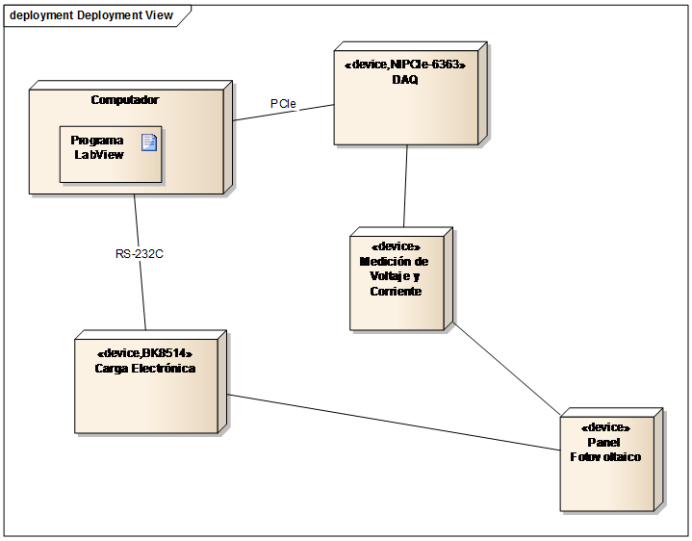


Figura 3: Diagrama de despliegue del sistema de caracterización de paneles fotovoltaicos

- *Paneles solares fotovoltaicos*. Los paneles fotovoltaicos utilizados son Hybrytec FV8512 [32] de características prácticamente iguales al clásico BP585, esto es: 18.75 V a máxima potencia, 22.16 V en circuito abierto, 4.8A a máxima potencia, 5.15 A en corto circuito y 13.04% de eficiencia.
- *Carga Electrónica*. Como carga resistiva se hace uso del módulo de carga BK PRECISION 8514; La carga es un dispositivo de dos terminales que pueden conectarse a fuentes de DC. Este tipo de cargas electrónicas se utilizan para el diseño, la fabricación y la evaluación de las fuentes de alimentación, tales como baterías químicas, pilas de combustible, pruebas de celdas fotovoltaicas o cualquier elemento generador. La carga puede operar bajo cuatro diferentes modalidades constantes: voltaje, corriente, potencia y resistencia. Para la operación de este sistema la carga es controlada desde el software, que le indica los valores de corriente que debe consumir, desde 0A hasta un valor definido con una conmutación de milisegundos y con un total de muestras predefinidas. La comunicación se hace de forma remota a través de una interfaz serie (RS-232).
- *Tarjeta de adquisición de datos (DAQ).* Se utiliza la tarjeta DAQ NIPCIe-6363 [33] del mismo fabricante del software LabVIEW, National Instrument, y que es completamente compatible con él, facilitando la captura y procesamiento de datos.

3.2 Software del sistema

LabVIEW es un entorno de desarrollo gráfico utilizado por la industria y la academia en la adquisición de datos y el control de instrumentos por software. Con la integración de LabVIEW, el hardware y las funciones de adquisición de datos, el sistema puede dibujar las curvas V/I, V/P y I/P automáticamente, manipulando la carga para calcular los parámetros de rendimiento de los paneles. La Figura 4 muestra el flujo de datos del sistema de caracterización de paneles.

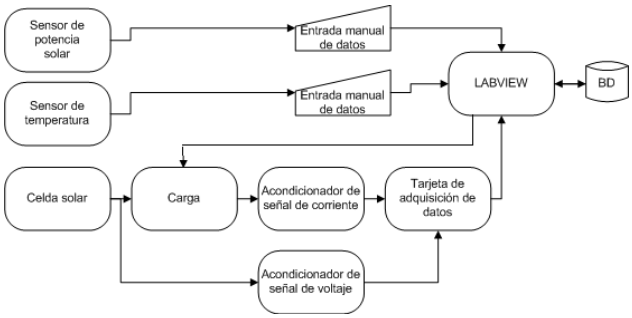


Figura 4: Esquema de flujo de datos

3.3 Algoritmo de caracterización de paneles

El diagrama de flujo de la Figura 5 representa los pasos que se deben seguir, ya sea para consultas de caracterizaciones pasadas o para ingresar datos nuevos. El algoritmo evidencia la flexibilidad de la plataforma a la hora de caracterizar los paneles, porque permite almacenar la información con todas las características necesarias para su posterior recuperación.

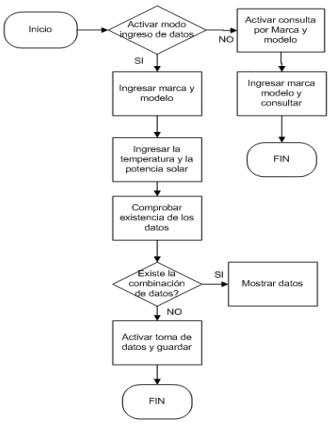


Figura 5: Algoritmo de caracterización complejo

4 Resultados y Conclusiones

La interfaz del aplicativo de la plataforma puede verse en la Figura 6, en la que se puede seleccionar el puerto donde está conectada la carga y se puede observar que se identifican los campos para ingresar las condiciones de prueba, tales como la cantidad de muestras y el incremento de corriente, la radiación solar y temperatura, así como los valores de corriente de corto circuito, voltaje de circuito abierto y potencia nominal del panel, esto con el fin de configurar la carga electrónica para que se proteja de problemas de funcionamiento de la plataforma. Igualmente, tiene campos donde se muestra la corriente, el voltaje medido y la potencia calculada. Esta información se organiza en una tabla y con ella se construye una gráfica de voltaje contra corriente. A través de un botón el usuario puede o no guardar la información.

Los resultados obtenidos con la utilización de la solución en una celda fotovoltaica Hybrytec FV8512 se guardan en la base de datos y luego se llevan a Matlab para crear las gráficas de caracterización que se muestran en la Figura 7. Se observan curvas típicas dadas por los fabricantes de paneles y en concreto para este panel según [32], y con un distanciamiento de no más del 5% en la información obtenida por el aplicativo contra los proporcionados por la hoja de datos. Por lo anterior, la plataforma presenta resultados

satisfactorios, que en caso de ser necesario pueden ser mejorados con sensores más precisos y de mayor ancho de banda.

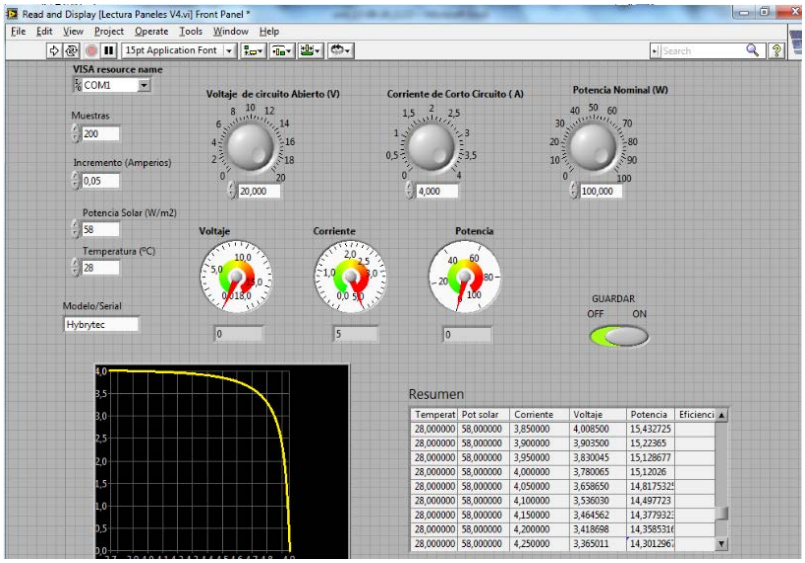


Figura 6: Interfaz del aplicativo en LabVIEW

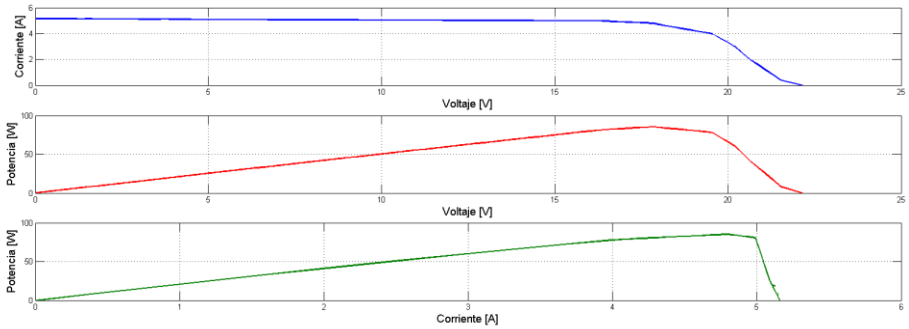


Figura 7: Graficas de caracterización del panel

Agradecimientos

Este trabajo se realizó en el marco de los proyectos de investigación P14-215 y P14-220 del grupo de investigación en Automática, Electrónica y Ciencias Computacionales del Instituto Tecnológico Metropolitano. Igualmente, hace parte de la tesis de maestría PM13102.

Referencias

[1] Drew, D. et al. (205). The impact of future offshore wind farms on wind power generation in Great Britain. Resources 4(1), pp. 155–171.

[2] Drude, L.; Pereira, C. & Rüther, R. (2014). Photovoltaics (PV) and electric vehicle-to-grid (V2G) strategies for peak demand reduction in urban regions in Brazil in a smart grid environment. Renewable Energy 68, pp. 443–451.

[3] Reid, G. & Wynn, G. (2015). The future of solar power in the United Kingdom. Energies 8(8), pp. 7818–7835.

[4] Ribeiro, A.; Arouca, M. & Coelho, D. (2016). Electric energy generation from small-scale solar and wind power in Brazil: The influence of location, area and shape. Renewable Energy 85, pp. 554–563.

[5] Sanz, E. et al. (2014). Production, consumption and research on solar energy: The Spanish and German case. Renewable Energy 68, pp. 733–744.

[6] Silva, S. et al. (2013). A stand-alone hybrid photovoltaic, fuel cell and battery system: A case study of Tocantins, Brazil. Renewable Energy 57, pp. 384–389.
[7] Villicaña, E. (2015). Solar energy potential in the coastal zone of the Gulf of Mexico. Renewable Energy 81, pp. 534–542.
[8] Waewsak, J.; Landry, M. & Gagnon, Y. (2015). Offshore wind power potential of the Gulf of Thailand. Renewable Energy 81, pp. 609–626.
[9] Zhao, H. & Guo, S. (2015). External benefit evaluation of renewable energy power in China for sustainability. Sustainability 7(5), pp. 4783–4805.
[10] Zhao, X. & Ren, L. (2015). Focus on the development of offshore wind power in China: Has the golden period come? Renewable Energy 81, pp. 644–657.
[11] Iverson, Z. et al. (2013). Optimal design of hybrid renewable energy systems (HRES) using hydrogen storage technology for data center applications. Renewable Energy 52, pp. 79–87.
[12] López, J. et al. (2015). Environmental impact associated with the substitution of internal combustion vehicles by fuel cell vehicles refueled with hydrogen generated by electrolysis using the power grid. An estimation focused on the Autonomous Region of Murcia (Spain). Renewable Energy 77, pp. 79–85.
[13] Piumsomboon, P. et al. (2013). Road testing of a three-wheeler driven by a 5 kW PEM fuel cell in the absence and presence of batteries. Renewable Energy 50, pp. 365–372.
[14] Topriska, E. et al. (2015). Solar hydrogen system for cooking applications: Experimental and numerical study. Renewable Energy 83, pp. 717–728.
[15] Santos, F. et al. (2015). Combining wind farms with concentrating solar plants to provide stable renewable power. Renewable Energy 76, pp. 539–550.
[16] Ebrahimi, F.; Khayatiyan, A. & Farjah, E. (2016). A novel optimizing power control strategy for centralized wind farm control system. Renewable Energy 86, pp. 399–408.
[17] Hu, Y. et al. (2015). Three-Port DC-DC converter for stand-alone photovoltaic systems. IEEE Transactions on Power Electronics 30(6), pp. 3068–3076.
[18] Manganiello, P. et al. (2014). Optimization of perturbative PV MPPT methods through online system identification. IEEE Transactions on Industrial Electronics 61(12), pp. 6812–6821.
[19] Villa, L.; Raison, B. & Crebier, J. (2014). Toward the design of control algorithms for a photovoltaic equalizer: Detecting shadows through direct current sampling. IEEE Journal of Emergin and Selected Topics in Power Electronics 2(4), pp. 893–906.
[20] Cai, C. & Aliprantis, D. (2013). Cumulus cloud shadow model for analysis of power systems with photovoltaics. IEEE Transactions on Power Systems 28(4), pp. 4496–4506.
[21] Al, A. & Dhaouadi, R. (2012). Efficiency optimization of a DSP-Based standalone PV system using fuzzy logic and Dual-MPPT control. IEEE Transactions on Industrial Informatics 8(3), pp. 573–584.
[22] Ishaque, K. et al. (2012). An improved Particle Swarm Optimization (PSO)-Based MPPT for PV with reduced steady-state oscillation. IEEE Transactions on Power Electronics 27(8), pp. 3627–3638.
[23] Gao, L. et al. (2009). Parallel-Connected Solar PV System to address partial and rapidly fluctuating shadow conditions. IEEE Transactions on Industrial Electronics 56(5), pp. 1548–1556.
[24] Patel, H. & Agarwal, V. (2008). MATLAB-Based modeling to study the effects of partial shading on PV array characteristics. IEEE Transactions on Energy Converstion 23(1), pp. 302–310.
[25] Yang, G. & Chen, M. (2009). LabVIEW based simulation system for the output characteristics of PV cells and the influence of internal resistance on it. Proceedings WASE International Conference on Information Engineering, pp. 391–394.
[26] National Instruments (2015). LabVIEW Help. Online [Feb. 2014].
[27] Jie, M.; Zhizhang, L. & Qingzhu, Z. (2010). Solar photovoltaic panels wind load testing and analysis. Proceedings International Conference on Mechanic Automation and Control Engineering, pp. 1632–1635.
[28] Li, S. & Zheng, H. (2011). Energy extraction characteristic study of solar photovoltaic cells and modules. Proceedings IEEE Power and Energy Society General Meeting, pp. 1–7.
[29] Magsi, K.; Macdonald, E. & Shterengas, L. (2010). Calculating the efficiency of dye sensitized solar cells with industry based silicon solar cells. Proceedings Applications and Technology Conference, pp. 1–7.
[30] Ma, L. et al. (2011). The measurement of series and shunt resistances of the silicon solar cell based on LabVIEW. Proceedings International Conference on Electrical and Control Engineering, pp. 2711–2714.
[31] Serna, S. et al. (2012). Laboratorio de fuentes de energía renovable. Avances en Ingeniería Eléctrica 3, pp. 28–36.
[32] Hybrytec (2013). Paneles Solares Hybrytec. Medellín: Hybrytec energía solar.
[33] National Instruments (2014). DAQ X Series-X Series User Manual: NI 632x/634x/636x/637x Devices. Austin: National Instruments.

Dinámicas de inventarios en el sector productivo

Federico Liévano M.
Universidad Católica del Norte – Colombia

1 Introducción

Para ser competentes en los mercados locales y globales, las empresas necesitan mejorar el servicio al cliente y reducir sus costos. Una manera estratégica de abordar este desafío es mediante la construcción de modelos dinámicos que ayuden a comprender los comportamientos de los centros de almacenamiento e inventarios de las empresas productivas. Para así conocer de manera precisa la cantidad de productos que se deben fabricar, para satisfacer la demanda y establecer un sistema de inventarios que garantice un nivel de servicio al cliente eficiente y que disminuya la cantidad de ventas perdidas además de los costos [1].

Uno de los factores más importantes para el éxito de la planeación de la producción y la gestión de inventarios, es que el sistema permita predecir o conocer en el tiempo la cantidad suficiente a producir para satisfacer la demanda [2]. Este factor de éxito el que motiva a simular un sistema dinámico de producción e inventarios, con el propósito de estudiar diferentes escenarios que permitan identificar estrategias logísticas, que repercutan en cumplimento a los clientes, permanencia en el mercado, ahorros, entre otros aspectos. Para lograrlo se utilizó la dinámica de sistemas, debido a que es una herramienta que permite esclarecer las relaciones involucradas entre las variables más importantes del sistema, teniendo presente la evolución del mismo a lo largo del tiempo, algo que muy posiblemente con otro tipo de técnicas no es posible identificar [3]. La estructura del modelo propuesto y las conclusiones que surgen de este estudio pueden ayudar a comprender mejor la toma decisiones respecto de la gestión de inventarios, porque se podrán conocer con anticipación los impactos que la elección de ciertas preferencias y decisiones tienen sobre el sistema de logística y distribución.

2 Los sistemas de inventarios

El control y la administración de inventarios es uno de los procesos más importantes en todo ambiente organizacional, que compete cualquier cadena de suministros o distribución. Dada la competitividad de los actuales mercados y la apertura de nuevas conexiones empresariales, la facilidad de la transmisión de la información, la expansión tecnológica y los tratados económicos entre los países, obliga a que las organizaciones optimicen todos sus procesos y en particular que posean una cadena de suministro óptima y competitiva [4].

Si bien los problemas de inventarios son abordados desde múltiples metodologías, tales como la simulación de eventos discretos, la investigación de operaciones y la economía operativa, con los modelos clásicos como el EOQ y a partir de una minimización de la función de costos se obtiene la cantidad óptima de pedido y los puntos de re-orden, entre otros [5]. La complejidad que posee este tipo de sistemas amerita que se continúe desarrollando estudios generales que contribuyan al entendimiento de los mismos, dado que su impacto en la cadena de suministros y la cantidad de costos asociados en el proceso específico de operación son altas.

Asimismo, los sistemas de inventarios están altamente ligados a las actividades específicas de cada organización, lo cual hace complejo su modelamiento. La captura de información relevante dentro de los distintos procesos, las condiciones específicas de cada elemento de negocios o producto que se produce o comercializa en una cadena, hace que la generalidad en el control de inventarios sea un reto difícil de alcanzar. Además, la cantidad de variables que pueden estar asociadas a los procesos de administración o gestión de inventarios hacen que el sistema, en muchas ocasiones, se comporte de forma no lineal, se desconozcan las dinámicas internas y, lo más grave, que el sistema responda con comportamientos emergentes descontrolados que alteren toda la cadena y repercutan en altos costos operativos en las organizaciones. Esto amerita aproximaciones de modelamiento no-lineal que involucren estas dinámicas.

2.1 El papel del modelado en los sistemas de inventarios

Actualmente, la modelación de sistemas dinámicos es de amplia utilidad para comprender procesos organizacionales en cualquier nivel de agregación y entorno. Los sistemas pueden presentar diversos grados de complejidad, por esto es necesario observar a la empresa como un todo, para luego llevarla a un nivel más detallado, es decir, pasar de lo general a lo particular para realizar un análisis que permita entender su realidad.

Es común escuchar a los administradores, gerentes y analistas de logística afirmar que uno de los principales problemas a los que se enfrentan es a la administración de los inventarios. Con frecuencia en las organizaciones del sector industrial o comercial se presenta el problema de excesos y faltantes. Los excesos se presentan cuando las expectativas están por encima de la demanda, esto fundamentalmente no genera complicaciones, porque se satisface al cliente con respecto al cumplimiento de su requerimiento, pero si en los altos costos que generalmente tiene el almacenaje de productos y que inevitablemente afecta las utilidades en las ventas. Los faltantes pueden presentarse cuando las expectativas que resultan de un análisis de inventario están por debajo de la demanda. Este comportamiento, además de la reducción en la satisfacción de los clientes debida una disminución en el nivel de servicio, genera costos de oportunidad que pueden surgir por diferentes factores: compras fallidas, nuevas compras, o nuevos mercados.

El problema de planeación de la demanda es fundamentalmente un problema de análisis de la información del mercado y de cómo impacta toda la cadena de suministro. Generalmente, se enfoca o aborda mediante técnicas econométricas o de Inteligencia Artificial, que intentan realizar pronósticos de lo que ocurrirá con el mercado en el futuro. Muchas de estas técnicas se basan en modelos determinísticos, que involucran retrogresiones lineales y no-lineales intentando ajustar curvas de comportamiento. Otros intentan, por medio de técnicas computacionales, esclarecer qué factores afecta la dinámica de los mercados y la demanda y de esta forma anticiparse a las adversidades del futuro [6].

La demanda es un factor clave en todo modelo que intenta representar la dinámica de los inventarios de una empresa, en este estudio se asume una porción de demanda o de mercado que le pertenece inherentemente a la organización y que genera unas ventas promedio. Sin embargo, para representar la realidad de los mercados volátiles que se presentan en diferentes entornos, se incorporan escenarios de cambio de demanda, y de esta forma se modela el impacto que tienen en toda la estructura de planeación y de gestión de inventarios diseñada [7].

En lugar de realizar estimaciones y modelos para predecir el comportamiento de la demanda, los autores se enfocan más en las consecuencias que implican los cambios abruptos a nivel de inventarios y en la toma de decisiones a nivel productivo. Asimismo, se analiza cuándo una empresa no produce la cantidad suficiente para satisfacer la demanda, lo cual genera un impacto en el nivel de servicio, la participación en el mercado y los costos por faltantes y excesos.

Para desarrollar un modelo de inventarios es necesario comprender que este ocurre cuando el producto terminado es mayor a las entregas. Esta acumulación del producto se usa para satisfacer a tiempo la demanda de los clientes, brindándoles un mejor servicio [8]. La cantidad de productos que se tienen en inventario no es constante, debido a variables como el comportamiento de la demanda a través del tiempo, la capacidad de producción de la empresa, la calidad de los productos fabricados, los daños de producto y el tiempo de fabricación de los productos [9]. Por lo tanto, y teniendo en cuenta lo mencionado anteriormente, la gestión de inventarios realizada en una empresa determinada debe ser considerada como un sistema, porque todas las variables que se definen para su modelamiento se afectan mutuamente y si se altera alguna de ellas tiene repercusiones en todo el sistema. La administración y la toma de decisiones de inventarios es un tema complejo, debido a la cantidad de variables que influyen en su comportamiento y las variaciones que estas sufren continuamente.

Como resultado, se presenta una necesidad de desarrollar modelos de simulación para determinar dinámicas de este complejo sistema y con diferentes escenarios de demanda, y realizar una aplicación correcta de los datos históricos para explorar el efecto de la planeación en la cadena de distribución o suministros.

3 Modelo de simulación

3.1 Hipótesis dinámica

El modelo desarrollado en este estudio puede considerarse de abstracción alta, porque considera elementos muy puntuales y de detalle riguroso que han sido explorados por varios autores. Asimismo, se incluyen algunas variables aleatorias que permiten ver las diferentes fluctuaciones que presentan algunas variables incorporando incertidumbre y estocasticidad. El objetivo con el modelo es tratar de demostrar que el nivel de inventarios de las empresas no debe seguir un comportamiento oscilatorio de sube y baja constante, sino un patrón asociado a las dinámicas de la demanda y las interacciones entre las variables más significativas del sistema [10].

Para el modelo se asumen organizaciones que fabrican o producen algún bien no perecedero, usando un sistema MTS, (Make to stock), es decir, que producen para tener reservas en inventarios. Adicionalmente, todas las cifras económicas e impositivas que se asumen son regidas bajo los estándares colombianos y basados en la legislación colombiana. Las variables usadas en el modelo son las siguientes [11]:

- Inv actual: la cual indica el inventario actual en bodega
- Brecha: La diferencia entre el inventario actual y el deseado
- Producción: Indica el nivel de producción en planta
- Licitaciones: Se refiere a solicitudes por licitación y de carácter urgente
- Costos producción: costos asociados a la producción del bien
- Costos totales: Costos que incluyen no solo los de producción sino también almacenamiento
- Precios: Precios del producto
- Demanda: Pedidos de producto solicitados
- Ventas: Perdidos convertidos en ventas y dinero efectivo

- Inv. Deseado: Inventario deseado por los planeadores logísticos
- Porción del Mercado: participación del mercado por parte de la empresa
- Calidad: Calidad del producto
- Percepción costo-beneficio: percepción del valor del producto por el consumidor
- Percepción del dinero del empresario: percepción del flujo de caja de la empresa
- Costos inventarios y contingencias: costos asociados a almacenamiento y contingencias

En el ciclo de la Figura 1 izquierda se puede ver que tener inventario de más afecta los costos y precios de venta, que a su vez afectan la demanda y las ventas, haciendo que la percepción del dinero del empresario reduzca la cantidad de inventario deseado a disminuya la producción y el inventario actual. Asimismo, se observa como la cantidad de inventario que se tiene aumenta los costos de inventario y contingencia, afectando los costos totales que posteriormente afectarán los precios, la demanda y las ventas y haciendo que el inventario no se venda.

Figura 1: Ciclos de realimentación asociado al impacto en precios

En el ciclo de la Figura 2 se ve lo que sucede cuando se tiene suficiente inventario actual, lo que hace que la brecha de lo que se necesita sea menor reduciendo producción y costos, lo que influye en el precio de venta que mejora la percepción costo beneficio, y de esta forma vendiendo más y reduciendo el inventario mismo. Asimismo, se ilustra que mientras más se tenga inventario actual menor va a ser la brecha con la producción, al tener una menor brecha la producción se reduce y al reducirse la producción disminuye de igual manera el inventario actual, generando así un ciclo de balance.

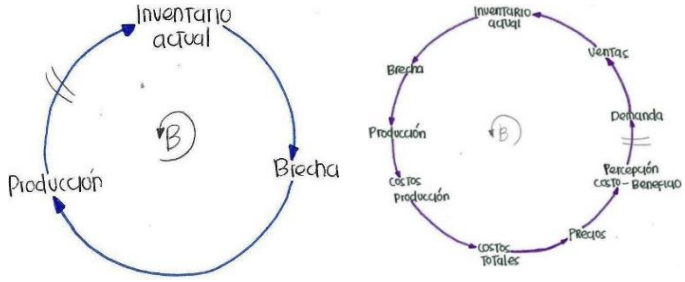


Figura 2: Ciclos de realimentación asociados al impacto en producción

Al tener un mayor inventario actual se tiene un mayor costo de inventario y contingencias, como se observa en la Figura 3, aumentando así los costos totales y dando como resultado un incremento de los precios finales de los productos, afectando de manera negativa la percepción costo-beneficio, disminuyendo la demanda y reduciendo las ventas y la percepción del dinero del empresario, lo que reduce el inventario deseado, reduciendo así la brecha y disminuyendo la producción, pero afectando de manera negativa el inventario actual. Cuando se tiene un mayor inventario actual los precios de los productos finales disminuirán, aumentando de esta forma la percepción costo-beneficio por parte de los compradores e igualmente su demanda, lo cual conlleva a un

aumento directo de las ventas totales aumentando la percepción del dinero del empresario, lo que se ve reflejado en un mayor inventario deseado y generando una mayor brecha. Para el cubrimiento de esta brecha se requiere un aumento de producción, lo que termina en un inventario actual o mayor.

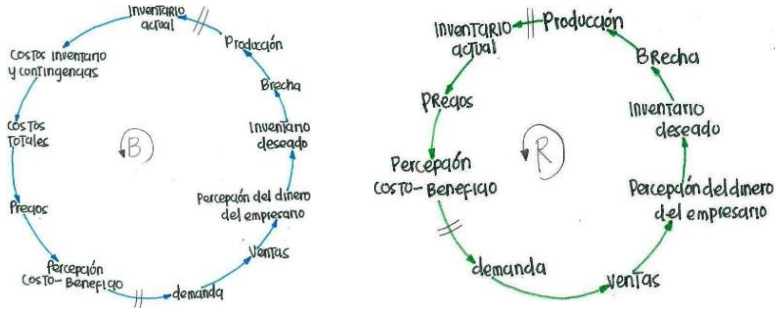


Figura 3: Ciclos de realimentación asociados a impactos en el valor del producto

Un inventario actual mayor genera una menor brecha, lo que disminuye de manera directa la producción, como se observa en la Figura 4, disminuyendo sus respectivos costos y por ende los costos totales. Los precios finales disminuyen debido al bajo costo de producción, lo cual conlleva a un aumento en la demanda y de la misma manera se aumentan las ventas y el inventario actual se disminuye debido a una mayor cantidad de ventas. Cuando el inventario actual es mayor los precios disminuyen, lo que aumenta su demanda y respectivamente sus ventas, reduciendo de manera directa el inventario actual.

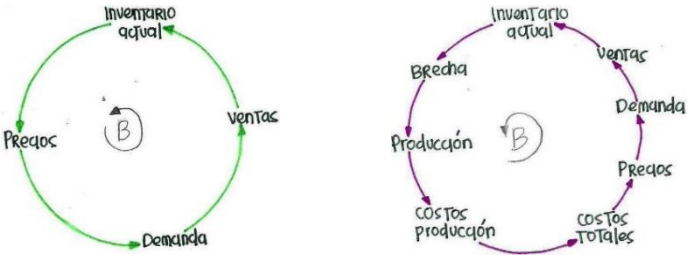


Figura 4: Ciclos de realimentación asociados al impacto en ventas

Finalmente, el ciclo de la Figura 5 muestra como al momento de hacer los descuentos o liquidaciones de inventario la percepción del costo-beneficio aumenta haciendo que haya más demanda y ventas, reduciendo el inventario actual. En este ciclo se puede observar que la cantidad del inventario afecta los costos de inventario y contingencia, que a su vez afectan costos totales y los precios haciendo que la percepción de costo-beneficio se vea reducida, lo que reduce la demanda y a su vez la ventas haciendo que el inventario no se venda.

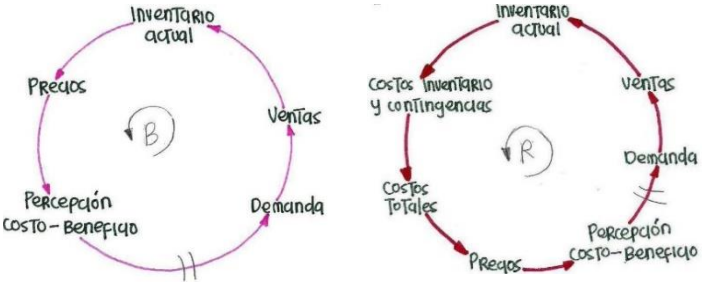


Figura 5. Ciclos de realimentación asociados al impacto en la demanda

La Figura 6 corresponde al diagrama causal del sistema completo para el objeto de estudio.

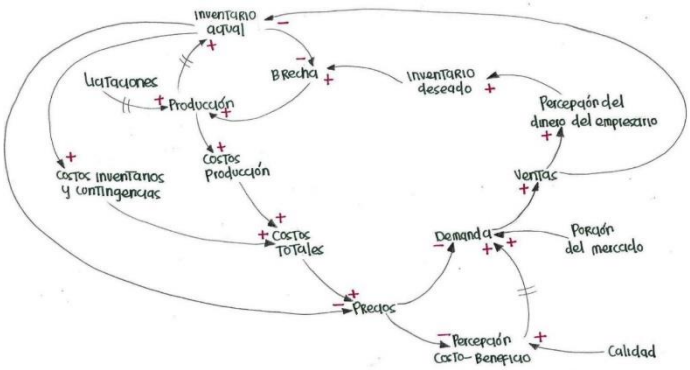


Figura 6: Diagrama Causal completo

3.2 Diagrama de flujos y niveles

Para el diagrama de flujos y niveles se incorporaron nuevas variables que aportaron en el modelamiento y en el diseño conceptual estas se listan a continuación [12]:

- Porción del Mercado: que hace alusión a la porción de Mercado que tiene la empresa y por ende el índice de ventas
- Ventas deseadas: es el nivel de ventas que desea o espera el administrador
- Retardo ventas: Es el tiempo de retraso entre las ventar reales y esperadas
- Producción especial: Es la producción orientada a negocios especiales
- Negocios especiales: Son los negocios con prioridad como licitaciones o urgencias
- Demanda mercado total: Es la demanda que se efectúa en todo el Mercado
- Descuento: es el precio de descuento por volumen de ventas
- Precio Unitario: Es el precio final del producto
- Esfuerzos Marketing: Dinero invertido en marketing y publicidad
- Costos almacenamiento: costos de almacenamiento
- Costos totales: costos que involucran todos los factores
- Costo unitario: costo de producción del producto
- Costo cotizado: costo discriminado por producto para establecer el precio
- % gastos fijos: Gastos fijos como arriendo o servicios públicos
- Gastos no variables: Gastos variables asociados al producto
- Rentabilidad esperada e impuesto de renta

El modelo de flujos y niveles se ilustra en la Figura 7. Allí se muestra como se relacionan todas estas variables a través de la lógica de flujos y niveles. También se observa que solamente se ha modelado un nivel principal asociado al inventario real de la organización. Las dinámicas y cambios de esta variable están relacionados con todas las interacciones de las demás variables. Las ecuaciones formuladas para la construcción del modelo son:

Nivel_de_inventarios(t) = Nivel_de_inventarios(t - dt) + (Tasa_de_producción + Tasa_de_producción__especial - Tasa_de_ventas_normales - Ventas_especiales) * dt
INIT Nivel_de_inventarios = 600
Tasa_de_producción = (IF Brecha>10 THEN Vble_puente ELSE 0)
Tasa_de_producción__especial = Variable_Puente_neg_esp*0.3
Tasa_de_ventas_normales = if Brecha<20 then Ventas_deseadas else Retardo_ventas
Ventas_especiales = IF Variable_Puente_neg_esp>10 THEN Variable_Puente_neg_esp *0.03 ELSE 0
Brecha = Ventas_deseadas-Nivel_de_inventarios

Costos_de_almacenamiento = normal(350,25)*Nivel_de_inventarios
Costos_totales = Tasa_de_ventas_normales*Costo_unitario+Costos_de_almacenamiento
Costo_cotizado = Costo_unitario*(1+Porcentaje_de_gastos_fijos * 0.05) + Gastos_operacionales_variables
Costo_unitario = normal(800,45)
Demanda_Mercado_Total = normal(1300,200)-0.03*Precios_unitario
Descuento = if Nivel_de_inventarios>300 then 0.15/exp(-1+300/LN(Nivel_de_inventarios)) else 0.25
Esfuerzos_de_marketing = IF Utilidad>2000000 THEN 1 ELSE IF Utilidad<2000000 AND Utilidad>1500000 THEN 0.5 ELSE 0
Gastos_operacionales__variables = normal(1750,200)
Negocios_especiales = if Esfuerzos_de_marketing=1 then normal(200,40) else if Esfuerzos_de_marketing=0.5 then 125 else 0
Porcentaje_de_gastos_fijos = normal(0.4,0.02)
Porcion_del_mercado = 0.7
Precios_unitario = Costo_cotizado/((1-Descuento)*(1-Rentabilidad_esperada_e_impuesto_de_renta))
Producción_especial = if Brecha<0 and Negocios_especiales<abs(Brecha) then 0 else Negocios_especiales
Rentabilidad_esperada_e_impuesto_de_renta = 0.25/(1-0.35)
Retardo_ventas = delay(Nivel_de_inventarios,0.5,0)
Utilidad =Precios_unitario*Tasa_de_ventas_normales-Costos_totales * (1 - Tasa_de_ventas_normales-Costos_totales)
Variable_Puente_neg_esp = delay(Producción_especial,3,0)
Vble_puente = delay(Brecha,6,0)
Ventas_deseadas = Demanda_Mercado_Total*Porcion_del_mercado * 0,05

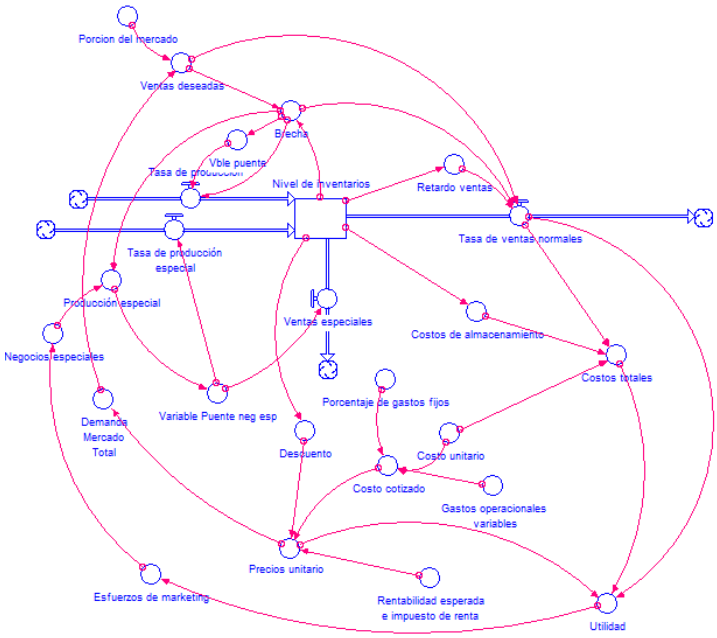


Figura 7: Diagrama de Flujos y Niveles para la simulación del sistema

4 Análisis de resultados

Para el análisis se consideran dos variables de interés: el nivel de inventarios y las utilidades. Los resultados se obtuvieron de la simulación del modelo anterior usando como unidad de tiempo semanas, para un periodo de 54 semanas, es decir, 1 año. La Figura

8 representa el movimiento de los inventarios durante el periodo señalado. Se realizó una prueba base con parámetros establecidos por la empresa *Fourier SAS* con el inventario de dispositivos tecnológicos; los resultados son muy cercanos a los presentados en sus datos históricos, donde se pudo observar un patrón cíclico alrededor de 15 semanas por trimestre. Las variables aleatorias incluidas intentaron modelar las percepciones de los administradores de inventarios y los cambios volátiles del mercado, lo cual acercó mucho los resultados a la realidad.

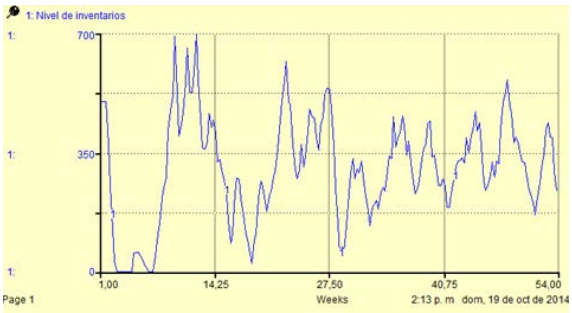


Figura 8: Simulación de nivel de inventarios en empresa Fourier SAS

La Figura 9 muestra las utilidades. Como se puede observar está altamente correlacionada al nivel de inventarios. En algunos periodos de tiempo se presentan utilidades negativas, esto es debido al retardo en las ventas y los costos de oportunidad por los bajos niveles de inventarios. También se nota que si el inventario inicial fuera 0 o menor entonces el retardo en ventas haría que las utilidades disminuyeran.

Figura 9: Utilidades del proceso productivo, almacenamiento y ventas

La Figura 10 muestra que las tasa de producción y las decisiones productivas son altamente dependientes del nivel de inventarios, lo cual hace que las expectativas de los administradores logísticos cambien y se apresuren a tomar decisiones sobre el sistema productivo.

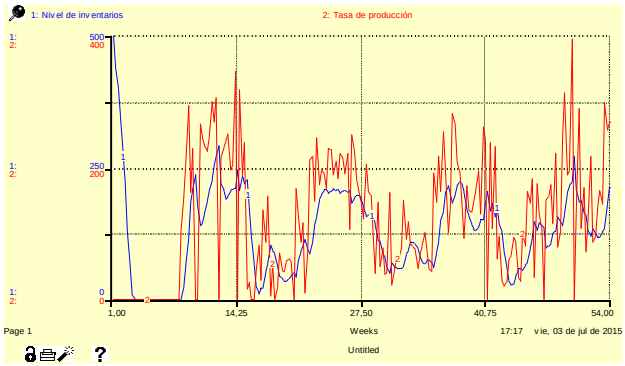


Figura 10: Inventarios versus decisiones productivas

En la Figura 11 se observa el impacto que tienen las licitaciones o negocios especiales sobre los inventarios. En el primer escenario se observa el comportamiento con una tasa de negocios especiales periódicos alta, que provienen con una probabilidad determinada. En el segundo escenario se disminuye la probabilidad de negocios especiales y se observa que las licitaciones son un factor determinante en la dinámica de los inventarios. Estos hacen que aumenten el nivel de inventarios en el tiempo y también que cambie la periodicidad de los ciclos.

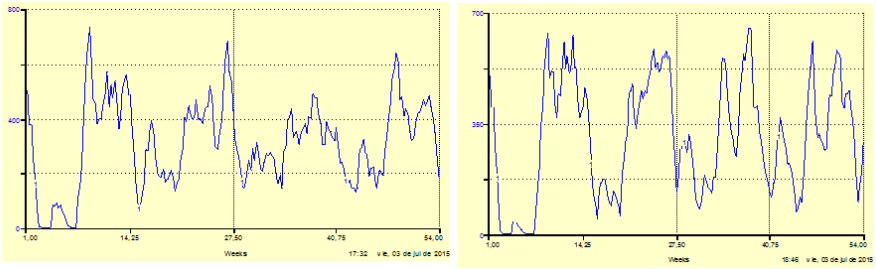


Figura 11: Efecto de las licitaciones y negocios especiales sobre los inventarios

Finalmente, en la Figura 12 se ilustra el impacto de los retardos en las ventas, demostrando lo sensibles que son los inventarios a retardos en ventas y lo importante de tener una fuerza comercial activa y efectiva que produzca ventas tangibles y rápidas. La imagen de la derecha muestra un nivel de inventarios con un patrón oscilatorio de 12 semanas, y crítico debido a un alto retardo en las ventas, lo cual hace que en el tiempo no se cuente con suficiente producto en las bodegas y que disminuya la capacidad productiva. Con un retardo mayor la escases de productos es menor y se mantiene el nivel de inventario más estable oscilando en un ideal de 350 unidades.

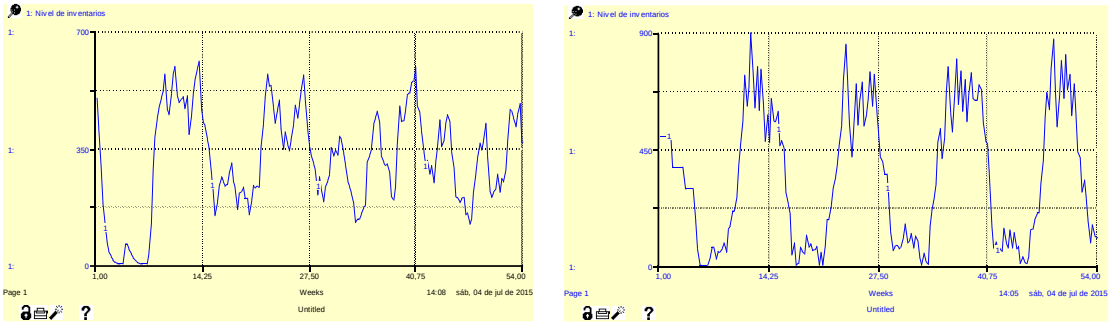


Figura12: Efecto de los retardos de las ventas en los inventarios

5 Conclusiones

En este trabajo se estudia la potencialidad que tiene la dinámica de sistemas para formular y analizar problemas relacionados con el manejo y control de sistemas de inventarios. Los resultados de diversos modelos de simulación demuestran que es una alternativa para enfrentar problemas de administración y gestión. También se plantea una hipótesis dinámica para afrentar el complejo sistema de inventarios, y la problemática de tener una administración adecuada del mismo bajo diferentes condiciones de mercado y operación.

Los resultados demuestran que la dinámica de sistemas es efectiva para la toma de decisiones y para comprender las dinámicas, a veces incomprensibles, en los sistemas de inventarios porque permite la formulación de modelos dinámicos que representan el proceso involucrando diferentes variables de mercado y de operación [13]. A partir de la

formulación de distintos escenarios de operación y el estudio realizado, se identificaron las siguientes dinámicas en el sistema de inventarios:

- Al realizar pruebas con los parámetros de una empresa real las dinámicas presentadas son similares a los comportamientos históricos de registro de inventarios de la empresa Fourier SAS, donde se realizó la validación.
- Las dinámicas presentadas en este modelo son oscilatorias con alta variabilidad asociada a la estocasticidad de la demanda, pero altamente sensibles a los retardos de las ventas y las licitaciones que la compañía está dispuesta atender.
- La productividad y las utilidades de una organización son altamente dependientes de la correcta gestión de inventarios y la posibilidad de mantenerlos bajos, pero satisfaciendo siempre la demanda y las ventas de los clientes.
- Un incremento ostensible en los costos de la materia prima afecta directamente la producción y las ventas, dado el encarecimiento del producto y la incapacidad para sostener la producción a largo plazo.
- El inventario presenta un patrón oscilatorio asociado a los retardos en toda la cadena de producción, lo cual hace que las percepciones de los administradores varíen de forma oscilatoria sobre el ideal de ventas y de stock.
- Tener un inventario costoso conlleva a disminución significativa del punto promedio de inventario y en múltiples puntos de re-orden. La estrategia más apropiada que enseña el modelo es a reducir los retardos en las ventas.

Referencias

[1] Thornassey, S. et al. (2002). Three complementary sales forecasting models for textile distributors. Proceedings International Conference on Systems, Man and Cybernetics.

[2] Lodding, H. & Lohmann, S. (2010). INCAP – Applying short-term flexibility to control inventories. International Journal of Production Research 50(3), pp. 909–919.

[3] Chan, F. & Prakash, A. (2011). Inventory management in a lateral collaborative manufacturing supply chain: A simulation study. International Journal of Production Research 50(1), pp. 4670–4685.

[4] Baganha, M. & Cohen M. (1998). The stabilizing effect of inventory in supply chains. Operations Research 46(3), pp. S72-S73.

[5] Singh, S. & Panda, G. (2015). An inventory model for generalized weibull deteriorating items with price dependent demand and permissible delay in payments under inflation. Scientific Journal of Logistics 11(3), pp. 259-266.

[6] Cardona, D. et al. (2001). Análisis de un sistema de inventarios en una PYME de confección. Memoris II Encuentro Colombiano de Dinámica de Sistemas, pp. 102–109.

[7] Harris, J. & Swatman, P. (1997). Efficient Consumer Response (ECR): A survey of the Australian grocery industry. Supply Chain Management: An International Journal 4(1), pp. 137-148.

[8] Feng Y. (2009). Study on model of supply chain inventory management based on system dynamics. Proceedings International Conference on Information Technology and Computer Science, pp. 209-212.

[9] Giannoccaro, I. & Pontrandolfo, P. (2002). Inventory management in supply chains: A reinforcement learning approach. International Journal Production Economics 78(2), pp. 153-161.

[10] Vergin, R. (1998). An examination of inventory turnover in the Fortune 500 industrial companies. Production and Inventory Management Journal 39(1), pp. 51-56.

[11] Vo, H. et al. (2007). Developing unbounded systems thinking: Using causal mapping with multiple stakeholders within a Vietnamese company. The Journal of the Operational Research Society 58(5), pp. 655-668.

[12] Alia, S. et al. (2013). Inventory management of perishable products: A time decay linked logistic approach. International Journal of Production Research 51(13), pp. 3864–3879,

[13] Wei, L. et al. (2015). A comparative simulation study about supply chain inventory management strategies based on system dynamics. Applied Mechanics and Materials 701-702, pp. 1280-1285.

Métricas del software - Herramientas de apoyo cuantificables para la toma de decisiones

Piedad Metaute P.
Alexei Serna A.
Corporación Universitaria Remington – Colombia

1 Introducción

En este documento se describe qué es medir, medida, medición, métricas del software e indicadores, así como la importancia que tienen las métricas en el software. Además, se presentan métodos utilizados para la medición, especialmente en los procesos relacionados con la Programación Orientada por Objetos POO; algunas herramientas utilizadas; su eficiencia para contribuir en la calidad del producto; y expectativas que se generan con la norma ISO/IEC 25000 SQuaRE - System and Software Quality Requirements and Evaluation, especialmente en lo relacionado con ISO/IEC 2502n [3] acerca de la medición de la calidad del producto.

Cuando se habla de métricas del software es importante tener claros los conceptos sobre medir, medida y medición: *medir* es comparar una cantidad con su respectiva unidad, con el fin de averiguar cuántas veces la primera contiene la segunda, es igualar y comparar una cosa no material con otra [4]; *medida* es la cantidad que resulta de medir una longitud, hecho a propósito con unas dimensiones determinadas [5]; por su parte, *medición* consiste sustancialmente en una observación cuantitativa, atribuyendo un número a determinadas características o rasgos del hecho o fenómeno observado, donde la medición, se refiere generalmente a la asignación de números a observaciones, de modo que sean susceptibles de análisis por medio de manipulaciones u operaciones de acuerdo con ciertas reglas [7]. En relación con el software, IEEE define que medición es una medida del grado en el que un sistema, componente o proceso posee un atributo determinado [8]. Por su parte, las *métricas del software* son un buen medio para entender, monitorear, controlar, predecir y probar el desarrollo del software y los proyectos de mantenimiento [9].

Roger Pressman [11] argumenta que se debe medir para evaluar los modelos de análisis y diseño, para tener indicios de la complejidad de los diseños procedimentales y del código fuente y para facilitar el diseño de las pruebas del software. De igual forma, hace referencia a que medir sugiere un proceso compuesto por cinco actividades: 1) formulación, que consiste en derivar medidas y métricas del software apropiadas al que se está construyendo; 2) recolección, se relaciona con acumular datos para derivar las métricas formuladas; 3) análisis, se realiza el cálculo de métricas con la ayuda de herramientas matemáticas; 4) interpretación, se evalúa la métrica resultante para comprender la calidad de la representación; y 5) retroalimentación, donde se realiza recomendaciones del resultado de la interpretación de la métrica y se socializa con el equipo de trabajo. Además, sugiere para estas actividades hay que tener presente los siguientes principios:

- Automatizar en lo posible la recolección y el análisis de datos
- Aplicar análisis estadístico que permita establecer relaciones entre los atributos internos y las características de calidad externas
- Establecer lineamientos y recomendaciones para cada métrica

Para abordar las métricas del software es necesario hacer alusión al indicador [12], definido como una métrica o una combinación de métricas que proporciona conocimientos acerca del proceso del desarrollo del software, un proyecto de software o el propio producto. Un indicador proporciona conocimientos que les permiten a los ingenieros ajustar el proceso, el proyecto o el producto buscando mejoramiento. El indicador permitirá cuantificar conceptos medibles, ofreciendo información para la toma de decisiones. En la Figura 1 se observa el proceso de recopilación de indicadores del software.

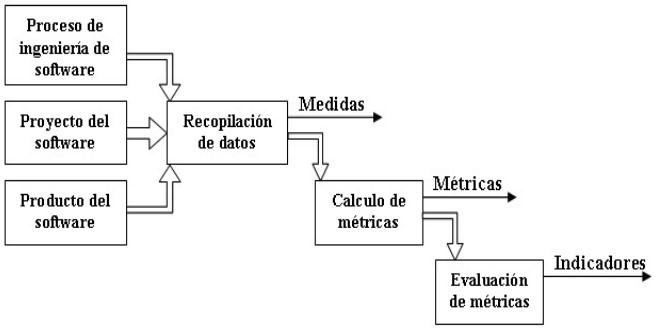


Figura 1: Indicadores del software [20]

Por lo tanto, para los procesos de la Ingeniería del Software es importante la aplicación de métricas en cada una de las fases, buscando con ello el fortalecimiento en la calidad del producto. Es así como ISO/IEC 25000 [14] tiene como objetivo principal guiar el desarrollo de los productos software mediante la especificación de requisitos y evaluación de características de calidad, como se observa en la Figura 2.

Figura 2: ISO/IEC 25000 (SQuaRE) [14]

ISO/IEC 25000 [14] evoluciona desde ISO/IEC 9126, que describe las particularidades de un modelo de calidad de productos software, e ISO/IEC 14598, que aborda el proceso de evaluación de productos software, convirtiéndose en la norma (*SQuaRE)*.

En relación con las métricas del software, ISO/IEC 25000 define la medición de la calidad del producto, definiciones de medidas de calidad (interna, externa y en uso) y guías prácticas para su aplicación [3]. Actualmente, esta división se encuentra conformada por un modelo de referencia común a los elementos de medición de la calidad, y un conjunto recomendado de métricas base y derivadas que puedan ser usadas a lo largo de todo el ciclo de vida del desarrollo. Además, define específicamente las métricas para realizar la medición de la calidad en uso del producto; las métricas para realizar la medición de la calidad de productos y sistemas software; y las métricas para realizar la medición de la calidad de datos.

Esta actualización es un aporte importante a la temática relacionada con las métricas del software, que en definitiva buscan evitar retrasos en el desarrollo de los proyectos, reducir los problemas inherentes a su proceso de construcción, detectar riesgos potenciales, evaluar la calidad de cada uno de sus procesos, establecer planes de mejoramiento, generar líneas base óptimas, motivar al aprendizaje continuo, fortalecer la productividad individual así como en los diferentes trabajos colaborativos, realizar estimaciones adecuadas, suministrar datos estadísticos que permitan la optimización de recursos, la evaluación de la productividad, el control del proyecto, y las certificaciones de calidad, en las que a través del aprendizaje que generan las buenas prácticas se puedan diseñar nuevos instrumentos de medición, acordes con las necesidades de optimización, competitividad y prospectiva de la empresa desarrolladora de software.

2 Modelos, métodos y herramientas relacionadas con las métricas del software

2.1 Sobre modelos y métodos

Es común que se presente confusión entre términos como modelo y método, por lo tanto, se hace necesario realizar una revisión a dichos conceptos. En [17] se define modelo como una representación de un objeto, sistema o idea, de forma diferente al de la entidad misma. El propósito de los modelos es ayudar a explicar, entender o mejorar un sistema; un modelo de un objeto puede ser una réplica exacta de éste o una abstracción de las propiedades dominantes del objeto. RAE [18] lo define como un arquetipo o punto de referencia para imitarlo o reproducirlo; y en [5] se define como una manera de proceder estructurada y ordenada para obtener un resultado, conjunto de reglas prácticas.

Entre los modelos propuestos y sobre los cuales se han fundamentado las mediciones actuales se encuentra COCOMO, que se basa en la estimación, el esfuerzo, el tiempo utilizado y el tamaño del software [23]. La propuesta se ha venido perfeccionando desde su versión inicial en los años 80, y actualmente está en la versíon COCOMO II. Considerado como el modelo de estimación de costos más ampliamente utilizado en el mundo, es el preferido para la estimación del esfuerzo cuando no se tiene información histórica. Además, es el más documentado de los modelos de estimación de esfuerzo de las actividades de diseño, codificación, pruebas y mantenimiento del software. COCOMO II se compone de tres sub-modelos [19] para ofrecer mayor fidelidad en el proceso de planificación y diseño de proyectos, y se denominan composición de aplicaciones, diseño temprano y modelos de post-arquitectura. El modelo se fundamenta en estimaciones matemáticas, donde para las respectivas mediciones es importante el tamaño del proyecto, representado en líneas de código. En la Figura 3 de describe la metodología COCOMO II.

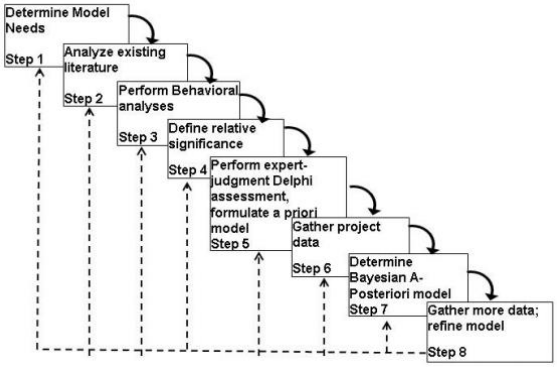


Figura 3: Metodología modelo COCOMOII [24]

Con la generalización de la Programación Orientada por Objetos [24] para el análisis, el diseño y el desarrollo de sistemas, se extendió el uso de la técnica Casos de Uso, que si bien no es primitiva de la orientación a objetos, es claramente preponderante. Desde su surgimiento demostró ser eficiente para la estimación, como una técnica que tiene coincidencias con la metodología utilizada. Los puntos por Casos de Uso clasifican los casos, de manera similar a puntos de función, en simples, medios y complejos según la cantidad de transacciones.

En la Ingeniería del Software se han abordado modelos y métodos alrededor de las métricas del software, algunos suficientemente argumentados desde el punto de vista teórico y otros con aplicación práctica. La mayoría cubre de forma general su uso, aunque se continúa replanteando la temática a través de propuestas que buscan que la medición se asuma como herramienta esencial en los diferentes procesos de construcción de software. Es trabajo se hace alusión especialmente a modelos y métodos propuestos en los procesos de la POO.

Las métricas proporcionan la comprensión necesaria para crear modelos efectivos de requisitos y diseño, código sólido y pruebas amplias [11]. En las orientadas a objetos se aplican métodos direccionados a las características de clase y de diseño (localización, encapsulación, ocultamientos de información herencia y técnicas de abstracción de objetos), que las hacen únicas. Otras consideran varias características de calidad, tales como cohesión, acoplamiento, complejidad, tamaño, reutilización, mantenimiento, entre otras [2]. La cohesión es el grado de relación entre elementos de módulo, donde mientras más alta es mejor. El acoplamiento es el grado de relación entre los módulos, es decir, la conexión, donde es deseable que sea bajo. La complejidad son los esfuerzos necesarios para analizar el código, además de los necesarios en la modificación de los módulos. La mantenibilidad es la capacidad de modificación del producto software después de la entrega [2]. La reutilización se refiere a cuánto se puede usar el módulo nuevamente, para reducir la codificación.

Varios investigadores han realizado estudios a este respecto, por ejemplo, Chidamber y Kemerer [16] utilizaron un grupo de métricas de acoplamiento para estimar la complejidad de los sistemas orientados a objetos, tales como RFC, CBO, DIT y NOC. Fenton y Pfleeger [16] demostraron que las métricas del software, tales como la complejidad, el acoplamiento y la cohesión, se podrían utilizar para evaluar su calidad. Badri [16] propuso indicadores multidimensionales para el acoplamiento de captura entre las clases, utilizando diferentes atributos internos del software, como la complejidad. Yadav y Khan [16] propusieron una métrica de complejidad de acoplamiento que minimiza la complejidad mediante el control del acoplamiento en los productos software. Misra [16] propuso una nueva métrica, Código Cognitivo de la Complejidad (CCC), que se basa en la herencia y la estructura de control básico. Observó que está métrica se relaciona con y está correlacionada con RFC, DIT, NOC, CBO y WMC, lo que muestra que los indicadores de métricas están relacionados entre sí, especialmente en términos de la información que recogen, porque proporcionan información útil para el mejoramiento de los procesos de calidad.

En POO, las métricas son estrategias necesarias para mejorar los procesos de construcción del producto. Tarcísio y Kecia [13] presentan un conjunto de datos estadísticos sobre los parámetros de medición del software de sistemas orientados a objetos. Los datos fueron generados a partir de la Qualitas.class Corpus y reunieron una amplia cantidad de datos de métricas de los sistemas incluidos. También aplicaron EasyFit a un número de distribuciones de cada conjunto de datos, que a su vez proporcionaron la

mejor distribución estadística de ajuste, así como el Ranking T (Transact-SQL). Disponían también de un volcado de base de datos MySQL para normalizar las métricas y facilitar la manipulación de datos, tales como el filtrado y la agregación. Al hacer disponible el conjunto tenían la intención de ayudarles a los investigadores en su trabajo sobre métricas.

El modelo de evaluación consiste típicamente en correlacionar el valor de la métrica con otros indicadores de calidad, tales como la probabilidad de cambio o su capacidad para predecir la presencia de errores. En otros casos, la evaluación es un análisis de los valores de una métrica para un conjunto de sistemas de código abierto, ya sea para una actividad o durante un período de tiempo dado [6]. Otros enfoques teóricos de la evaluación se basan en inspeccionar las propiedades matemáticas de las métricas o se centran en las propiedades metodológicas de las mismas, así como en el equilibrio de componentes y perfiles de dependencia, diseñados para cuantificar la capacidad de análisis y la encapsulación de datos dentro de un sistema software. La evidencia de la validez de constructos de estas métricas se ha reunido previamente en experimentos a pequeña escala, aplicando la evaluación en una simulación de construcción de software.

Como resultado de estos experimentos se procedió al análisis de los datos recogidos, y que produjo dos tipos de resultados [6]. En el primero se identificaron las situaciones y condiciones en la que las métricas son útiles. En el segundo se descubrió cómo mejorar las métricas en sí mismas y la manera de aplicarlas mejor. Además de informar sobre la evaluación de estas métricas específicas en este contexto particular, se reflexiona sobre un método general para la evaluación de métricas del software en un entorno práctico. Los retos implicados en el diseño y la ejecución de un estudio de este tipo de métricas, se deben describir y discutir con la posibilidad de generalizar los resultados [6].

La construcción de modelos de predicción de defectos es un método para utilizar métricas de software recogidas durante el ciclo de vida y el mantenimiento [1]. Estos modelos pueden guiar a los encargados del control de calidad directamente hacia módulos con más probabilidades de ser defectuosos, mejorando la calidad del software mientras se disminuyen los costos. Además, es uno de los favoritos de los profesionales en software por trabajar con subconjuntos de características más pequeñas, como por ejemplo datos de entrada, de salida, conformación de la clase, comunicación interna, entre otros, porque se puede entender directamente cómo contribuye cada característica a la calidad del software, permitiendo recopilar las métricas más importantes en los esfuerzos futuros [1]. Por lo tanto, la selección de características es un componente integral en la aplicación de la métrica, porque puede afectar de manera significativa la interpretación de los resultados. De ahí la importancia de manejar el pre-procesamiento de datos, debido a que es un paso crítico para el proceso de modelado, donde se debe tener cuidado en la eliminación de redundancias y el manejo que se le dé a los atributos menos relevantes.

Es importante mencionar a la Programación Orientada a Aspectos (POA), que se basa en la POO [26], es una extensión de ella y que busca soportar la separación de competencias para los Aspectos (propiedades de un software que tienden a atravesar sus funcionalidades principales) en el diseño e implementación del software. Esta nueva metodología busca que el software sea de fácil desarrollo y mantenimiento debido a la separación de competencias entre sus componentes y los aspectos. Estas unidades no son totalmente independientes, porque existen puntos de enlace que permiten su conexión, donde el tejedor, ayudado por los puntos de enlace, logra la mezcla entre componentes y aspectos.

POA enfoca y proporciona una forma eficiente para la separación de preocupaciones transversales (caché, localización, seguridad, puesta en común de recursos, sincronización, control de excepciones, etc.). A través de ella se reduce el sistema con código difuso [2], además, facilita entender la funcionalidad del núcleo. También pueden ser aumentados la reutilización y el mantenimiento. De igual forma POA proporciona un aspecto que contiene código ágil y las instrucciones sobre cuándo, dónde y cómo invocar el aspecto. Con la ayuda de lenguajes tales como AspectJ, AspectC, AspectC ++, AspectXML, CasearJ, Hyper/J, se implementan como aspectos las preocupaciones transversales.

Las métricas definidas por Chidamber y Kemerer [16] para POO (WMC, DIT, NOC, CBO, RFC), son aplicables a POA, donde la unidad es el Aspecto en vez de la clase o el procedimiento. Baigorria y Montejano [26] proponen utilizar métricas desde el diseño, aplicando las extensiones realizadas a UML en relación a la OA (Orientación a Aspectos), y en lo referente a: la relación entre clases y aspectos, clase tejida, puntos de enlace, conocimiento de mediciones como cantidad de aspectos, cantidad de relaciones entre aspectos y una clase, cantidad de clases relacionadas con un mismo aspecto, cantidad de puntos de enlace en una clase, cantidad de clases tejidas, reutilización de una clase, cantidad de puntos de enlace con la reutilización de la clase, cantidad de relaciones existentes entre un aspecto y una clase con la reutilización de la misma y cantidad de clase tejidas con la cantidad de aspectos, siendo una propuesta que requiere ser analizada de forma objetiva y buscando la calidad en el diseño y el desarrollo.

2.2 Sobre herramientas para medición

Los modelos y métodos propuestos para las métricas se automatizan buscando eficiencia, seguridad, reusabilidad, fiabilidad y exactitud tanto en datos como en información, porque en un alto porcentaje se debe recurrir a datos históricos que permitan aprender de las experiencias, siendo las herramientas de software especializadas en medición un recurso necesario que contribuye con esto.

Las herramientas automáticas de análisis estático de código fuente es un modelo compuesto por cuatro partes: algoritmos de modelo de construcción, análisis y reconocimiento de patrones, patrones de conocimiento y representación de resultados [15]. El modelo se construye a partir del código fuente o el archivo binario del programa bajo análisis y los algoritmos de análisis y reconocimiento de patrones tienen la función de buscar en el modelo, con el fin de encontrar probables problemas. Los patrones se almacenan en un bloque separado llamado Patrones/Conocimiento. Los patrones de diseño describen soluciones genéricas y las mejores prácticas recurrentes en problemas de software, incluyendo tanto aspectos estructurales como de comportamiento del programa. El enfoque del análisis estático tiene como intención revisar el código fuente, comprobando el cumplimiento de normas específicas y el uso de argumentos, entre otros, La promesa del análisis estático es identificar automáticamente los problemas de codificación antes de que el programa se lance. El análisis estático tiene como objetivo examinar estáticamente el texto de un programa, sin intentar ejecutarlo. El enfoque dinámico consiste esencialmente en ejecutar el código y comprobar dinámicamente la comprobación las inconsistencias de los resultados dados. Esto significa que las pruebas y el código de revisión son cosas separadas y distinguibles, pero no deben producirse separadamente [15].

Herramientas como FindBugs [15] se utilizan en el análisis estático para la detección de defectos de software, y se ha popularizado ampliamente. Existen diversos contextos para realizar análisis estático para la detección de defectos, tales como: 1) cuando se

realiza la revisión de código de un módulo recién escrito, y 2) cuando se desea buscar defectos en una base de código existente. FindBugs es una herramienta de análisis estático de código abierto que analiza los archivos Java Class buscando defectos de programación. El motor de análisis de los informes consta de cerca de 300 diferentes patrones de errores, y cada patrón se agrupa en una categoría y cada informe de un patrón de error se le asigna una prioridad: alta significa que debe corregirse rápidamente, media debe tomarse en cuenta para corregir o baja que no existe riesgo pero igual se debe corregir para un mejor rendimiento.

Otra herramienta para la medición, Scade, utiliza métricas HSS (Halstead's Software Science) y define nociones tales como volumen del programa, esfuerzo y dificultad, con base en el número total de operadores y operandos. La arquitectura del programa difiere significativamente de intérpretes clásicos [10]. Tiene cinco componentes de alto nivel: el analizador, el intérprete, interfaz para las métricas estáticas, interfaz para las métricas dinámicas y la implementación de las métricas. Estos cinco componentes están organizados en tres capas. El analizador y el intérprete son la capa más baja y tienen acceso directo a toda la información proporcionada por el modelo Scade. Las dos interfaces son la capa media, que tienen la función de proporcionar información en un formato fácil de utilizar por la tercera capa la aplicación de las métricas. En la Figura 4 se muestra a Scade con métricas HSS.

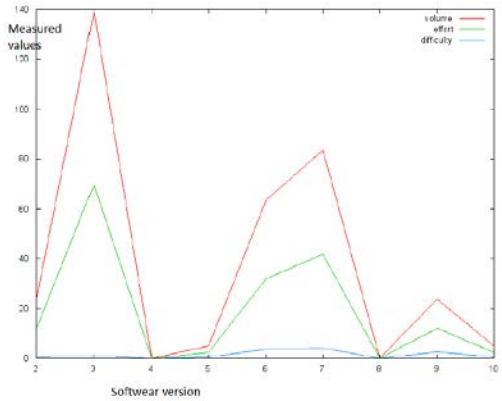


Figura 4: Scade con métricas HSS [10]

Lugo García [20] propone un proceso de colección automática de medidas del software, utilizando la herramienta Open Source StatSVN, que permite recolectar estadísticas de desarrollo a partir de un repositorio SVN. Este autor propone basar la medición del tamaño en líneas de código, con las que se puede medir la productividad de los trabajadores al controlar el tiempo. Esta herramienta es aplicable en la etapa de codificación y también en la fase de diseño. Sin embargo, requiere pasos manuales para obtener los indicadores de interés.

También existen herramientas independientes o integradas como plug-in con otras, tales como Cyvix, JDepend, Metric1.3.6, NDepend, SDMetrics, Refactorit, Design Advisor, GenMETRIC y SAAT [25], que se basan en código fuente y/o binario. Este proceso se realiza después de la programación, lo que es una desventaja para prever errores a tiempo, porque no están diseñadas para medir la calidad en etapas anteriores a la programación.

3 Conclusiones

Si bien existen modelos y métodos para aplicar medición, por ejemplo los puntos de función, los puntos de casos de uso, las líneas de Código (LOC), las métricas de COCOMO

II, entre otras, que se centran en el tamaño del software sin tener en cuenta el tipo de lenguaje utilizado para el desarrollo del producto [20], es importante aprovechar la teoría para generar herramientas de automatización.

La automatización de procesos conlleva ventajas tales como la eficiencia en el desarrollo de las diferentes tareas, la reducción de costos, la optimización de tiempos, la flexibilidad para realizar ajustes a los procesos, la racionalización de recursos, entre otras. En relación con esto, Rodríguez y Piattini [21] plantean que un denominador común de los estudios que presentan casos de aplicación práctica, es la importancia que atribuyen a disponer de un entorno tecnológico que automatice las actividades. La mayoría carecen de él y se plantean como trabajo futuro abordar su construcción para que les permita medir, procesar los datos y analizar los resultados de una manera más rápida y sencilla.

El tema de las métricas sigue siendo un reto en los procesos de Ingeniería del Software, porque cada vez crece más la demanda por productos informáticos con alta calidad, y que deberán cumplir las expectativas cambiantes de los clientes y usuarios, donde el software se ha convertido en producto transversal como soporte de las actividades sociales. Esto obliga a actualizar normatividades internacionales orientadas a lineamientos, pero sin perder de vista los requisitos del producto. En este sentido, las normas ISO/IEC 25000 [14] contemplan un modelo para la calidad del producto software que define las características y sub-características de calidad que se pueden medir [22]. El problema es no determina un conjunto concreto de indicadores, métricas y umbrales que puedan ser tomados por las organizaciones como referente a la hora de evaluar y certificar la calidad de sus productos.

Si bien los Puntos de Función y las Líneas de Código (LOC) han sido las técnicas más utilizadas para estimar tamaño, tiempo y esfuerzo de desarrollo de un proyecto [23], empiezan a surgir otras como los Puntos de Características, MK II, Puntos de Objeto y Puntos de Casos de Uso, ideadas para fortalecer la estimación en los actuales modelos de desarrollo de software. A nivel de métricas se ve la necesidad de crear herramientas dinámicas (automatizadas) que puedan integrar todo el ciclo de vida del software, cumpliendo características como altos niveles de usabilidad, pero sobre todo que sus resultados sean fiables, minimizando los índices de error durante el proceso de construcción, así como los defectos que en la mayoría de veces pueden ser catastróficos.

Como tema de estudio futuro se sugiere abordar el concepto de la Programación Orientada a Aspectos (POA) [2] como un nuevo enfoque de programación para desarrollar software, porque los enfoques existentes contienen limitaciones al momento de separar adecuadamente las preocupaciones transversales, como por ejemplo el almacenamiento en caché, la localización, la seguridad, la puesta en común de recursos, la sincronización y el control de excepciones, entre otros.

La sociedad actual demanda productos confiables, funcionales y fáciles de usar, que presenten disponibilidad y que sean flexibles a modificaciones, con altos beneficios, que sean seguros, intuitivos, fiables, ágiles, y muchos otros aspectos orientados a la calidad. El software como producto no es ajeno a estas características y requiere de mayor cuidado, porque los datos y la información constituyen el activo más importante de la sociedad actual. Es recomendable que las personal en los diferentes procesos de la Ingeniería del Software tengan la convicción de Peter Drucker: *Si no se mide lo que se hace, no se puede controlar, y si no se puede controlar, no se puede dirigir, y si no se puede dirigir no se puede mejorar*. Por eso es que la aplicación de la medición en la industria del software no es una opción, es una obligación que se verá reflejada en la sostenibilidad, la competitividad y el crecimiento de la industria.

Referencias

[1] Wang, H. et al. (2012). A comparative study on the stability of software metric selection techniques. Proceedings 11th International Conference on Machine Learning and Applications, pp. 303-307.

[2] Arora, K; Singhal, A. & Kumar, A (2012). A study of cohesion metrics for aspect- oriented systems. International journal of engineering science & advanced technology 2(2), pp. 332–337.

[3] ISO (2014). División de medición de calidad. ISO/IEC 2502n.

[4] Wordreference Diccionario (2005). Medir. Online [Feb. 2015].

[5] Thefreedictionary. (2013). Medida. Online [Feb. 2015].

[6] Bouwers, E.; Deursenz, A. & Visser, J. (2013). Evaluating usefulness of software metrics: an industrial experience report. Report TUD-SERG-2013-003. Delft University of Technology.

[7] Mendoza, H. & Bautista, G. (2002). Probabilidad y estadística. Universidad Nacional de Colombia.

[8] IEEE (1990). Standard glossary of software engineering terminology. IEEE Std 610.121990.

[9] Lionel C.; Morasca, S. & Basili, V. (1999). Defining and validating measures for object-based high-level design. IEEE transactions on software engineering 25(5), pp. 722–743.

[10] Attila, K. & Laszlo, B. (2013). Scade interpreter for measuring static and dynamic software metrics. Proceedings 11th International Symposium on Intelligent Systems and Informatics, pp. 26-28.

[11] Pressman, R. (2010). Ingeniería del Software - Un enfoque práctico. USA: McGraw.

[12] Cendejas, J. (2014). Modelo de desarrollo de software Integral colaborativo. Tesis Maestría. Universidad Popular Autónoma del Estado de Puebla

[13] Tarcísio G. et al. (2014). Statistical dataset on software metrics in object-oriented systems. Software Engineering Notes 39(5), pp. 1-6.

[14] ISO (2014). SQuaRE - System and Software Quality Requirements and Evaluation. ISO/IEC 25000.

[15] Zoubi, Q.; Alsmadi, I. & Huda, B. (2012) Study the impact of improving source code on software metrics. Proceedings International Conference on Computer, Information and Telecommunication Systems, pp. 1-5.

[16] Abuasad, A. & Alsmadi, I. (2012). The correlation between source code analysis change recommendations and software metrics. Proceedings 3rd International Conference on Information and Communication Systems, Article No. 2.

[17] Bermón, L. (2014). Simulación. UNAL. Online [Mar. 2015].

[18] RAE (2013). Modelar. Diccionario de la lengua española. Online [Mar. 2015].

[19] CSSE (2014). COnstructive COst MOdel II. Online [Feb. 2015].

[20] García, L. (2011). Colección automática de métricas hacia un repositorio de mediciones. Rev. Fac. Ing. Univ. Antioquia 58, pp. 199-207.

[21] Rodríguez, M. & Piattini, M. (2015). Experiencias en la industria del software - Certificación del Producto con ISO/IEC 25000. XVIII Congreso Iberoamericano en Ingeniería de Software, pp. 1-14.

[22] Rodriguez, M.; Pedreira, O. & Fernández, C. (2015). Certificación de la mantenibilidad del producto software: Un caso práctico. Revista Latinoamericana de Ingeniería de Software 3(3), pp. 127-134.

[23] Cordero, R. (2013). Estimación en proyectos de software integrando los métodos de Boehm y Humphrey. Revista Cubana de Ciencias Informáticas 7(3), pp. 23-36.

[24] Boehm, B. & Ricardo, V. (2005). Achievements and challenges in software resource estimation. IEEE Software 25(5), pp. 74-83,

[25] Rodriguez, C. (2012). Herramienta para aplicar métricas al diagrama de clases del diseño orientado a objetos. Revista Avanzada Científica 15(3), pp. 39-52.

[26] Baigorria, L. & Montejano, G. (2006). Métricas aplicadas a la programación orientada a aspectos Proceedings Workshop de Investigadores en Ciencias de la Computación, pp. 1-5.

Descripción de un proceso de autenticación de una firma manuscrita usando procesamiento digital de imágenes

José Espitia
Cesar Torres
Marjorie Jácome
William Lasso
Universidad Popular del Cesar – Colombia

1 Introducción

La firma humana ha sido desde siempre la forma más simple de autentificar la vida y las pertenencias, y es la más común y utilizada de aprobación en el mundo físico. En el mundo digital ha sido descartada para estos mismos fines por su inherente complejidad. A diferencia de los sistemas basados en características morfológicas del cuerpo humano, la firma es una señal de naturaleza distinta de tipo psicomotriz, lo que para el mundo digital la hace una señal variable e inestable con respecto a cada ejecución. Es decir, la firma digitalmente generada no contiene la misma información digital equivalente; pero si se combina con un adecuado método propio de extracción de características físicas, propiedades como la forma puede llegar a establecerse como patrones [1], candidatos hacer extraídos y definitivamente reconocibles, como en el caso que se presenta a continuación.

2 Metodología

El objetivo de esta investigación es extraer la firma por completo del espacio de la digitalización, eliminando todas las zonas que no hayan sido utilizadas por el firmante, con base en un algoritmo de barrido digital diseñado y desarrollado para detectar la ubicación de los bordes externos de la firma. Para lograrlo, primero se pasa la imagen a escala de grises, para convertir la hipermatriz de color a una matriz bidimensional con valores que van de 0 a 255. Hay que tener en cuenta que este proceso no afecta el objetivo central del proceso, que es la detección de la forma de la firma.

El proceso o algoritmo de barrido resultó ser un mecanismo multipropósito, que ha permitido tanto eliminar la zona de trabajo no utilizada por el firmante, como extraer los bordes externos de la firma. Además, no permite ninguna distinción en cuanto al tamaño de la firma o su localización dentro del espacio de trazo, una característica importante y relevante en el proceso de autenticación y reconocimiento de patrones. Dentro de las condiciones normales de funcionamiento, este algoritmo siempre detectará y extraerá la característica. En estas condiciones se utilizó otro proceso, que permite eliminar pequeñas suciedades que se encuentran dentro de la zona de interés y que son ocasionadas por el mismo roce inadecuado de los nudillos de los dedos, al ejercer presión inconsciente sobre el mecanismo digitalizador. Estas difuminaciones se clasifican como ruido inherente y pueden ser un distractor para los algoritmos mencionados.

Con la aplicación de estos algoritmos se puede garantizar que el patrón resultante brinde una representación real y limpia de cada ejemplar de la firma que ejecute el usuario. El resultado de la aplicación del estos procesos da como resultado una imagen bidimensional de dimensiones desconocidas (filas, columnas), lo que obliga a la estandarización de las matrices a una dimensiones fijas de 100 filas por 300 columnas

para las pruebas piloto. En la realidad estas dimensiones irían directamente relacionadas con la resolución que maneje el agente o dispositivo digitalizador.

Para este propósito se puede utilizar cualquier algoritmo de interpolación numérica especializado en matrices. Para aumentar, si las dimensiones de la matriz están por debajo de 100x300, o para reducir, si están por encima de estas dimensiones, dando como resultado una matriz estándar de dimensiones fijas de 100 filas por 300 columnas. Antes de pasar a la etapa de extracción de bordes se debe aplicar nuevamente el algoritmo de eliminación de ruido, con el objetivo de retirar residuos de imagen o ruidos circundantes a los bordes de la firma, producidos por la interpolación numérica y que pueden entorpecer el proceso de extracción.

Posteriormente, se aplica el algoritmo multipropósito para recorrer las columnas de izquierda a derecha consecutivamente, e iniciando desde la fila uno, preguntado pixel a pixel de la columna si el valor contenido es igual a 0. Si esto se cumple quiere decir que se está tocando un punto perteneciente al borde superior de la firma, y en este caso se toma la distancia recorrida, se interrumpe el recorrido y se pasa a la siguiente columna con el mismo propósito, hasta llegar a la columna 300 de la imagen. Como resultado se obtiene un vector que contendrá punto apunto el margen superior de la imagen. Este mismo proceso se realiza para obtener el margen inferior de la firma, que se logra girando la imagen 180 grados. Como resultado, por ejemplar se obtiene dos vectores para trabajar y procesar a través del proceso central de la investigación, denominado proceso de autenticación de una firma manuscrita, y que permitirá realizar la evaluación.

2.1 Proceso de detección de la firma a través de la eliminación de ruido

Asumiendo que lo que realmente importa es la forma de la firma, los valores internos del pixel en este caso particular servirán solamente para detectar los bordes externos que la describen, por lo tanto se puede realizar una simplificación de la misma convirtiéndola a escala de grises, porque lo que define una imagen en formato RGB es la composición del color [2]. Esta característica es despreciable en la extracción del trazo del firmante, por lo que la imagen de entrada será una matriz bidimensional “A (i, j)”.

La intención principal de este procedimiento es retirar del espacio de trabajo las zonas que no utiliza el firmante para el trazo. Por otro lado, y debido a la naturaleza del proceso, agrega una ventaja adicional es que funciona indirectamente como mecanismo de rastreo, detectando y encontrando la firma donde quiera que se encuentre dentro del área de trabajo. Esto resuelve de manera indirecta el inconveniente provocado por las diferentes dimensiones en que la firma se puede presentar entre cada ejemplar realizado, como se observa en la Figura 1.

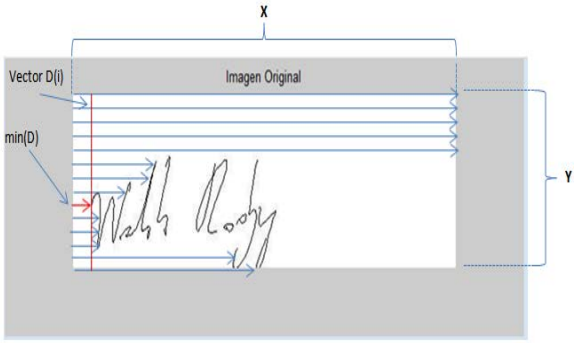


Figura 1: Explicación grafica proceso de detección a través de la eliminación del ruido

2.2 Descripción del proceso algorítmico para la detección de la firma a través del proceso de eliminación de ruido

```
A=imread('FIRMAS\firma1.jpg')
A=rgb2gray(A)
D=0;
CONT=0;
A_AUXILIAR=0
[Y,X]=size(A);
for i=1:Y
for j=1:X
if A(i,j)<=128
CONT=CONT+1;
D(CONT)=j;
break;
end
end
end
D_MENOR=min(D);
for i=1:Y
COLUMNA=0;
for j=D_MENOR:X
COLUMNA=COLUMNA+1;
A_AUXILIAR(i,COLUMNA)=A(i,j);
end
end
```

En A se carga la imagen original; luego se pasa a escala de grises vector de distancias recorrida y contador para llenar el vector D; de A se obtiene la altura y la base para recorrer por filas la imagen, lo que permite desplazarse por las columnas; si el valor del pixel es menor o igual 128 guarda su posición en el vector D, y termina el recorrido por las columnas pasa a la siguiente fila; el proceso *toco la firma* avanza así por cada una de las filas que componen la imagen. Como resultado se tiene un vector D con el mismo tamaño de la altura de la imagen y las distancias en pixeles. En A_AUXILIAR se obtiene la imagen sin el ruido existente en la parte izquierda de la original A. Para el retiro del ruido en las demás direcciones se gira la imagen 90 grados para la zona superior, 180 para la zona izquierda, 270 para la zona inferior, y se aplica este mismo proceso consecutivamente. El proceso se puede observar en la secuencia de la Figura 2.

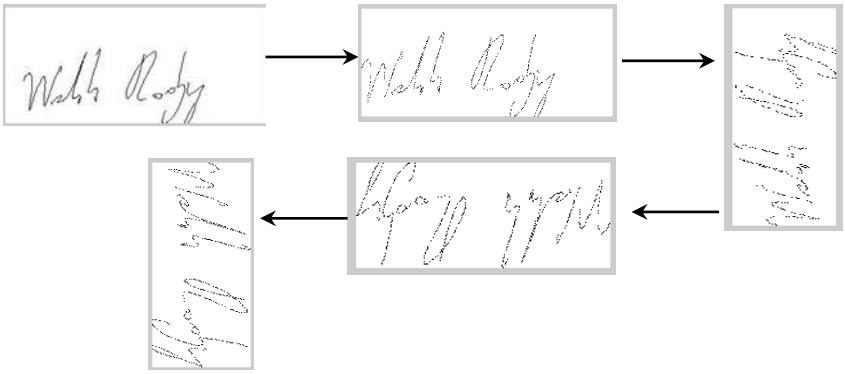


Figura 2: Proceso desde la imagen digital firma a escala de grises hasta la eliminacion de ruido

Una vez extraída la firma se realiza un redimensionamiento de la imagen resultante, es decir, se establecen dimensiones específicas en cuanto al tamaño se refiere. Para ello se

utiliza un algoritmo de interpolación numérica [3], en este caso se utiliza una función de MatLab especializada en este tipo de tareas, llamada Imresize (A,[100 300],'bicubic'). Cabe resaltar que para realizar la interpolación es apropiado retirar las zonas no implicadas en el contorno de la firma. Esto evita que el algoritmo de interpolación genere excesivas brotes de ruido con valores dispersos, que luego serían difíciles de remover o ajustar. Como resultado de este proceso se tiene una matriz de 100x300 pixeles, como se observa en la Figura 3.

Figura 3: Imagen resultante del algoritmo de interpolación

Debido a la naturaleza del algoritmo se generaron zonas opacas en los alrededores de los bordes que describen el contorno de la firma. Estos valores son producidos por la combinación que realiza el algoritmo de interpolación entre pixeles vecinos, para generar los nuevos valores que conformarán la imagen resultante. A través de un análisis visual, como el de la Figura 3, se aprecia que las zonas encerradas por los círculos son las equivalentes al ruido producido. La ventaja es que los niveles de luz de estas zonas son, en comparación a los pertenecientes al contorno de la firma, elevados, hasta el punto que son de fácil identificación numérica, como lo que realiza la siguiente rutina de recuperación, y cuya imagen resultante se aprecia en la Figura 4. De esta manera se resuelve el problema y entonces se obtiene la zona de interés estandarizada y en perfectas condiciones.

```
[a,b]=size(imagen1);
for i=1:a
 for j=1:b
 if imagen1(i,j)<=128
 imagen1(i,j)=0;
 else
 imagen1(i,j)=255;
 end
 end
 end
```

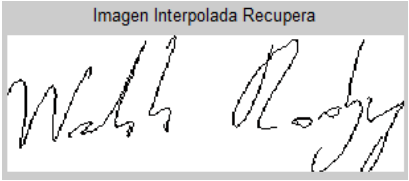


Figura 4: Imagen resultado de aplicar algoritmo de recuperación

2.3 Proceso de extracción de la forma física de la firma

Para este proceso se diseñó un algoritmo inteligente capaz de detectar los bordes inferiores y superiores de la firma [4], que se encuentra inmerso en la zona de interés e independientemente va recreando de forma precisa dos vectores que contienen las distancias en las que esta dibujado cada punto o píxel que conforma la firma de la imagen. La metodología operacional de este algoritmo se ilustra en la siguiente Figura 5.

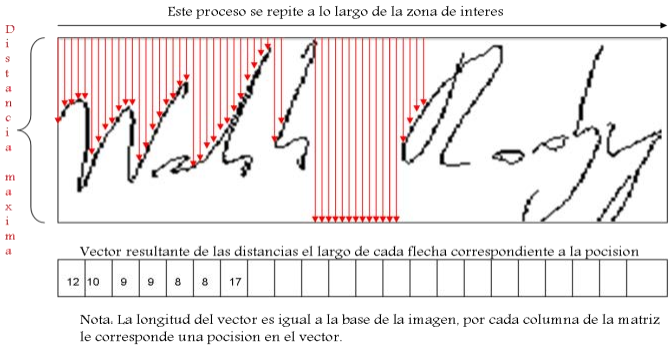


Figura 5: Metodología del algoritmo de extracción de bordes

El resultado de este algoritmo son dos vectores: vector "A", que posee el borde superior de la figura de la firma y que si grafica corresponderá de forma exacta al borde superior de la firma, como se muestra en la siguiente figura 5a; y el vector "B", que al graficarlo se obtiene la forma del borde inferior de la firma, como se ilustra en la Figura 5b.

```
b=rot90(A);
d=0;
cont=0;
[y,x]=size(b);
for i=1:y
 for j=1:x
 if b(i,j)<=128
 cont=cont+1;
 d(cont)=j;
 break;
 end
 end
end
contador=length(d)
for i=1:length(d)
 daux(i)=d(contador);
 contador=contador-1;
end
A1=daux;
```

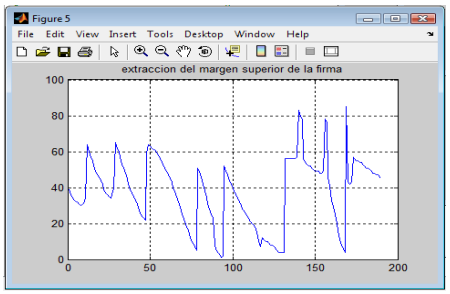


Figura 5a: Vector "A" graficado, borde superior de la firma

```
b=rot90(A,3);
figure(5),imshow(A);
d=0;
cont=0;
[y,x]=size(b);
for i=1:y
```

```
for j=1:x
if b(i,j)<=128
cont=cont+1;
d(cont)=j;
break;
end
end
end
A2=d;
```

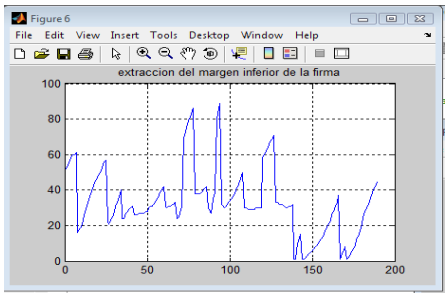


Figura 5b: Vector "B" graficado, borde inferior de la firma

2.4 Proceso de autenticación de la firma

El proceso de autenticación de la firma, o control de calidad como se conoce dentro de esta investigación, es un proceso estadístico diseñado en el proyecto con el fin de garantizar y controlar que las personas, que utilicen este sistema de seguridad, posean un patrón de trazo estable entre cada repetición que realicen de su firma. Todas las personas pueden decir que tienen una firma, pero cada vez que la utilizan lo hacen de forma diferente al trazo original o no posee un patrón estándar para cada repetición. En este caso, un mecanismo de seguridad debe evaluar y controlar que las personas que vayan a utilizar un sistema de seguridad especial, como el de esta naturaleza, sean idóneas y que cumplan a cabalidad con todos los requerimientos físicos, motrices y de destreza que el sistema exige para un óptimo funcionamiento.

El proceso de control de calidad tiene como entrada seis vectores (A1, A2, B1, B2, C1, C2), que son el resultado de los procedimientos de tratamiento digital de la señal. El vector A1 posee el margen superior de la firma original, A2 el margen inferior de la misma, B1 el margen superior del ejemplar uno, B2 el margen inferior del mismo, C1 el margen superior del ejemplar dos, C2 el margen inferior del ejemplar dos, y como salida un mensaje de texto en un cuadro de dialogo que emite un diagnóstico de rechazo o aceptación, acompañado del margen de error arrojado por el proceso. Los pasos que componen este procedimiento secuencialmente son:

Entradas

Vectores, A1, A2, B1, B2, C1, C2;

Se calculan las desviaciones típicas de cada uno de los vectores [5], para obtener por cada vector un coeficiente de variación sobre el comportamiento de sus valores.

DtA1 = Desviación típica de (A1);
DtA2 = Desviación típica de (A2);
DtB1 = Desviación típica de (B1);
DtB2 = Desviación típica de (B2);
DtC1 = Desviación típica de (C1);
DtC2 = Desviación típica de (C2);

Luego se comparan las desviaciones típicas de los límites superiores [6], calculando las tres posibles diferencias (R1, R2, R3):

R1 = valor absoluto de (DtA1-DtB1);
R2 = valor absoluto de (DtA1-DtC1);
R3 = valor absoluto de (DtB1-DtC1);

Ahora se calculan las medias generales o geométricas de las desviaciones superiores y las diferencias entre desviaciones típicas (R1, R2, R3) [7]:

Media general Dts = (DtA1 + DtB1 + DtC1) / 3;
Media general Rs = (R1 + R2 + R3) / 3;

Luego se calcula el error porcentual de los márgenes superiores (Error superior):

Error superior = (Media general Rs * 100) / Media general Dts;

Ahora se encuentra el valor absoluto de las diferencia para evitar posibles valores negativos, producidos por la resta entre las desviaciones típicas [8]. Posteriormente se comparan las desviaciones típicas de los límites inferiores, calculando las tres posibles diferencias (R4, R5, R6):

R4=valor absoluto de (DtA2-DtB2);
R5=valor absoluto de (DtA2-DtC2);
R6=valor absoluto de (DtB2-DtC2);

Ahora se hallan las medias generales o geométricas de las desviaciones inferiores y las diferencias entre desviaciones típicas (R4, R5, R6):

Media general Dti = (DtA2 + DtB2 + DtC2) / 3;
Media general Ri = (R4 + R5 + R6) / 3;

Luego se calcula el error porcentual de los márgenes inferiores (Error inferior):

Error inferior = (Media general Ri * 100) / Media general Dti;

Por ultimo, se calcula una tolerancia, que dependiendo del valor del error que se estime será el permitido para aceptar o rechazar un usuario durante el control de calidad de las firmas. Esta tolerancia se encuentra en las políticas de seguridad del sistema. Para este ejemplo se supondrá que el margen de error existente entre cada ejemplar de la firma debe de ser inferior al 5% para ser aceptado por el sistema.

Entonces,
Tolerancia = (Media general Dts * 5) / 100;
Tolerancia = (Tolerancia * 100) / media general Dts;
Error gs = (Media general Rs * 100) / Media general Dts;
Error gi= (Media general Ri * 100) / Media general Dti;
Error Genera= (Error gs + Error gi) / 2;

Si (Tolerancia >= Error General), firmas aceptadas, sino, firmas Insuficientes, con sus respectivos errores y análisis grafico opcional sobre las discrepancias entre ejemplares.

Si el proceso de verificación es positivo (aceptado), quiere decir que el usuario ha sido diagnosticado por el control de calidad como persona apta para utilizar este tipo de método de autenticidad, y por ende se le permitirá intentar crear su perfil de usuario dentro del sistema de seguridad de datos. Los resultados de este diagnostico se le darán a conocer al usuario por medio de un cuadro de dialogo que lo ilustrará a través de un mensaje texto sobre los umbrales alcanzado durante la prueba, por ejemplo: error producido por la discrepancia entre cada ejemplar, acompañado de un mensaje

comunicándole si fue aceptado o no ("Firmas Aceptadas con un error del:"). Tal como se ilustra en la secuencia de la Figura 6.

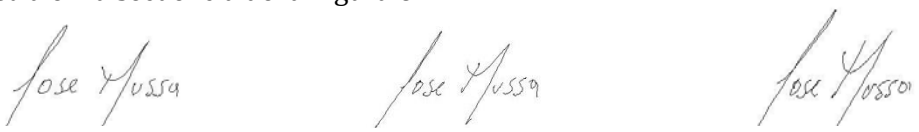


Figura 6: Firma original **Figura 6a:** Ejemplar numero uno **Figura 6b:** Ejemplar numero dos

Estos tres ejemplares fueron aceptados por el sistema por poseer una discrepancia de trazo del 3.2793%. Si el proceso de verificación es negativo (rechazado) [9], quiere decir que el usuario ha sido diagnosticado por el control de calidad como persona no apta para este tipo de sistema de autenticidad, y por ende no se le permitirá ni siquiera crear su perfil. Los resultados de este diagnostico se le darán a conocer al usuario por medio de un cuadro de dialogo que lo ilustrara a través de un mensaje texto sobre los umbrales alcanzado durante la prueba, por ejemplo: error producido por la discrepancia entre cada ejemplar, acompañado de un mensaje comunicándole si fue aceptado o no ("Firmas Insuficientes con un error del:"). El rechazo se ilustra en la secuencia de la Figura 7.

Figura 7. Firma original **Figura 7a.** Ejemplar numero uno **Figura 7b.** Ejemplar numero dos

Estos tres ejemplares fueron rechazados por el sistema por poseer una discrepancia de trazo del 20.9853%. Se han realizado diversos trabajos de evaluación o verificación biométrica de algoritmos de autenticación de firma manuscrita [10], pero los resultados obtenidos en esa investigación han mostrado que el algoritmo que combina DTW y GMM consigue mejores resultados en cuanto a tasas de error (EER): 4.14 % en la detección de firmas falsificadas entrenadas, siendo del 0.93 % para firmas falsas aleatorias. Los resultados que se obtuvieron son los esperados, porque se tienen tasas de error mayores al analizar las firmas falsas entrenadas, que en las firmas falsas aleatorias. Estos resultados tienen lógica porque las firmas falsas entrenadas han sido recogidas para usarlas en su base de datos con un ensayo previo, es decir, el usuario poseía el conocimiento de cómo es la firma original y pudo practicarla, por lo que la tasa de error debe ser mayor que si se compara con la tasa de error obtenida para las firmas falsas aleatorias, que fueron recogidas sin que el usuario tuviera conocimiento alguno.

3 Conclusiones

La tolerancia se convierte en el punto de referencia de esta investigación y depende del nivel que el usuario final quiera darle a su perfil de reconocimiento, porque el control de calidad se encarga de encontrar similitudes producidas por extracción de características morfológicas a través de la aplicación de métodos estadísticos. Es decir, se puede determinar la seguridad requerida en el sistema.

La detección del patrón de trazo permanente entre cada ejemplar del firmante es necesario y debe de establecerse como requisito indispensable al momento de su puesta en marcha. Esto con el objetivo de evitar posibles bloqueos o estancamientos producidos por el no-logro de la tolerancia previamente establecida.

Unos de los temas importantes inherentes en el proceso y que juega un papel importante, es la selección idónea del mecanismo o dispositivo de digitalización que se

vaya a emplear para la obtención de la firma. Es recomendable utilizar la misma referencia de dispositivo para la creación del perfil de usuario y para la validación del mismo. Además, se debe procurar que el dispositivo digitalizador posea propiedades que permitan la obtención de todas las características morfológicas y motrices del firmante.

Referencias

[1] Rodríguez, R. & Sossa, J. (2012). Procesamiento y análisis digital de imágenes. Madrid: Editorial Alfaomega.

[2] Báez, D. & Cervantes, O. (2012). Matlab con aplicaciones a la ingeniería, física y finanzas. Madrid: Editorial Alfaomega.

[3] Cuevas, E. & Zaldivar, D. (2010). Procesamiento digital de imágenes con Matlab y Simulink. Madrid: Editorial Alfaomega.

[4] García, A. (2012). Inteligencia artificial, fundamentos práctica y aplicaciones. Madrid: Editorial Alfaomega

[5] Martínez, C. (1999). Estadística y muestreo. México: Editorial Ecoe.

[6] Walpole, R. & Myers, R. (1999). Probabilidad y estadística para ingenieros. México: Editorial Prentice Hall.

[7] Maibaum, G. (1988). Teoría de probabilidades y estadística matemática. Buenos Aires: Editorial pueblo y educación.

[8] Triola, M. (2006). Estadística. México: Editorial Pearson.

[9] Ortega, M. (2014). Matlab aplicado a telecomunicaciones. Madrid: Editorial Alfaomega.

[10] Avila, M. (2010). Evaluación biométrica de algoritmos de autenticación de firma manuscrita. Proyecto fin de carrera. Universidad Carlos III Madrid.

Construcción de una herramienta usando software libre para la gestión de la red nacional de estaciones GNSS permanentes con propósitos geodinámicos de Colombia

Leonardo Cardona
Yuli Corchuelo
Héctor Mora
Servicio Geológico Colombiano – Colombia

1 Introducción

Un sistema de navegación está conformado por satélites que transmiten señales de radio, permitiendo a sus usuarios determinar con precisión su posición y velocidad las 24 horas del día y en cualquier parte del mundo, independientemente de las condiciones meteorológicas. Actualmente existen varios sistemas de navegación por satélite denominados Global Navigation Satellite System (GNSS), entre los se encuentran los sistemas en operación NAVSTAR-GPS y GLONASS, pertenecientes a EE.UU. y a la Federación Rusa, respectivamente. Existen otros sistemas en proceso de desarrollo e implementación, tales como los sistemas GALILEO de la Unión Europea, Beidou de la República Popular China, QZSS de Japón e IRNSS de India.

La navegación satelital ha revolucionado la forma en que se realizan las labores de topografía y cartografía. Un navegador GPS/GNSS permite obtener de forma rápida una ubicación con precisiones en un rango de 5 a 15 metros. Si se requiere mayor precisión, por ejemplo, a nivel sub-métrico, es necesario contar con una estación de referencia que le proporcione al usuario los datos de corrección diferencial, necesarios para obtener la precisión requerida y que puede darse en postproceso o en tiempo real. La mayoría de los usuarios de la navegación satelital deben contar con una estación GNSS de referencia para obtener las correcciones necesarias y las coordenadas en su procesamiento, esfuerzo innecesario si se tiene Estaciones GNSS permanentes de referencia de operación continua, comúnmente denominadas CORS.

El Servicio Geológico Colombiano (SGC), a través del proyecto Geodesia: Red de Estudios de Deformación (GeoRED) [1], viene desarrollando desde 2007 la implementación de una red nacional de estaciones GPS/GNSS para cubrir todo el territorio nacional, la cual considera realizar, entre otras, las siguientes acciones:

- Implementar una red nacional activa de estaciones permanentes geodésicas satelitales con propósitos geodinámicos.
- Utilizar los datos GNSS para el estudios geodinámicos, troposféricos, ionosféricos, entre otros.
- Obtener coordenadas de alta precisión y determinar las velocidades en las estaciones que constituyen la red.
- Contribuir a la definición de los nuevos Marcos Globales de Referencia (ITRF).
- Proporcionarle a la comunidad geodésica datos para la realización de levantamientos geodésicos, topográficos, cartográficos, catastrales y en general de posicionamiento.

Con la implementación de la red de estaciones permanentes se abre un mundo de posibilidades científicas, no solamente para el estudio de la dinámica terrestre, como en el caso de sismos, erupciones volcánicas, subsidencia entre otros fenómenos, sino también

para diferentes aplicaciones en diversos campos del conocimientos, tales como estudios atmosféricos para analizar la troposfera, mediciones de vapor de agua precipitable, retardo de la señal, y demás procesos meteorológicos y estudios ionosféricos para la interacción de la actividad solar y sus consecuencias mediante el análisis de GNSS.

Cada estación permanente GNSS se compone de receptor y antena GNSS, sistema fotovoltaico de alimentación, sistema de comunicación y sistema de protección a descargas eléctricas. Actualmente, el proyecto cuenta con 70 estaciones permanentes instaladas, que están continuamente recolectando datos de los satélites GNSS. Esto ha significado un reto, porque no se trata únicamente de instalar la estación permanente en un lugar con condiciones geológicas, geodésicas y de seguridad óptimas, sino que se debe garantizar la comunicación estación-centro de acopio para la descarga de los datos recolectados y así llevar a cabo el pre-procesamiento de la información. En ocasiones, la instalación de las estaciones permanentes GNSS de monitoreo continuo se realiza en lugares remotos y de difícil acceso, donde implementar un sistema de comunicación que permita establecer comunicación con las estaciones permanentes es un reto. Los sistemas instalados comúnmente en una estación GNSS permanente corresponden a uno o a la combinación de los siguientes sistemas:

- *Radio enlace*. Es una conexión entre diferentes equipos de telecomunicaciones usando ondas electromagnéticas. Los enlaces se hacen básicamente entre puntos visibles, es decir, puntos altos de la topografía.
- *Radiomodem celular*. Corresponde a un equipo que permite la comunicación con la estación permanente a través de la red de datos de un operador de telefonía móvil. La emisión y recepción de señales electromagnéticas de este equipo es equivalente a las de un Smartphone de reciente tecnología, en las bandas de frecuencia 900, 1800 y 2100 MHz. Este sistema tiene como limitación la cobertura celular por parte del operador.
- *Vía satélite*. La información adquirida por algunas de las estaciones permanentes, instaladas con estaciones sismológicas, es transmitida mediante un sistema de comunicación vía el satélite INTELSAT VI como medio, el cual retransmite al SGC mediante antena terrestre localizada en las instalaciones de Bogotá.
- *Virtual Private Network (VPN)*. En algunas ocasiones, mediante convenios con entidades públicas o privadas, se realiza la instalación de las estaciones permanentes GNSS en sus predios, lo cual brinda un sistema de comunicación con la estación a través de la red local de la entidad. Para establecer comunicación con la estación desde el SGC es necesario utilizar un cliente VPN suministrado por la entidad, esto permite conectarse con la estación y descargar los datos almacenados. Este sistema tiene la desventaja que requiere disponer e implementar servicios de descarga de datos en un equipo de cómputo adicional.

Mediante el uso del software TEQC (Translate/Edit/Quality Check) [2], desarrollado por UNAVCO, los datos nativos recolectados diariamente por las estaciones permanentes GNSS se convierten al formato universal RINEX (Receiver INdependent Exchange), que se utiliza como fuente para el procesamiento y análisis de la información. Asimismo, se facilitan a diferentes usuarios para que los utilicen en tareas de georreferenciación en proyectos de cartografía, infraestructura, SIG, entre otros.

Uno de los desafíos del proyecto GeoRED es mantener la red nacional activa de estaciones permanentes geodésicas satelitales con transmisión de datos en tiempo casi real, lo cual ha llevado a implementar diversas estrategias que permitan el almacenamiento, procesamiento, administración y difusión de los datos capturados por

las estaciones permanentes. El constante aumento y densificación de la red de estaciones permanentes GPS/GNSS conlleva un aumento en la gestión de la información y también implica retos para conservar el funcionamiento adecuado de la infraestructura, verificar el estado de la comunicación y la calidad de los datos recolectados.

Teniendo en cuenta que los datos GNSS que se producen por el desarrollo de las actividades en el proyecto GeoRED son un activo del SGC, y que adicionalmente dentro de la institución se están adelantando políticas de gestión de la información geocientífica, fue necesario crear un sistema que responda a las necesidades de la gestión de la información dentro del proyecto a la vez que éstas debe ser compatibles con las políticas internas de la entidad. Bajo esta premisa se hizo indispensable el desarrollo de un sistema de gestión de información de datos GNSS siguiendo la filosofía del proyecto, que consiste en utilizar herramientas de software libre y de alto desarrollo científico. Esto implica que la plataforma debe integrar las diferentes herramientas de software disponibles y garantizar que la generación de datos GNSS se enmarque dentro de las siguientes políticas básicas de información:

- Custodia de la información geocientífica
- Estandarización de datos y procesos
- Calidad de datos
- Integración e intercambio de información

Existen diferentes organizaciones a nivel mundial que realizan investigación asociada a las tecnologías GNSS, entre las que se destaca UNAVCO [3], un consorcio sin ánimo de lucro establecido en Colorado, Estados Unidos, y que opera redes GPS a nivel mundial. Esta organización realiza la descarga diaria de los datos generados por las estaciones permanentes distribuidas en todo el mundo y pone a disposición de la comunidad científica los archivos convertidos en formato RINEX. También brinda una plataforma web [4], en la cual se puede ver el estado de conexión de las estaciones permanentes, consultar la información asociada y ver de forma gráfica el respectivo análisis de calidad de los datos generados. También administra cientos de estaciones permanentes, para lo que han tenido que automatizar sus procesos, algo que, como parte del proceso de crecimiento del proyecto GeoRED, también generó la necesidad de incursionar en estos frentes de investigación.

2 Metodología

Para llevar a cabo la creación del sistema de gestión de información de datos GNSS, se empleó la metodología de desarrollo de software definida por Pressman [5], y que abarca las fases de análisis de los requisitos, de diseño, de codificación, de pruebas y de mantenimiento. El enfoque de desarrollo de software utilizado fue el incremental, con el que se desarrolló la plataforma mediante iteraciones. Se comenzó con una implementación simple del sistema y en cada iteración se realizaban los cambios y se agregaban funcionalidades, evaluando que el desempeño en conjunto fuera correcto y que se lograra una versión funcional del sistema, de tal forma que para la siguiente iteración existía una retroalimentación de la fase anterior.

2.1 Fase de análisis

Se realizó la identificación de las actividades que se ejecutan al interior del proyecto GeoRED, entre las que se encuentran:

- Verificación del estado de la comunicación y el funcionamiento de las estaciones permanentes GNSS.
- Descarga de información de las estaciones permanentes GNSS de monitoreo continuo.
- Ejecución del pre-procesamiento de los datos nativos recolectados por las estaciones permanentes para convertirlos a datos en formato estándar RINEX.
- Verificación de la calidad de los datos recolectados y generación de su respectivo reporte de calidad.
- Distribución de los datos generados diariamente a usuarios, organizaciones nacionales e internacionales.
- Implementación de un estándar de almacenamiento de la información GNSS recolectada por el proyecto GeoRED.

Se realizó un inventario de la información GNSS existente en el proyecto GeoRED y se encontró la recolectada desde 1994 (año en que fue instalada la primera estación permanente en el territorio colombiano) hasta la fecha actual y en diferentes estructuras de almacenamiento, lo cual implicó evaluar una estrategia para el almacenamiento de la información. Por otro lado, en el proceso de implementación de la red de estaciones permanentes se han utilizado diferentes instrumentos geodésicos GNSS, aspecto que se debe tener en cuenta para tener una plataforma automatizada que permita la adecuada gestión de los datos, estandarizando la forma en que son capturados, así como las tasa de rastreo, entre otros parámetros. Adicionalmente es necesario tener presente que la plataforma a desarrollar no abarca ninguna funcionalidad relacionada con el procesamiento de datos, porque para esta actividad existen diferentes tipos de software científico, tales como GIPSY-OASIS II [6] y GAMIT-GLOBK [7], que son utilizados en el proyecto GeoRED.

2.2 Fase de diseño

El volumen de información generado por las estaciones permanentes GPS/GNSS de operación continúa, llevó a considerar una estrategia óptima para realizar el almacenamiento de todos los datos. En el diseño del sistema de almacenamiento se tuvo en cuenta los siguientes requerimientos:

- Es indispensable que los receptores GNSS de las estaciones permanentes tengan una configuración estándar, tanto en la forma de rastreo como en el almacenamiento de los datos nativos, con el fin de estandarizar la forma como se capturan y descargan los datos del receptor GNSS. Además, al momento de generar los archivos RINEX se debe realizar de forma estándar. Para el caso del proyecto GeoRED se generan archivos RINEX a tasas de rastreo de 15 segundos y se tiene todas las observaciones de las constelaciones GNSS habilitadas.

- Es necesario conservar los datos nativos generados por las estaciones GNSS permanentes, porque permite su conversión a datos RINEX con diferentes tasas de rastreo, y a su vez usarlos en múltiples aplicaciones donde se requieran archivos RINEX a una tasa de rastreo diferente a 15 segundos.

- Los archivos nativos no se pueden comprimir debido a que frecuentemente se están accediendo y utilizando en múltiples tareas. Los archivos RINEX se comprimen usando HATANAKA [8].

Teniendo en cuenta estos requisitos, se diseñó la estructura de almacenamiento de datos GPS/GNSS que se observa en la Figura 1. Este diseño está basado en la propuesta de UNAVCO [9].

Estructura del servidor Centro de datos GNSS del proyecto GeoRED	
Metadato de las estaciones permanentes	
/logs	Logs de cambios de las estaciones permanentes según el estilo IGS
Datos rinex	
/rnx	En este directorio se archivos rinex diarios con un intervalo de muestreo mayo a un segundo
/rnx/obs/yyyy/ddd	Archivos rinex de observación comprimidos usando UNIX-compress Hatanaka.
/rnx/nav/yyyy/ddd	Archivos de navegación (si está disponible).
/rnx/met/yyyy/ddd	Archivos de meteorología (si está disponible).
/rnx/qc/yyyy/ddd	Archivos de calidad de datos correspondiente al reporte qc del teqc (si está disponible).
Datos nativos	
/raw	En este directorio se almacenan los archivos nativos diarios con un intervalo de muestreo al segundo
/raw/id/yyyymm	Archivos nativos (T00, T01, T02)
Nota: En Donde **yyyy** corresponde al año **ddd** corresponde al día juliano **mm** corresponde al mes **id** corresponde al identificador de la estaciones GNSS permanente	

Figura 1: Estructura del servidor de datos

También se realizó una evaluación de los parámetros necesarios para estandarizar la captura de datos en todas las estaciones permanentes GNSS y se definieron estos parámetros:

- El intervalo de medición de las señales emitidas por los satélites de los diferentes sistemas GNSS se realizará a la taza de un segundo y 15 segundos dependiendo de la calidad de la comunicación con el sitio.
- Se configurará el receptor GNSS para que genere un archivo con las observaciones capturadas cada hora, con el propósito de transferir archivos pequeños cada hora y no dejar la transferencia de un archivo con más datos al final del día. Adicionalmente, si llega a ocurrir un problema en la captura de los datos solamente se volverá corrupto el archivo en la hora específica en que ocurra.
- La información capturada por los receptores GNSS se almacena en el equipo, por lo que se realizó la estandarización en la estructura de los directorios generados y del nombre del archivo.

En la Figura 2 se muestra la configuración del registro de datos estándar utilizado en los receptores GNSS.

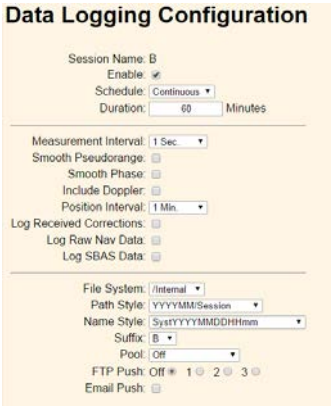


Figura 2: Configuración de los receptores GNSS

2.3 Fase de codificación

Teniendo en cuenta que la adquisición de software especializado no es posible por factores de costo y restricciones de interoperabilidad, se optó por software libre bajo Linux. Se dispuso un servidor con sistema operativo CentOS 6.6 en arquitectura de 64 bits, que se denominará Servidor Central de Datos-SCD, se procedió a implementar la

estructura de almacenamiento de los datos GNSS basados en la estructura diseñada y se desarrollaron los siguientes servicios y se registraron como tareas programadas:

Servicios construidos

- Servicio de descarga de datos cada hora. Es un *script* escrito en lenguaje Bash y registrado como tarea programada, que se encarga de realizar la descarga cada hora de los datos generados por las estaciones permanentes GNSS y de almacenarlos en la ruta definida en el SCD. La descarga de los datos se realiza a través de una petición al receptor GNSS de la estación permanente usando el protocolo FTP bajo el comando *wget*, y se deja en segundo plano usando el comando *nohup*. El tiempo de descarga de un archivo puede variar dependiendo del estado de la comunicación con la estación permanente.
- Servicio de sincronización del centro de datos GNSS del proyecto GeoRED con las estaciones permanentes. Es un *script* escrito en lenguaje Bash y registrado como tarea programada, que se ejecuta una vez a la semana. Se encarga de realizar la sincronización del directorio local del centro de datos GNSS, en el que se encuentran los archivos descargados para una estación permanente con el directorio remoto, donde se almacenan los datos capturados por la estación permanente. Para realizar la sincronización de los directorios se utilizó la aplicación LFTP de Linux, que permite realizar la transferencia de datos mediante diferentes protocolos: FTP, HTTP, FISH, SFTP, FTPS y HTTPS.
- Servicio de conversión de datos nativos a RINEX. Para realizar la conversión de los datos al formato estándar RINEX, se realizó un *script* en lenguaje Bash que se ejecuta a las 4:00 a.m. UTC. Se encarga de revisar la integridad y completitud de los datos y de realizar la conversión a formato RINEX, generando adicionalmente el respectivo reporte de calidad de los datos mediante el programa TEQC, y finalmente almacenando los archivos comprimidos usando la librería HATANAKA en su respectivo lugar en el SCD. En el proceso de conversión se realiza una estandarización en el dato y se dejan todos los archivos generados a una tasa de rastreo de 15 segundos, registrando en el metadato la información de la entidad que realizó el procesamiento.
- Servicio de monitoreo de las estaciones permanentes. Es importante conocer el estado de comunicación con las estaciones permanentes, para que en un caso de pérdida de comunicación con una estación se realice el monitoreo y se establezca si es por horas o días, o si es permanente y si es necesario programar una comisión de mantenimiento que revise el sistema de comunicación de la estación. Se debe tener presente que las estaciones permanentes tienen autonomía para el registro de datos de un año a una tasa de muestreo de un segundo, dándole al proyecto GeoRED el tiempo suficiente para solicitar los permisos y realizar los trámites necesarios para programar la visita al sitio donde se encuentra instalada la estación.

Para realizar el monitoreo del estado de la comunicación con las estaciones se escribieron dos *scripts* en Python, que se encargan de revisar si las estaciones permanentes responden al *ping* y si responde peticiones utilizando el protocolo HTTP. Es necesario revisar estos dos protocolos de forma complementaria para chequear la conectividad y la operatividad de los equipos. El *ping* permite establecer el funcionamiento del sistema de comunicación con la estación y la petición HTTP permite establecer la operación del receptor, de tal forma que se puede conocer si existen fallos en el equipo. Las librerías de Python utilizadas en la construcción de estos *scripts* son *requests*

y *psycopg2*, que permiten respectivamente el manejo de peticiones HTTP y el registro de información del estado de comunicación en la base de datos PostgreSQL.

- Servicio de construcción de reporte de calidad de datos. Para realizar mayor control sobre la calidad de los datos recolectados se construyó un *script* en Python, que se encarga de tomar el archivo de calidad de datos generado en el proceso de conversión a RINEX y extraer los valores de multitrayectoria de la señal, los valores de la relación señal/ruido, los saltos de ciclo y el tiempo de observación realizada. Los datos recolectados se almacenan en la base de datos y son utilizados para generar gráficas en forma de serie de tiempo, que permiten evaluar en el tiempo el funcionamiento de la estación. Para este *script* se utilizó la librería *psycopg2* que permite registrar la información extraída del resumen de calidad en la base de datos.

Se creó una base de datos para llevar un registro adecuado de la información generada en el proyecto GeoRED, que se desarrolló bajo el sistema manejador de base de datos PostgreSQL junto con su componente para el manejo de información espacial PostGIS. Entre la información a almacenar en la base de datos se encuentra:

- Características de las estaciones permanentes: nombre, ubicación, configuración receptor y antena, fecha de instalación, etc.
- Reporte de calidad de datos para cada archivo RINEX generado por la estación permanente.
- Estado de comunicación de la estación.
- Bitácora para el registro de los cambios, mantenimientos y sucesos extraordinarios ocurridos en cada estación.

Plataforma web

Para articular cada uno de los desarrollos y componentes creados y facilitar la gestión y administración, se creó una aplicación web con las siguientes funcionalidades:

- Mapa dinámico del estado de comunicación de las estaciones permanentes, que se actualiza cada hora y que genera alertas en problemas de comunicación. Se representa en la Figura 3.

Figura 3: Mapa dinámico del estado de la comunicación con las estaciones permanentes

- Consulta de información general de las estaciones permanentes.
- Consulta de las variables de calidad de datos de cada estación, tales como efectos de multitrayectoria, número de observaciones esperadas vs número de observaciones obtenidas, saltos de ciclo, etc. El resultado se observa en la Figura 4.
- Disponibilidad de datos en un determinado período de tiempo, cono se observa en la Figura 5.

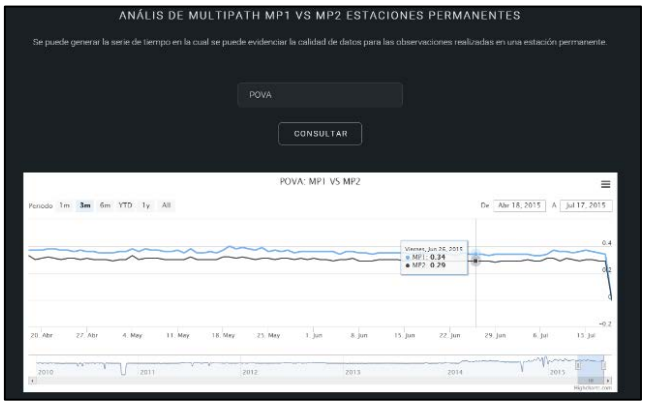


Figura 4: Análisis de multitrayectoria M1 y M2 para las estaciones permanentes

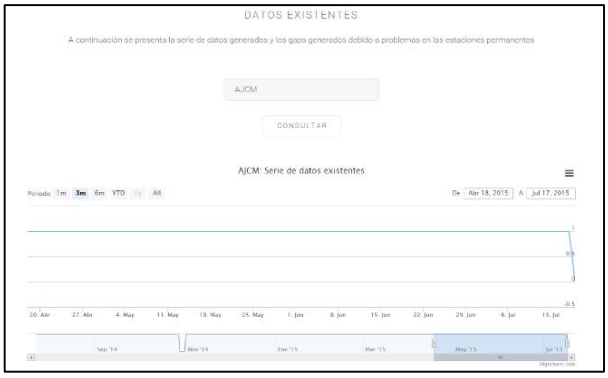


Figura 5: Gráfico en forma de serie que presentan los datos existentes y faltantes

Para el desarrollo de la plataforma web se utilizó el patrón de arquitectura de software modelo-vista-controlador (MVC) con componentes de software libre.

- Capa del modelo. Se implementó una base de datos PostgreSQL con su extensión PostGIS para el manejo de información espacial. El modelo de datos realizado es entidad-relación. Para establecer comunicación y ejecutar las operaciones sobre la base de datos se utiliza el API de Java Database Connectivity SQL.
- Capa del controlador. Se encuentran implementados el contenedor de *servlets* Apache Tomcat, en el que se despliega un archivo WAR que contiene el conjunto de servicios utilizados por la aplicación y que realiza las transacciones con las base de datos y da respuesta a las solicitudes utilizando el formato JSON. Adicionalmente, se encuentran configurados algunos servicios que permiten generar reportes en formato PDF utilizando la librería iText y JasperReports. También se encuentra implementado el servidor de mapas GeoServer, en el que se publican servicios de mapas utilizados por la aplicación, y que se publican como servicios bajo el estándar Web Map Services WMS [10], definido por Open Geospatial Consortium OGC [11] (Figura 6).
- Capa de la vista. Esta capa se desarrolló bajo HTML5 (Figura 7) y se utilizaron diferentes librerías *javascript* para dinamizar la plataforma. Para que la vista tuviera un diseño web adaptable se utilizaron las librerías Bootstraps y Awesome, que permiten que la plataforma puede ser visualizada adecuadamente desde diferentes dispositivos, como un teléfono inteligente. Para realizar la visualización de la información cartográfica se utilizó la librería Leaflet junto con el API de Google Maps para *javascript*. Para el manejo de las gráficas estadísticas interactivas se utilizó la versión libre de la librería Highcharts.

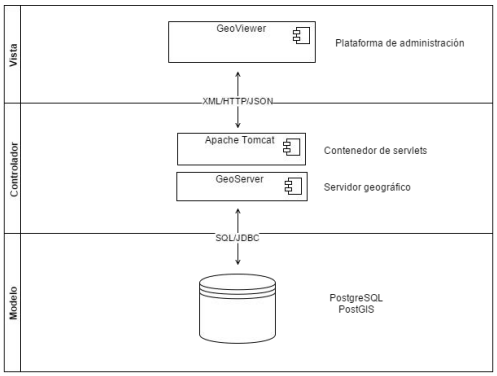


Figura 6: Diagrama de arquitectura de la aplicación

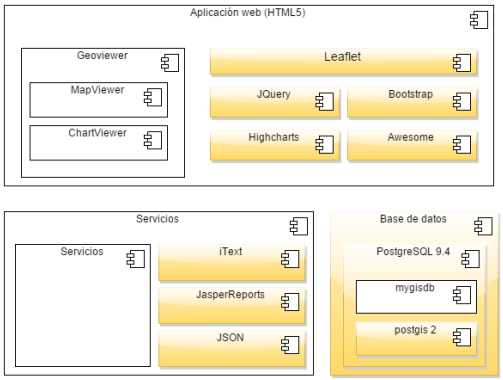


Figura 7: Componentes de software utilizados

2.4 Fase de pruebas

Se realizaron pruebas unitarias a cada una de las funcionalidades desarrolladas y se evidenció que cumplían de forma exitosa la actividad para la que fueron construidas. El servidor en el que se implementó la plataforma tuvo actualizaciones de diversas librerías, lo que provocó que algunos *ccripts* dejaran de funcionar debido al uso de Shell TCSH. Esto se debió porque al actualizarse perdió soporte en el manejo de número octales [12], por lo que fue necesario ejecutar nuevamente todos los *scripts* y ajustar los que fueron afectados por la actualización. Una vez finalizadas las pruebas unitarias se empezaron las pruebas de integración, para verificar cómo funcionaba en conjunto toda la plataforma, y se encontró que funcionaba como se esperaba.

2.5 Fase de mantenimiento

El mantenimiento corresponde a las actividades que evitan que la plataforma caiga en desuso por la obsolescencia de la información y por el funcionamiento incorrecto. Por esta razón, una de las principales actividades a realizar para tener un correcto funcionamiento es el ingreso de la información de las nuevas estaciones, así como la actualización de los cambios que tienen las estaciones permanentes, tales como cambios en los parámetros de comunicación y en el instrumental geodésico reportado, entre otros. Pero, para que se adapten a nuevos requerimientos de los usuarios, es necesario realizar la evaluación y optimización del código de los *scripts* y de la plataforma web. También es necesario realizar una evaluación periódica del comportamiento de la plataforma y en caso de tener algún error realizar los ajustes necesarios. Además, se debe prestar atención a los nuevos requisitos que puedan surgir o a los ajustes que sean necesarios debido a cambios tecnológicos en el instrumental geodésico

3 Resultados

Antes de implementar la plataforma de gestión de la red nacional de estaciones GNSS permanentes, el proyecto GeoRED realizaba muchas actividades de forma manual. La descarga de datos se realizaba dos veces por semana utilizando la herramienta FileZilla, un proceso que implicaba esperar a que se estableciera la comunicación con la estación y que generalmente tomaba mucho tiempo, para luego proceder a ingresar al directorio de los datos, seleccionar los datos faltantes e iniciar la descarga. Muchas veces, al intentar establecer la comunicación con la estación permanente, no era exitosa y se debía esperar hasta la próxima descarga de datos para intentar establecer nuevamente la comunicación. A partir de la implementación de los *scripts* para la descarga de datos, estos se obtienen desde las estaciones permanentes a medida que se generan. Además, cuando una estación tiene problemas de comunicación, los datos de descargan en los momentos en que se restablece la comunicación con la estación, que en muchas ocasiones tiende a ser a altas horas de la noche.

Realizar la conversión de los archivos nativos a formato RINEX era una actividad que tomaba cerca de 4 minutos por RINEX diario generado. Actualmente, esta conversión toma cerca de 27.359 segundos, cuando los datos recolectados por la estación permanente están a una tasa de rastreo de 1 segundo y 2.212 segundos cuando están a una tasa de rastreo de 15 segundos. El proceso de conversión a formato RINEX de todos los archivos nativos recolectados diariamente tarda aproximadamente 9 minutos con 17.844 segundos, para un total de 70 estaciones permanentes.

La revisión del estado de comunicación con las estaciones permanentes se realizaba dos veces por semana, para verificar el estado de comunicación se ingresaba desde un navegador web a la dirección IP de la estación y se identificaban las estaciones con las cuales no era posible establecer comunicación. Actualmente, a través de los *scripts* desarrollados para realizar el monitoreo del estado de comunicación con las estaciones permanentes, se puede saber de forma fácil y rápida el estado en que se encuentra cada estación. La ejecución de los *scripts* se realiza cada hora y en promedio tarda cerca de 30 minutos.

La plataforma web desarrollada tarda en promedio 12.99 segundos en ser desplegada por un navegador web y transfiere 2.356.47 kB de información (Figura 8). El despliegue no requiere ningún complemento adicional, porque la plataforma está basada en HTML 5 y JavaScript.

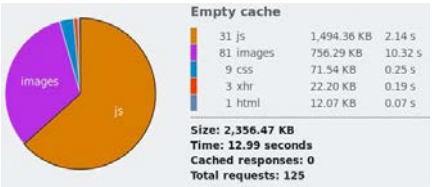


Figura 8: Estadísticas de despliegue de la aplicación

Aparte de la información transferida con el cargue inicial de la plataforma, se puede necesitar información adicional cuando el usuario empieza a interactuar con la aplicación. Para los análisis de calidad de datos cada petición que se realiza toma en promedio 494 ms, en los que se transfiere un archivo JSON de 77.68 kB, y para conocer la disponibilidad de datos de una estación se realiza una petición que toma en promedio 565 ms, en los que se transfiere un archivo JSON de 6.85 kB en promedio.

Para realizar la transferencia de datos en tiempo real se utiliza el protocolo NTRIP [13], que permite distribuir flujos de datos GNSS a través de Internet utilizando el

protocolo HTTP. Actualmente, el proyecto GeoRED se encuentra evaluando esta tecnología y su posible implementación.

4 Conclusiones

El uso de software libre permitió la creación del sistema de gestión de información de datos GNSS con bajos costos, integrando eficientemente las diversas herramientas del proyecto GeoRED.

La automatización de procesos permite realizar un control minucioso de las actividades realizadas, así como eliminar posibles fuentes de error humano en su desarrollo.

Con la automatización de procesos, el tiempo de dedicación de talento humano en la ejecución de actividades monótonas puede ser invertido en actividades de investigación y optimización de procesos.

Es útil contar con una plataforma de gestión que permita saber el estado de operación de la red GNSS en tiempo casi real, porque al tener conocimiento de los problemas con el sistema de comunicación, en la pérdida de información o en la calidad de los datos capturados por las estaciones, se puede tomar decisiones pertinentes para resolver los inconvenientes presentados.

Referencias

[1] Mora, H. (2010). Implementación red Nacional de Estaciones Geodésicas satelitales con propósitos geodinámicos. Documento BPIN para el Departamento Nacional de Planeación INGEOMINAS.

[2] UNAVCO (2009). TEQC - The Toolkit for GPS/GLONASS/Galileo/SBAS/Beidou/QZSS Data. Online [Nov 2014].

[3] UNAVCO (2009). Coconet en español. Online [Nov 2014].

[4] UNAVCO (2010). Plataforma de consulta web UNAVCO. Online [Dec 2014].

[5] Pressman, R. (2010). Ingeniería de Software - Un enfoque práctico. México: McGraw-Hill.

[6] NASA (2014). GIPSY-OASIS II, GIPSY-OASIS II. Online [Oct 2014].

[7] GAMIT-GLOBK (2013). Home page. Online [Nov 2014].

[8] Hatanaka, Y. (2008). A compression format and tools for GNSS observation data. Bulletin of the Geographical Survey Institute 55, pp. 21-30.

[9] UNAVCO (2014). UNAVCO Data - FTP Server Layout. Online [Nov 2014].

[10] Geoserver (2012). WMS reference. Online [Dec 2014].

[11] OGC (2014). About OGC. Online [Dec 2014].

[12] Tavkelishvili, G. (2013). Handling of octal numbers (since tcsh v6.14) causes script breakage. Online [Nov 2014].

[13] GDC (2004). Ntrip - Networked Transport of RTCM via Internet Protocol. Online [Dec 2014].

Desarrollo de una ontología sobre estudios de impacto ambiental orientada a sistemas de recuperación documental

Jaime Guzmán L.[1]
Jennifer Cartagena O.[2]
Edna Rodríguez G.[3]
[1]*Universidad Nacional de Colombia – Colombia*
[2, 3]*Institución Universitaria Colegio Mayor de Antioquía – Colombia*

1 Introducción

La contaminación ambiental es un tema al que se le ha prestado mayor atención en los últimos años, debido a que su aumento genera afectaciones a la salud y el medio ambiente que han obligado en cierta medida a crear instrumentos apropiados que incorporen la dimensión ambiental en la planificación del desarrollo territorial, con la intensión de reducir las afectaciones nocivas al entorno y a los seres vivos.

En la búsqueda de mejorar los procesos de licenciamiento, que es la autorización que otorga la autoridad ambiental colombiana para la ejecución de una obra o actividad que pueda ocasionar deterioro grave al medio ambiente [1], las autoridades ambientales adoptaron los criterios generales definidos en el Manual de Evaluación de Estudios Ambientales de Proyectos expedido por el Ministerio de Ambiente y Desarrollo Sostenible, y que permite evaluar los estudios ambientales que se desarrollen en el país [2], con el objetivo de hacer veeduría a los procesos que pueden afectar la salud humana y al medio ambiente. Por esto se establecen criterios para el desarrollo de proyectos, obras o actividades de gran envergadura que generen deterioro grave a los recursos naturales renovables, enumerados en los artículo 8 y 9 del decreto 2041 de 2014 [2].

Para obtener la licencia ambiental colombiana, se debe presentar ante la autoridad ambiental competente la evaluación ambiental previa a la ejecución o puesta en marcha del proyecto, y que está compuesta por un Diagnóstico Ambiental de Alternativas (DAA) y un Estudio de Impacto Ambiental (EsIA), ambos procesos deben ser documentados y sustentados mediante evaluación. A EsIA se le considera el instrumento básico para tomar decisiones sobre un proyecto, obra o actividad que afecte el medio ambiente natural o artificial [3]. Este estudio deberá tener en cuenta estudios tales como la identificación y evaluación de los impactos ambientales (EIA) que puedan ocasionar el proyecto, obra o actividad, indicando cuales pueden prevenirse, mitigarse, corregirse o compensarse; una descripción de las acciones propuestas y de su entorno; una valoración de importancia de cada impacto; una descripción de los programas de monitoreo y medidas propuestas de mitigación [4]. A esto se debe la importancia que tiene la realización del EsIA como proceso de gestión ambiental, y que ha sido probada a lo largo de años de aplicación, y también por lo que se han propuesto herramientas informáticas para automatizarlo porcompleto o partes esenciales del mismo.

En este orden de ideas, un esfuerzo para la realización de este tipo de trabajos es poder accesar fácilmente la documentación de estudios de caso asociados al concepto investigativo, académico y de contratación, que ayuden a su entendimiento en el contexto y teniendo en cuenta la legislación colombiana que los regula, para reconocer qué metodologías existen y se han aplicado con éxito en el territorio colombiano. Es decir, acceder a la documentación de los diferentes tipos de estudios realizados en el país en un marco de categorización del mismo con el objetivo de simplificar su búsqueda. Sin

embargo, la información que manejan los distintos sistemas es diversa y generalmente resulta complicado intercambiar de información entre los mismos [5].

Los profesionales en distintas áreas especializadas se refieren a conceptos o términos de maneras diferentes, por lo que es importante formalizar el conocimiento en el tema de EsIA, con la finalidad de lograr un lenguage común y unificado que sea comprensible, y para que se logre un verdaero intercambio de saberes e información. En este sentido se pueden utilizar las ontologías. Una ontología es un conjunto de términos estructurados jerárquicamente para describir un dominio [6], que pueden ser utilizados como el soporte fundamental para una base de conocimiento. En este caso, para el tema de estudios de impacto ambiental, no solamente permite que una comunidad en dicha área del saber y de manera consensada entiendan claramente la terminología, sino que esa terminología pueda ser entendida por las máquinas, es decir, agentes que haciendo uso de la ontología pueden realizar una tarea específica, tal como buscar documentos especializados.

Para encontrar documentación relacionada con los estudios de impacto ambiental en el contexto colombiano se utilizan diferentes buscadores y sistemas de recuperación documental. En este proceso se identifican dos principales problemas: 1) existen diferentes tipos de información asociada a contextos externos al de Colombia, y 2) se puede presentar el problema de la polisemia terminológica. Para solucionarlos, en este trabajo se presenta el desarrollo de una ontología sobre estudios de impacto ambiental en el contexto colombiano, que permite la representación de la información y el conocimiento de manera coherente y entendible para humanos y máquinas, sin ambigüedades terminológicas y bajo el contexto de las leyes y normativas colombianas. Esta ontología podría constituirse en el soporte estructural de un sistema de recuperación documental sobre estudios de impacto ambiental en el país, para accesar la información referencial y en texto completo acerca del tema.

Este trabajo está organizado de la siguiente manera: la sección 2 presenta el marco teórico que soporta el trabajo; en la 3 se presenta la metodología seguida para su desarrollo; en la 4 se detalla los resultados obtenidos en la investigación; en la 5 se revisan los diferentes trabajos relacionados con el desarrollo de ontologías en el campo de estudios de impacto ambiental y se comparan con los resultados presentados logrados; y en la sección 6 se presentan las conclusiones y se proponen trabajos futuros.

2 Marco Teórico

2.1 Estudio de Impacto Ambiental

EsIA muestra la descripción pormenorizada de las características de un proyecto, obra o actividad que se quiera llevar a cabo. Para ello se expiden los términos de referencia, se avoca conocimiento del caso, se fija el valor de los costos de evaluación y seguimiento y se hace entrega de los términos de referencia a los solicitantes [7]. A continuación se detallan este tipo de estudio a la luz del contexto colombiano. En el decreto 2041 de 2014 [2] se define este tipo de estudio como el instrumento básico para la toma de decisiones sobre los proyectos, obras o actividades que requieren licencia ambiental, y que se exigirá en todos los casos en que de acuerdo con la ley y el presente reglamento se requiera. La Autoridad Nacional de Licencias Ambientales (ANLA) [3] describe el contenido y características de lo que debe contener EsIA y nombra los principales requerimientos que debe incluir:

- Se establece el objeto y alcance del estudio
- Elaboración de un resumen ejecutivo

- Delimitación del área de influencia directa e indirecta del proyecto
- La descripción del proyecto, obra o actividad
- Información de compatibilidad del proyecto con el uso del suelo establecidos en el POT
- La información sobre los recursos naturales renovables que se pretenden usar, aprovechar o afectar para el desarrollo del proyecto, obra o actividad
- Identificación de las comunidades y de los mecanismos utilizados para informarles sobre el proyecto, obra o actividad
- La descripción, caracterización y análisis del medio biótico, abiótico, socioeconómico en el cual se pretende desarrollar el proyecto, obra o actividad
- La identificación y evaluación de los impactos ambientales que pueda ocasionar el proyecto, obra o actividad, indicando cuáles pueden prevenirse, mitigarse, corregirse o compensarse
- La propuesta de Plan de Manejo Ambiental del proyecto, obra o actividad

2.2 Acerca del concepto ontología

Una ontología constituye una especificación formal y explícita de una conceptualización compartida [8, 9]. De acuerdo con esto se aclara la definición haciendo énfasis en cada concepto: *conceptualización* se refiere a un modelo abstracto de algún fenómeno del que se identifican los conceptos que son relevantes; *explícito* hace referencia a la necesidad de especificar de forma consciente los distintos conceptos que conforman una ontología y las restricciones sobre ellos; *formal* indica que la especificación se debe presentar en un lenguaje de representación formalizado; y *compartido* refleja que la ontología debe capturar conocimiento consensual aceptado por una comunidad, como mínimo, por el grupo de personas que deben usarla.

Clásicamente, el término ontología ha sido empleado como referencia a la rama de la filosofía que trata el estudio del ser y de sus propiedades trascendentales [9], aunque las ontologías tienen aplicación en distintos dominios y dependiendo del contexto en el que se emplee proporciona diversas formas de compartir el conocimiento. La ontología es una tecnología para el desarrollo de la web semántica y es proporciona conocimiento y entendimiento en determinado domino o área del saber al producir interacción entre personas y aplicaciones. En la ontología, las definiciones asocian los nombres de las entidades en el universo del discurso bajo un contexto legible desde el punto de vista humano, describiendo el significado de los nombres y axiomas formales que restringen la interpretación y el uso de estos términos [8]. También facilita la especificación de los conceptos, sus propiedades y sus restricciones, así como la organización de los mismos con sus relaciones.

2.3 La web Semántica

Las ontologías tienen un papel fundamental en el desarrollo de la web semántica, debido a que permiten organizar y categorizar la información dispuesta en ella. La web semántica también resuelve dos importantes problemas: la sobrecarga de información y heterogeneidad en fuentes de información. En el contexto de la web, una ontología describe formalmente un dominio del discurso, facilitando un entendimiento compartido [10]. Por otras parte, la web semántica está constituida por una serie de capas y donde el desarrollo de las ontologías suponen el uso de dos de ellas: RDF y RDF Schema, así como se muestra en la Figura 1.

Sin embargo, para dotar la web semántica de una mayor expresividad en la representación del conocimiento asociado a las ontologías, es necesario una nueva capa

que asociada a Web Ontology Language (OWL), que permite la creación de ontologías con mayor capacidad de expresividad semántica y convirtiéndose en el lenguaje estándar para su implementación en la web.

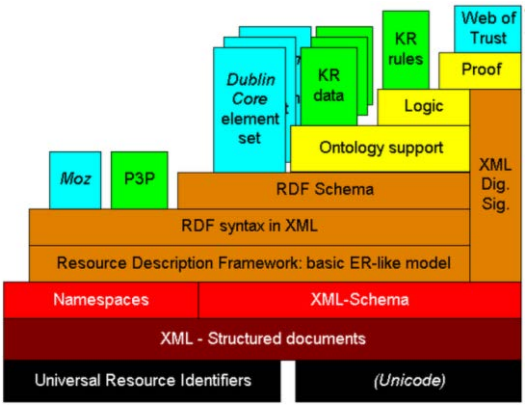


Figura 1: Capas de la Web Semántica [11]

3 Metodología

Esta investigación se planteó como objetivo general construir una ontología que representara conceptualmente la terminología sobre el Estudio de Impacto Ambiental. Para la estructuración de los conceptos y relaciones semánticas se aplicó un enfoque basado en *tesauros* y ontologías preexistentes, que permitieran hacer un modelado de este tipo de conocimiento. UNESCO [12] define *tesauro* como una lista controlada y estructurada de términos para el análisis temático y la búsqueda de documentos y publicaciones en los campos de la educación, la cultura, las ciencias naturales, sociales y humanas, la comunicación y la información.

Las relaciones semánticas básicas entre los conceptos modelados para este trabajo se definieron de tres tipos: equivalencia, asociación y jerárquicas [13]. La metodología propuesta para la creación de esta ontología siguió las siguientes fases:

1. Captura del conocimiento básico del dominio asociado al *corpus* bibliográfico en el contexto colombiano, con el fin de identificar el conjunto de términos relevantes, sus sinónimos y sus descripciones en el dominio. Básicamente se hizo la captura de los textos más representativos en el área EsIA como motor de conocimiento, que a su vez permitiera articularse con áreas afines. Para este proceso se tuvo en cuenta publicaciones entre 2010 y 2015 con el objetivo de encontrar información actualizada y acorde con la normatividad vigente.
2. Caracterización y complementación del conocimiento adquirido haciendo uso de la técnica de modelado UML [14]. En esta etapa se identificaron los componentes básicos de la ontología del dominio, tales como sus términos, relaciones, instancias y reglas de inferencias. Para la implementación de este tipo de conocimiento se utilizó la ontología Skos core, utilizada ampliamente en la creación de *tesauros*, y que provee una serie de propiedades que permite representar el conocimiento de forma flexible, describiendo el contenido y la estructura organizacional del concepto. La clase y las propiedades del Skos core utilizadas en el desarrollo de esta ontología bajo el enfoque de un *tesauro* se describes en la Tabla 1.
3. Implementación de los componentes básicos de la ontología utilizando el lenguaje OWL y ontologías existentes, con el fin de formalizar el conocimiento del dominio y produciendo definiciones precisas y no ambiguas

4. Validación de la ontología desarrollada mediante el diseño y elaboración de pruebas propias de este tipo de desarrollo, con el fin de valorar sus fortalezas y debilidades.

Tabla 1: Propiedades de Skos core

	Propiedades	Definición
Clase Concepto	Concept	Unidad de conocimiento
Etiquetas	prefLabel	Termino preferente
Léxicas	altLabel	Termino no preferente
Propiedades	definition	Definición de termino
Semánticas	narrower	Termino especifico
	broader	Termino general

Teniendo esta metodología es necesario establecer el procedimiento para el desarrollo de las fases y seguir un orden lógico que permita obtener de manera coherente la ejecución y formalización de la ontología. Cada desarrollador sigue sus propios principios, criterios de diseño y fases en el proceso de desarrollo, debido a la ausencia de un estándar para la construcción de ontologías, aunque existen algunas directrices desarrolladas para la creación de la ontología. Gómez y Benjamins [15] afirman que la mayoría de los métodos para la construcción de ontologías se centran en el desarrollo de las actividades, especialmente en la aplicación, la conceptualización y la ontología, lo que indica que no se presta suficiente atención a otros elementos importantes relacionados con la gestión, la fusión, el aprendizaje, la integración, el seguimiento y la evaluación de las mismas.

4 Resultados

En el desarrollo de la ontología se utilizaron varios componentes para la manipulación y posterior extracción de conocimiento. Uno de los componentes básicos fue la creación del *tesauro*, que permitió describir conceptualmente las relaciones existentes en los términos que refiere el estudio de impacto ambiental.

4.1 Criterios para la construcción de la ontología

Gruber [16] establece los siguientes criterios para la construcción de una ontología y que se retoman en este trabajo:

- Claridad: comunicar de manera efectiva el significado de sus términos
- Coherencia: hacer inferencias para que sean consistentes con las definiciones. La coherencia debe aplicarse a los conceptos sean informales o bien documentados
- Aplicabilidad: toda ontología debe ser planeada teniendo en cuenta la posibilidad de ser aplicada y extendida a otras áreas
- Compromiso ontológico y con la implementación: una ontología debe hacer pocas imposiciones sobre el dominio en el que se basa, tener flexibilidad y permitir la aplicación en otras áreas

Con base en estos criterios se articulan los conceptos para generar una ontología especializada en el área de Estudio de Impacto Ambiental que permita recuperar documentos especializados en el contexto colombiano.

4.2 Corpus bibliográfico

Para la captura del conocimiento básico del dominio asociado al corpus bibliográfico en el contexto colombiano, correspondiente a la primera fase de la metodología propuesta, se seleccionó un corpus de textos producidos por algunas instituciones importantes que trabajan en el tema, que regulan u orientan el desarrollo y que controlan

o monitorean EsIA. Estas son las instituciones que se eligieron para la extracción de la documentación:

- Autoridad Nacional de Licencias Ambientales (ANLA)
- Universidad Nacional de Colombia sede Medellín
- Institución Universitaria Colegio Mayor de Antioquia
- Universidad de Antioquia
- Corporación Autónoma Regional (CAR) [17]
- Área Metropolitana del Valle de Aburra (AMVA) [18]

De las bases documentales de estas instituciones se seleccionaron investigaciones, informes, tesis, artículos y boletines, para un total de 36 textos. El periodo comprende entre 2010 y 2015, y la búsqueda se limita a la producción en temas colombianos. En la Tabla 2 se relaciona la documentación que sirvió como marco de referencia para la primera etapa de la metodología.

Tabla 2: Clasificación de los textos seleccionados

Académicos	
Libros	26
Cursos	6
Planes de estudio	4
Contratación	
Documentos	6
Leyes	18
Investigación	
Tesis de investigación	6
Total	36

4.3 Árbol de dominio inicial

Una vez seleccionados los textos se realizó una segunda depuración de los documentos, se hizo una lectura analítica y se eligieron los más representativos entre las distintas instituciones. En total se seleccionaron 9 documentos para análisis detallado, y luego de ser analizados con el especialista en el tema se identificaron los primeros 5 conceptos de clasificación: normativas, características de un EsIA, tipos de diagnóstico ambiental, diferentes temáticas de una evaluación ambiental y diferentes estudios de alternativas de manejo asociados a los impactos, como se muestra en la Figura 2.

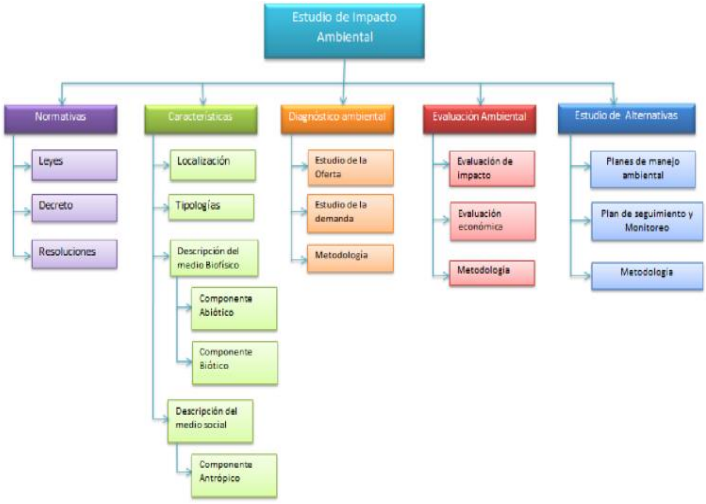


Figura 2: Árbol de dominio inicial

Con el fin de complementar el conocimiento básico del dominio asociado al corpus bibliográfico en el contexto colombiano, se definieron 19 nuevos conceptos básicos y sus relaciones de jerarquía, que se asociaron a los primeros 5 de la clasificación definidos previamente, para un total en este punto de 25 conceptos que complemen lo detallado en la Figura 2.

4.4 Rastreo de los términos en los textos especializados

En la búsqueda de la documentación de la primera fase también se tuvo en cuenta los textos físicos, aparte de la documentación digital que proporcionaron las bases de datos de las diferentes instituciones. Con el fin de obtener la caracterización y el complemento del conocimiento adquirido y haciendo uso de UML correspondiente a la fase 2 de la metodología, se realizó inicialmente un análisis de los términos y su frecuencia de aparición en cada documento. Para el caso de los documentos físicos se hizo un proceso de digitalización con el fin de lograr su análisis mediante herramientas computacionales. Para ello fue necesario escanear y extraer el texto de cada página mediante un Free Online OCR [19]. Los OCR son herramientas que permite digitalizar un artículo, revista, libros, cartas, entre otros documentos, convirtiéndolos a formato digital. Antes de esto se deben escanear o ingresar en formato imagen [20]. Básicamente es una tecnología que permite convertir diferentes tipos de documentos en texto editable.

Luego de tener los textos en formato digital, todos los textos se sometieron a un analizador de textos *Textalyser,* una aplicación *online* para análisis estadísticos de un texto [21]. Este análisis se hizo con el objetivo de encontrar las palabras más repetitivas y frecuentes relacionadas con el tema EsIA. Se alcanzó un conteo de las repeticiones por palabras de los 9 libros, identificando las que deben hacer parte de la jerarquía de conceptos de Estudios de Impacto Ambiental. La información arrojada por el sistema se almacenó en una hoja de cálculo utilizando distintas plantillas para cada libro analizado. En la Figura 3 se detalla una muestra de los términos y sus frecuencias en cada uno de los textos analizados.

Términos	Frecuencia									Total
Preferentes	LIBRO 1	LIBRO 2	LIBRO 3	LIBRO 4	LIBRO 5	LIBRO 6	LIBRO 7	LIBRO 8	LIBRO 9	repeticiones
acci¢n	8		47							55
actividad	46		45	4	20		27		136	278
actores	2									2
acuerdo	17	28		8	24		9		29	115
afectada	4									4
afectado	6									6
agua	31	56	61		36				95	279
aire	6		19							25
alteraci¢n	8		37							45
alternativa	6			4			6			16
ambiental	91	101	230	33	61	6		7	791	1320
ambiente	41	43	76		59	22	28		56	325
anlisis	54									54
antr¢pico	13									13
aprovechamiento	4			4	16	8	7			39
atm¢sfera	3		27							30
atmosferico	4									4
beneficio	2									2
bi¢tico	15									15
biofisico	20									20

Explotacion Hidrocarburos | Libro 5 | Resolucion | Libro 6 | Decreto | Libro 7 | Reglamentacion y Licencia | Libro 8 | est. Impacto Ambiental | lexicográfico | Lib

Figura 3: Depuración de los términos

Posteriormente se depuraron los términos, teniendo en cuenta la frecuencia y la importancia de las palabras y que a través de ellos se pudiera enriquecer de conceptos semánticos el árbol de dominio inicial. El resultado de este trabajo fue un árbol de 140 conceptos jerárquicamente organizados. Un detalle se muestra en la Figura 4.

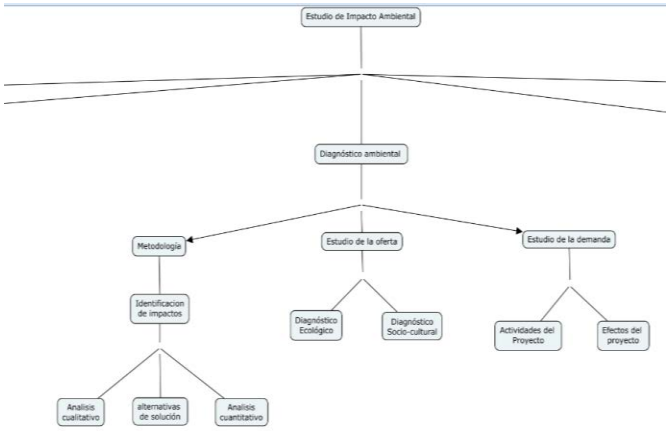


Figura 4: Árbol de dominio final enriquecido con los términos extraídos de Textalyser

Para cada uno de los términos extraídos se identificaron sus significados y sus relaciones entre términos y haciendo uso de una hoja electrónica. La terminología de la ontología Skos core se detalla en la Figura 5.

Conceptos generales incluidos en la ontologia de Skos- Core					
Concepts	**Definition**	**Broaders**	**Narrowers**	**Pref. Labels**	**Alt. Labels**
Estudio de Impacto Ambiental	Es el instrumento básico para tomar decisiones sobre un proyecto, obra o actividad que afecte el medio ambiente natural o artificial	N/A	Normativas Caracteristicas Diagostico Ambiental Evaluación Ambiental Estudio de alternativas	Estudio de Impacto Ambiental	N/A
Normativas	Norma o conjunto de normas por las que se regula o se rige determinada materia o actividad.	Estudio de Impacto Ambiental	Decretos Leyes Legislación Ambiental	Normativas	Reglamento
Leyes	Norma o regla, que dice cuál es la forma en la que debe ser el comportamiento o actuar en la sociedad	Normativas	Procedimientos de sanción Política ambiental Código de minas	Leyes	Legislación legalidad constitución norma
Política ambiental	Es el conjunto de los esfuerzos políticos para conservar las bases naturales de la vida humana y conseguir un desarrollo sostenible	Leyes	Campo de aplicación Principios ambientales	Política ambiental	N/A
Campo de aplicación	Se define en función del tema de estudio de investigación, u objeto de la asistencia técnica y/o capacitación, o elemento muestreado en cuestión, en relación a su aplicación en una actividad	Política ambiental	N/A	Campo de aplicación	Área de aplicación

Figura 5: Significados y relaciones de cada término del dominio final

4.5 Implementación del sistema de conceptos

Para la implementación de la ontología basada en un enfoque tesauros, correspondiente a la fase 3 de la metodología, se hizo utilizó la ontología Skos core implementada en lenguaje RDF [22]. En cuanto a la configuración, cada uno de los términos rastreados en el corpus bibliográfico se implementó sobre la ontología mediante el uso del Protégé [23], un ambiente interactivo para desarrollo de ontologías que ofrece una interfaz gráfica para editarlas y una arquitectura para la creación de herramientas basadas en conocimiento, para construir soluciones basadas en el conocimiento en áreas diversas. La Figura 6 muestra la ontología con los conceptos y relaciones en el ambiente de implementación Protégé.

4.6 Validación con los especialistas

La validación con los especialistas es importante para el desarrollo de cualquier terminología y gracias a sus observaciones y opiniones se pudo lograr la validación de los términos. En la primera fase del proyecto se convocó a dos especialistas en los temas de ingeniería ambiental e ingeniería de sistemas, para que tuviera una mirada más holística en el desarrollo de la ontología. Esto permitió discutir los conceptos propuestos inicialmente en el árbol de dominio, y que fueron el insumo principal para la generación

del tesauro en el área EsIA. A cada especialista se le envío previamente la documentación asociada del proyecto y árbol de dominio inicial. Posteriormente se tomaron acciones frente a las observaciones y sugerencias que tenían de los conceptos, y luego se buscó un árbol más amplio y enriquecido en términos. Al final se logró un consenso en relación con la estructura del mapa conceptual.

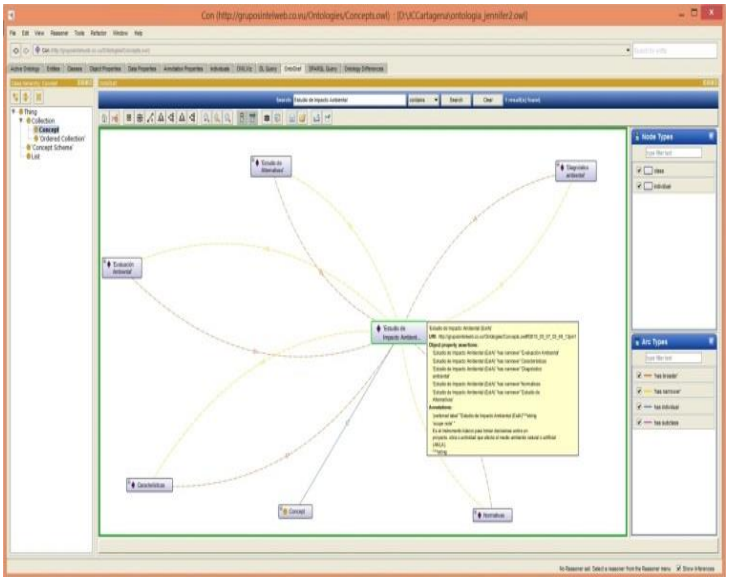


Figura 6: Estructura de la ontología vista desde Protégé

5 Trabajos relacionados

Se enoctró poca literatura relacionada con el caso de ontologías para la representación del dominio en el área de Estudios de Impacto Ambiental (EsIA). A continuación se relacionan los principales trabajos que implementan ontologías en el área EsIA contrastando sus características relacionadas con el alcance del dominio y su contexto geográfico, frente a la ontología detallada previamente. En esta comparación se entiende el alcance del dominio a las grandes temáticas que comprenden un estudio de impacto ambiental (Evaluación de Impacto Ambiental, Plan de Manejo Ambiental y el plan de Contingencias). El contexto geográfico hace referencia al país al que se asocia los conceptos que conforman el dominio de un Estudio de Impacto Ambiental en el marco de sus leyes y normas. La comparación de estos trabajos se resume en la Tabla 3.

Tabla 3: Comparación de trabajos relacionados

Referencia	Descripción	Alcance	Contexto geográfico
[24]	Propone una ontología de procesos y servicios que contiene los principales requisitos relacionados con la gestión ambiental.	Evaluación de impacto ambiental	Canadá
[25]	Muestra la integración efectiva de las cuestiones ambientales por medio de herramientas, que mejoran el intercambio de información y la comunicación, que desempeña un papel crucial para el funcionamiento de la empresa desde una perspectiva ambiental y proporciona una toma de decisiones de soporte de herramientas mediante la combinación de diferentes sistemas de información, que se adapta y reconoce los diferentes elementos relacionados con las funciones de la empresa.	Evaluación de impacto ambiental	España

[26]	Propone el uso de una ontología para la evaluación de impacto ambiental de las actividades humanas, básicamente define la estructuración de la terminología en EIA a partir de información obtenida de normas, legislación y la literatura.	Evaluación de impacto ambiental	España
[27]	Se muestra la recolección de la terminología relacionada a EIA con sus relaciones para facilitar la estructuración y desarrollo de metodologías, mediante el uso de ontologías breves y define cómo construir una nueva ontología sin la cantidad de información que tiene la ontología original.	Evaluación de impacto ambiental	Cuba
[28]	Elabora una ontología que comprende los conceptos fundamentales involucrados en las evaluaciones de impacto ambiental de proyectos mineros.	Evaluación de impacto ambiental	Cuba

Particularmente, de la Tabla 3 se desprende que se encontraron solamente ontologías en el área específica de Evaluación de Impacto Ambiental (EIA), que si bien hace parte de los estudios base que se deben presentar con EsIA, es un dominio conceptualmente más amplio en sus características y entorno que un EIA como el que abarca el presente trabajo. Aunque estas ontologías incluyen de manera general los conceptos referidos a las evaluaciones de impacto ambiental y a las relaciones entre ellos, no comprenden el análisis EsIA, en especial en el contexto colombiano. Esto se debe a que atacan un problema en particular dependiendo del contexto en el que se observe el problema en cuestión, pero pueden ser utilizadas como punto de partida.

6 Conclusiones y trabajo futuro

Este trabajo describe el diseño de una ontología para el desarrollo de EsIA con el propósito de crear una base de conocimiento accesible por usuarios interesados en realizar consultas en el contexto colombiano. La ontología se desarrolló mediante un marco conceptual que contienen definiciones formales, que solucionan problemas de ambigüedad semántica y por tanto permite asegurarse de que se reduzca la polisemia conceptual entre expertos.

La ontología ha sido revisada por especialistas en ingeniería ambiental e ingeniería de sistemas, con el fin de obtener un resultado con suficiente rigor científico. Esta ontología fue desarrollada teniendo en cuenta varias fuentes, muchas de las cuales aportaron grande información al llenado del conocimiento de la misma. Si bien se concibe para ser una herramienta de base de conocimiento de referencia que apoye los procesos de licenciamiento en la generación de EsIA, no es una herramienta que modifica los procedimientos descritos por las normas colombianas en la realización de los mismos.

Como trabajo futuro se propone realizar pruebas de validación específicas a la ontología y relacionadas con indicadores de recuperación de información en el marco del sistema SLO4DTV, que permite la generación de aplicaciones educativas basadas en televisión digital y usando objetos de aprendizaje semánticos [29]. Al interior de este sistema se implementa un buscador semántico de objetos educativos, donde la ontología desarrollada puede ser incluida fácilmente para las pruebas, y que tendrían cobertura temática, exhaustividad de la anotación y precisión en la recuperación de documentos.

Agradecimientos

Este trabajo se realizó en el marco del proyecto de investigación: Desarrollo de un sistema de recomendación de objetos de aprendizaje en el área de educación ambiental

basado en linked Data, con código Hermes 25847 y financiado por la Vicerrectoría de investigación de la Universidad Nacional de Colombia en la convocatoria del programa nacional de proyectos para el fortalecimiento de la investigación, la creación y la innovación en posgrados de la Universidad Nacional de Colombia 2013-2015, modalidad 2: nuevos proyectos de investigación, creación o innovación.

Referencias

[1] Min. Ambiente (1993). Ley 99 de 1993. Ministerio del Medio Ambiente.
[2] Min. Ambiente (2014). Decreto 2041 del 15 de Octubre 2014. Ministerio del Medio Ambiente.
[3] ANLA (2014). Estudio de Impacto Ambiental. Autoridad Nacional de Licencias Ambientales. Online [May 2015].
[4] Arguedas, S. et al. (2004). Lineamientos y herramientas para un manejo creativo de las areas protegidas. Programa de Política y Ciencias Ambientales, Costa Rica.
[5] Montes, A. & Ferrer, R. (2014). Ontología de evaluación de impacto ambiental para proyectos mineros. Mineria y Geología 30(1), pp. 104-117.
[6] Swartout, B. et al. (1997). Toward distributed use of large - scale ontologies. AAAI Technical Report SS-97-06. USC/Information Sciences Institutes.
[7] Zúñiga, H. (2009). Elaboremos un estudio de impacto ambiental. Bogotá: Universidad Distrital Francisco José de Cladas.
[8] Thomas, G. (1993). A traslation approach to portable ontology. Knowledge Acquisition 5(2), pp. 199-220.
[9] RAE (2012). Diccionario de la lengua española. Online [Feb. 2015].
[10] Cambridge Semantics (2004). Introduction to Semantic Web. Cambridge: The MIT Press.
[11] Berners, T. (2002). Bootstrapping the Semantic Web with Redland. Online [May 2015].
[12] UNESCO (2014). Tesauro de la UNESCO. Online [May 2015].
[13] IST (2004). SWAD-Europe. Online [Feb. 2015].
[14] Popkin Software and Systems (1999). Modelado de Sistemas con UML. Online [May 2015].
[15] Gómez, A. & Benjamins, R. (1999). Overview of knowledge sharing and reuse components: Ontologies and problem-solving methods. Proceedings Proceedings IJCAI9 workshopon Ontologies and Problem-Solving Methods, pp. 1-15. Stockholm, Sweden.
[16] Thomas, G. (1993). A traslation approach to portable ontology. Knowledge Acquisition 5(2), pp. 199-220.
[17] Min. Ambiente (2005). Tesauro ambiental para Colombia. Ministerio del Ambiente. Online [Feb. 2015].
[18] INVEMAR (2015). Directriz relacionada con el estilo de direccion en el Invemar. Online [May 2015].
[19] Free Online OCR (). Optical Character Recognition. Online (Feb. 2015].
[20] ABBYY (2015). Acerca de OCR. Online [Feb. 2015].
[21] Textalyser (2015). Text Analyzer. Online [Apr. 2015].
[22] Hansen, S. (2000). Tesauro de derechos económicos, sociales y culturales. Washington: Asociación Americana para el avance de las ciencias.
[23] Chesani, F. (2010). Introduction to Protégé. Online [Feb. 2015].
[24] Uschold, M. & Gruninger, M. (1996). Ontologies principles methods and applications. Knowledge Engineering Review 11(2), pp. 1-69.
[25] Muñoz, E. et al. (2013). Considering environmental assessment in an ontological framework for enterprise sustainability. Journal of Cleaner Production 57, 149-164.
[26] Garrido, J. (2011). Ontologies for environmental impact assessment in different human activities. Granada: Universidad de Granada.
[27] Garrido, J. & Requena, I. (2013). La ontología de la evaluación de impacto ambiental y sus ontologías breves. Revista Electrónica universidad de Jaén 5(6), pp. 1-12.
[28] Montes, A. & Ferrer, R. (2014). Ontología de evaluación de impacto ambiental para proyectos mineros. Minerología y geología 30(1), pp. 104-117.
[29] Guzman, J.; Torres, I. & Alvarez, J. (2014). Propuesta de un generador de aplicaciones educativas basadas en television digital usando arquitectura de computa en la nube. Revista colombiana de tecnologia de avanzada 2(24), pp. 33-41.

Arquitectura empresarial y metodologías ágiles - Una combinación efectiva para hacer frente a los frecuentes cambios en el negocio

Alejandro Álvarez C.
Albeiro Patiño B.
Fundación Universitaria Católica del Norte – Colombia

1 Introducción

La Globalización, los Tratados de Libre Comercio y la fuerte competencia que esto implica, demandan de las empresas modelos de desarrollo de software que les permitan ser rápidas y efectivas al tiempo que eficientes y oportunas. Refiriéndose a América Latina la Organización para la Cooperación y el Desarrollo Económico OCDE afirma que la innovación tecnológica resulta clave para promover el cambio estructural necesario, que favorezca la diversificación, incremente la productividad y reduzca las brechas tecnológicas en la región [5]. Una responsabilidad que no se le puede dejar solamente a las áreas de desarrollo, sino que debe partir, en primera instancia, de las áreas de arquitectura donde se deben dar los lineamientos para que las actividades de tecnología se puedan llevar a cabo de forma clara y ágil.

En procura de definir un modelo maduro que les permita a las áreas de tecnología aumentar la productividad de sus desarrollos y disminuir los costos y recursos (técnicos y humanos), requeridos para optimizar el *time to market*, la Arquitectura Empresarial y las metodologías ágiles se constituyen en importantes pilares. La primera, como promotora de un *framework* o metodología que guíe la actuación de las personas y las áreas, y las segundas, como método de trabajo.

La Arquitectura Empresarial propone la creación de un modelo que permita abordar la tecnología como un complemento y apoyo del negocio, de forma que las áreas de Negocio y de TI evolucionen de manera alineada, conjunta, dinámica y ágil. De acuerdo con Bernhard Hitpaas [1], la tendencia de toda organización es crecer y desarrollarse, y el desarrollo es un proceso lento y gradual que conduce al conocimiento exacto de sí misma y a la plena realización de sus potencialidades. Por eso es que la tecnología debe cumplir su papel de evolucionar al mismo ritmo que el negocio impone, permitiendo un modelo que de forma coordinada plasme la realidad presente de la compañía y modele su visión futura. La evolución del modelo debe ser continua y sostenible, aplicable y fácil de implementar, soportar, actualizar y mantener, sin que para ello se deba incurrir en costos elevados. Por su parte, desde su concepción, las metodologías ágiles surgieron como alternativa para subsanar los tantos y tan variados inconvenientes que enfrentan los desarrolladores al usar metodologías tradicionales. Pablo Lledó [3] relaciona algunas de ellas: falta de información, cambio de prioridades, esperas en colas, apagar incendios, falta de recursos, multi-tareas.

En este trabajo se abordan ambos temas: la Arquitectura Empresarial y las metodologías ágiles, y se muestra que su conjunción representa una combinación efectiva para hacerle frente a los frecuentes cambios de negocio. Se abordan algunas metodologías ágiles de Arquitectura en proceso de maduración y que están siendo usadas, que podrían servir para quienes planean diseñar y desarrollar software, con miras a mejorar la oportunidad tecnológica y el *time to market* de la empresa. Si bien es aplicable tanto a empresas grandes como medianas y pequeñas, podría ser más útil en las grandes y en

algunos casos en las medianas, que es donde las metodologías tradicionales se vuelven más burocráticas y paquidérmicas.

Para la discusión y el análisis de los temas planteados aquí, los autores se basan en buena medida en su experiencia real y práctica, así como en una amplia bibliografía que se ofrece. Además, puede servir para profundizar en los conceptos planteados. Para llegar a la lista final de la bibliografía seleccionada, se realizaron búsquedas con base en expresiones claves, tales como Metodologías de Desarrollo de Software, Arquitectura Empresarial, *Framework* de Arquitectura, *Framework* de Desarrollo, entre otras. También se recurrió a la lectura y el análisis de revistas, libros e informes de organismos multilaterales, tales como la OCDE, el Banco Mundial, el Fondo Monetario Internacional, y otros tantos, toda vez que el propósito fundamental es proponer herramientas que ayuden a las empresas a aumentar la productividad, reducir los costos y mejorar el *time to market*, con un enfoque ágil.

Aunque los temas de Arquitectura Empresarial y metodologías ágiles datan de hace más de diez años, se procuró que la documentación para soportar el análisis fuera lo más reciente posible (de preferencia posterior a 2012), considerando que, particularmente, la tecnología sigue una línea acelerada en la que lo vanguardista pasa rápidamente a ser caduco. A continuación se revisan los siguientes temas: qué es un *framework* de arquitectura, para qué sirve y cómo se puede aplicar, qué es un modelo actual, un modelo transitorio y un modelo objetivo de arquitectura empresarial, que es un GAP, cuál es el rol de un arquitecto en proyectos ágiles y cuál ha sido el enfoque de las metodologías ágiles con mejores resultados en la práctica para que los equipos de trabajo realicen sus análisis y diseños de arquitectura empresarial.

2 Metodologías de Arquitectura Empresarial

En la Figura 1 se muestran algunas de las razones por las cuales las arquitecturas se vuelven caóticas y difíciles de administrar en las empresas.

Figura 1: Razones por las que se complejizan las arquitecturas en las empresas (A partir de [4])

Gestionar y administrar una arquitectura regida por este tipo de antecedentes y premisas de crecimiento es difícil y complejo. De ahí que se haga necesario definir un mecanismo de gestión y administración, y en este punto es donde surge la Arquitectura Empresarial. Institute of Electrical and Electronics Engineers IEEE ratificó en 2000 el estándar 1471-2000 como una práctica recomendada para la descripción de sistemas intensivos de software, creando un modelo conceptual para la descripción de metodologías específicas de arquitectura, e incluyendo una aproximación fiable al *framework* de arquitectura empresarial. Este estándar sirvió como fundamentación conceptual para entender y modelar los puntos de vista individuales y las perspectivas y preocupaciones de los usuarios de negocio.

Por su parte, The Open Group Architecture Framework TOGAF [2] define la Arquitectura Empresarial como la organización fundamental de un sistema, conformada

por sus componentes, las relaciones entre ellos y su entorno y los principios que gobiernan su diseño y evolución. TOGAF propone trabajar con cuatro arquitecturas:

1. Arquitectura de negocio, se ocupa de la estrategia de negocio, gobierno, organización y procesos clave de la compañía.
2. Arquitectura de datos, es la responsable de definir la estructura de datos lógicos y físicos que posee una organización y sus recursos de gestión de datos.
3. Arquitectura de aplicación, debe atender todo lo concerniente a las aplicaciones individuales que se deben implementar, sus interacciones y las relaciones que tienen con los procesos de negocio.
4. Arquitectura tecnológica, debe garantizar las capacidades software y hardware para apoyar la implementación de los servicios de negocio, de datos y de aplicación.

Siguiendo con TOGAF, el *framework* de Arquitectura Empresarial proporciona un modelo detallado de productos de trabajo arquitectónicos, incluyendo entregables, artefactos dentro de los entregables y los Bloques de Construcción de la Arquitectura (ABB) que los entregables representan.

En cierta forma, Zachman [8] fue quien acuñó el concepto de Arquitectura Empresarial al que se refirió como Sistema de Información en su trabajo. Pero ante todo su propósito era sentar las bases para alinear la tecnología y el negocio, y procurar lograr mejores resultados en tiempo, costos, recursos y *time to market* para la compañía. De esta forma Zachman creó la punta de lanza para abrir el camino. De sus raíces crecieron otras visiones de Sistemas de Información y Arquitectura Empresarial. El Departamento de Defensa de los Estados Unidos creó en 1994 el Technical Architecture Framework for Information Management (TAFIM); posteriormente, en 1996, evolucionó al Federal Enterprise Architecture Framework (FEAF), que nuevamente en 2002 cambia el nombre por el actual Federal Enterprise Architecture (FEA). De esta base nacen los trabajos de The Open Group, conocido como TOGAF.

Quizá los *frameworks* de Zachman y TOGAF sean los más conocidos y utilizados a nivel mundial, tanto en el sector público como en el privado. Sin embargo, y a partir de ellos, muchas compañías han definido los suyos buscando siempre mantenerse dentro de los lineamientos generales dictados por los originales. También particularizan los suyos de forma que se ajusten a sus necesidades y a sus formas particulares de ver y aplicar la alineación de negocio y tecnológica. En la Figura 2 se muestra un *framework* definido para una empresa particular, que considera que la mejor forma de alinear negocio y tecnología es ver a la Arquitectura Empresarial con especial atención a las aristas de: 1) procesos, 2) información, 3) aplicativa, 4) integración, 5) técnica, y 6) de seguridad.

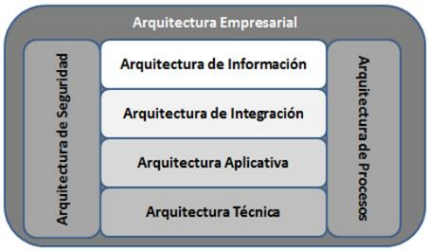


Figura 2: Framework de Arquitectura Empresarial para una particular empresa (Autor)

2.1 Utilidad de un *framework* de Arquitectura Empresarial

La Arquitectura Empresarial es una representación gráfica en la que se muestran diagramas y especificaciones teóricas, que facilitan la comprensión de los procesos de

negocio, así como la información, los Sistemas de Información, los servicios que permiten sus interacciones, etc. De forma que se puedan optimizar los recursos (técnicos y humanos), mantener un adecuado balance de la relación costo/beneficio y tener satisfechos a los usuarios. Una de las principales ventajas de la Arquitectura Empresarial es que facilita la adaptabilidad del negocio, y como afirman Sáez et al. [7], los Sistemas Adaptativos Complejos CAS pueden reorganizar su estructura interna sin la intervención de ningún agente externo. Esta característica, producto un aprendizaje inconsciente, aumenta la probabilidad de que el sistema sobreviva en entornos turbulentos e inestables Otras ventajas de tener claramente definida la Arquitectura Empresarial son:

- Los procesos se identifican y mantienen más fácilmente.
- Se garantiza de forma estructurada una permanente alineación entre el negocio y la tecnología.
- Se tiene identificados claramente los componentes tecnológicos que se usarán para soportar los procesos de negocio.
- Se hace más fácil la toma de decisiones en relación con la tecnología y en función de los procesos del negocio.
- Se pueden tomar decisiones más fácil y rápidamente, con oportunidad y garantizando una buena relación costo/beneficio a través del reuso de activos de TI (aplicaciones, datos e infraestructura)
- Se hace más fácil identificar dónde están los focos de interés y de mayor oportunidad, y los que presentan mayores riesgos y preocupación, lo que sirve como guía para la evolución del negocio a corto, mediano y largo plazo.
- Es más fácil determinar los impactos que se visualizan ante cambios en los procesos de negocio y en la tecnología.

2.2 Modelo actual, modelo transitorio y modelo objetivo de la Arquitectura Empresarial

De acuerdo con Ross, Weill y Robertson [6], la Arquitectura Empresarial, como conjunto integrado de procesos de negocio, puede mejorar los servicios para los clientes, suministrar mejor información para la toma de decisiones y permitir cambios en una parte de la empresa para alertar a otras sobre acciones que deben tomar. De lo anterior se desprende que hay diversos aspectos críticos que atañen a la Arquitectura Empresarial, tales como los procesos de negocio y los servicios para los clientes, la información y los cambios que impactan las decisiones que se deben tomar. Estos últimos, entre todos, son quizás los que más relevancia cobran cuando se trata de administrar una Arquitectura Empresarial, por lo que mantenerla actualizada es vital para el cumplimiento de los demás aspectos. Los cambios deben reflejarse en los modelos de arquitectura y si no sucede así, lo más seguro es que la Arquitectura misma fallará como modelo de alineación de negocio y tecnología.

Ross et al. [6] proponen que el mejor soporte para la estrategia de una compañía es el modelo que defina. Pero como tanto el negocio como la tecnología cambian permanentemente, es necesario garantizar una evolución sistemática en los modelos. No es fácil hacer un cambio radical en la Arquitectura Empresarial, pero de hacerlo es posible que se impacte negativamente al negocio por posibles fallas en la tecnología. Lo ideal es definir un *roadmap*, en el que se pueda visualizar claramente el hoy, el mañana cercano y el futuro a mediano plazo, tanto en los modelos de negocio como en los de Arquitectura, o lo que es lo mismo, en la Arquitectura Empresarial de la compañía. En otras palabras, la Arquitectura Empresarial debe procurar contar con diseños que reflejen la situación actual de la empresa y con diseños que muestren la situación que visualiza para su futuro

mediano. Pero, además debe diseñar un modelo que permita, de forma paulatina y gradual, ir acercándose a ese modelo objetivo. En la Figura 3 se representa el *roadmap* por el que debe propender la Arquitectura Empresarial para garantizar una efectiva evolución desde un modelo actual a un modelo objetivo.

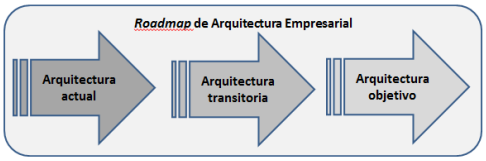


Figura 3: Roadmap de evolución de una Arquitectura Empresarial (Autor)

2.3 Brechas entre los modelos de Arquitectura Empresarial

Para progresar desde el modelo actual al modelo objetivo de la Arquitectura Empresarial se debe evidenciar las falencias y vacíos, es decir, las brechas o GAPs que se deben cubrir. Estos GAPs pueden deberse a aspectos puntuales en cualquiera de las vistas de la arquitectura: la técnica, la aplicativa, la de integración, la de información o la de seguridad. Las razones para que existan en la compañía pasan por la disponibilidad de recursos técnicos o humanos, falta de conocimiento, costos, oportunidad, etc., que se deben evidenciar y reflejar en la arquitectura transitoria, y definirse la forma y fecha para la cual se espera cubrir.

En cualquier caso, siempre se podrá considerar osado, sobre todo en proyectos grandes, intentar pasar de un modelo de arquitectura actual a un modelo de arquitectura objetivo sin definir un modelo transitorio que permita contar con todos los recursos técnicos y humanos requeridos, el personal debidamente capacitado, el presupuesto garantizado, etc. Algo que confirman Matsumura, Brauel y Shah [4] al recomendar que no se debe intentar una aproximación de *big bang*, es decir, hacer realidad el esquema global de la Arquitectura mediante un único, interminable y costosísimo proyecto. Lo mejor es seleccionar y establecer una secuencia de pequeños proyectos, de forma que cada uno aporte por sí mismo una ventaja cuantificable para el negocio.

Cubrir un GAP en la Arquitectura Empresarial podría ser la instalación de un canal de comunicación o una aplicación, o apropiar conocimiento en la gestión y administración de una determinada herramienta; también garantizar que una transmisión de archivos, que tradicionalmente se ha hecho de forma no-segura, se haga de forma segura, o cambiar el mecanismo de intercambio de información entre dos aplicaciones que típicamente han *hablado* mediante servicios web a un mecanismo basado en colas. Incluso, puede ser la conjunción de varios de estos u otros inconvenientes, o hasta el emprendimiento y finalización de uno o varios proyectos. En la Figura 4 se muestra la arquitectura actual y la arquitectura objetivo, y entre ellas los GAPs que se deben superar.

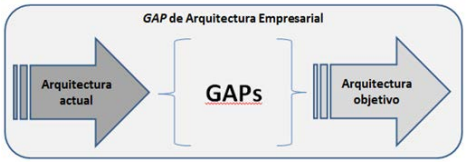


Figura 4: GAPs de la Arquitectura Empresarial (Autor)

Matsumura, Brauel y Shah [4] afirman que cada proyecto debe proporcionar un retorno de la inversión, y además motivar la realización de futuros proyectos que le permitan continuar surcando el espacio hacia sus objetivos. También dicen que a medida que se implementa se pueden ir perfeccionando y automatizando los procesos de

implementación de la Arquitectura, hasta alcanzar una condición en la que los esfuerzos desaparecen, y que se denomina estado de *ingravidez*.

2.4 Las metodologías ágiles sustentan la Arquitectura Empresarial

En este punto es donde cobran valor y relevancia las metodologías ágiles. En el proceso de sacar adelante uno, o en el peor de los casos, varios proyectos de forma simultánea, como ruta crítica para pasar de una arquitectura actual a una arquitectura objetivo, puede ser necesario tomar muchas y muy variadas decisiones, y probablemente de forma apresurada. Si a esta dificultad se le suma el que es una empresa grande, donde la burocracia pesa y relentiza los procesos, el asunto se agrava. En una situación como esta las metodologías tradicionales no son tan útiles, porque las actividades se realizan más en secuencia que en paralelo, es necesario involucrar a personas muchas veces solamente en virtud de su rango, o por escalamiento. En este caso, los procesos se hacen lentos, los reprocesos aumentan y las decisiones se retardan.

Las metodologías ágiles se presentan como una nueva apuesta para los desarrolladores de software, justamente en respuesta a las falencias de las metodologías tradicionales, incluso con mayor aceptación por parte de los líderes PMO. Esencialmente por ser simples y prácticas, por reunir en una misma mesa a todo el equipo de trabajo, ser más colaborativas y flexibles, y sobre todo por lograr resultados en menos tiempo. Una vez conocido el modelo actual y el objetivo de la Arquitectura Empresarial es necesario definir una forma controlada de avanzar. Lo mejor es hacerlo mediante iteraciones, es decir, de forma incremental, evolutiva, escenificada, o en espiral, porque no es algo nuevo. Lo único que se busca es un mecanismo de retroalimentación que permita saber con precisión cuál es la situación en intervalos cortos y frecuentes.

En la misma mesa de trabajo deben estar sentados, de preferencia, los integrantes de las diferentes áreas que conforman la cadena de desarrollo; y en conjunto, sin perder de vista ni desconocer la responsabilidad que recae sobre cada uno, serán los encargados y deben identificar: servicios de negocio, requisitos para la descripción de los servicios, métricas de rendimiento de los servicios, estándares de interoperabilidad, esquemas de datos, políticas, requisitos de clasificación y localización de los servicios [4]. Una vez identificados estos insumos el equipo debe trabajar en la resolución de los problemas planteados, documentarlos y proceder luego a su implementación.

2.5 Estado de las metodologías ágiles de Arquitectura Empresarial

Actualmente, las empresas buscan grandes transformaciones a través de un enfoque mucho más ágil de los modelos implementados. Parecería que se está formando una nueva tendencia que conlleva a una transformación en la forma como se aborda la práctica de la arquitectura empresarial. Esta tendencia está relacionada con conceptos actuales y relacionados entre sí, tales como metodologías ágiles, *cloud computing*, transformación digital, iteraciones continuas y desarrollo rápido de software. Todo esto está marcado por la necesidad de acelerar el negocio y en particular el ciclo de desarrollo y la salida al mercado de nuevos productos y servicios.

La arquitectura empresarial todavía es una práctica que pueden implementar solamente empresas grandes con capacidad de ejecución. Una de las principales razones para la no-adopción es que los tiempos de implementación y maduración son prolongados y a corto plazo no se obtienen beneficios tangibles, por lo que variables como la incertidumbre en tiempos y costos únicamente pueden ser asumidas por grandes empresas con grandes presupuestos. La metodología TOGAF contiene un conjunto de

guías y técnicas para desarrollar una arquitectura empresarial, y su desarrollo y puesta en marcha contempla ocho fases. Cada una requiere alrededor de 20 entregables, lo que necesita tiempos de implementación prolongados y una restricción de entrada en términos de agilidad en la adopción de la metodología para el desarrollo de una Arquitectura empresarial [9]. Sin embargo, la necesidad de responder rápidamente a los requerimientos del negocio y a las nuevas tendencias del mercado es una exigencia a la que todas las empresas están expuestas sin importar su tamaño, recursos o sector. Diversos marcos metodológicos generales y particulares están disponibles para cada sector industrial, y ofrecen una guía de implementación a alto nivel. Pero cuando se organiza una iniciativa en su desarrollo, siguiendo algunos de los modelos de implementación disponibles, los tiempos de puesta a punto e implementación y la cantidad de entregables y artefactos es una restricción para la mayoría de empresas, especialmente Pymes.

Las empresas y el mercado son cada vez más exigentes y demandan alta productividad, mayor flexibilidad a los constantes cambios, disminución de los ciclos de entrega de servicios de software y la implementación de capacidades de negocio que habiliten los nuevos productos y servicios [10]. La adopción de metodologías agiles es una forma para lograr estándares de eficiencia y satisfacción del negocio. Es común encontrarlas aplicadas a distintas disciplinas, tales como el desarrollo de software bajo metodologías SCRUM, eXtreme Programming, Lean y Six Sigma, entre otras, que ofrecen un modelo para gestionar eficientemente el ciclo de desarrollo de software, con una cantidad de actividades alcanzable y tareas en contextos ágiles.

Para la Arquitectura Empresarial el concepto de agilidad no ha evolucionado ni ha sido incorporado en sus modelos de referencia, como si lo han hecho las disciplinas descritas. Sin embargo, existen percepciones divergentes en la relación de uso entre las metodologías ágiles y la arquitectura. Si bien el agilismo sirve para mejorar la agilidad en el desarrollo de proyectos, también ha sido criticado por practicantes y académicos debido a la falta de un enfoque en la práctica en la Arquitectura [17, 18]. De acuerdo con Beck [19], algunas metodologías ágiles, como eXtreme Programming (XP), sugieren la eliminación de artefactos de Arquitectura que no estén relacionados con la versión actual del sistema que se está tratando en el proyecto. Algunos autores también sugieren que las metodologías ágiles son demasiado ligeras en cuanto a cumplir con las necesidades arquitectónicas y no son abordadas con suficiente profundidad y rigurosidad [20].

Estudios documentados sobre la relación entre metodologías ágiles y Arquitectura muestran que apenas si la abordan, y sugieren hacerlo de forma mas amplia y con nuevos métodos asociados directamente al trabajo arquitectónico, el cual es considerado como una de las tareas más importantes en el diseño inicial de los proyectos [21, 24, 25]. A esto se suma la creciente necesidad de reconocer la importancia del buen diseño en las etapas iniciales del proceso de desarrollo de software, y de integrar métodos de arquitectura en los enfoques ágiles [22]. En el mismo sentido, para Falessi et al. [23] el trabajo de Arquitectura es relevante, especialmente en el entendimiento y modelamiento del sistema, la documentación de las alternativas o escenarios de implementación, la oportunidades en el escalamiento de las prácticas ágiles de forma transversal en otros proyectos y la reusabilidad de componentes para soportar futuros requerimientos.

2.6 El rol del arquitecto ágil

En la ejecución de proyectos ágiles se observa la necesidad de tener en el equipo un rol que haga las veces de *architect owner,* quien será el responsable de diseñar la

arquitectura de una forma distinta y especializada para proyectos agiles. Es to sería similar al rol actual de *product owner,* quien es responsable de la definición y validación de los requisitos, o al *scrum master* como responsable de implementarlos. El *architect owner* tiene un perfil técnico y es responsable de la definición y el modelamiento de la arquitectura, con una visión transversal que facilite su evolución en el corto, mediano y largo plazo con criterios de simplicidad y flexibilidad [24]. Tradicionalmente, el arquitecto se ha encargado exclusivamente de su definición y de verificar su cumplimiento. En el enfoque ágil, realiza su función de forma colaborativa con el equipo, que tiene una función de co-ayudar a su evolución en el tiempo. De forma similar, el rol de arquitectura puede ser transicional a otros miembros del equipo. El *architect owner* posee un rol senior, lo cual le facilita servir de mentor a sus pares. Uno de los retos que tiene bajo el esquema de los equipos agiles es definir las guías de implementación de TI, bien sea de seguridad, información, integración, etc., con la agilidad necesaria que no estanque el normal desarrollo del equipo.

Si bien la arquitectura puede ser definida en consenso con el equipo, en algunas ocasiones no se llega a un acuerdo, y entonces el arquitecto debe ser quien tome decisiones con la oportunidad necesaria para no retrasar al equipo ni al proyecto. En la metodología tradicional de cascada el arquitecto construye las definiciones en las etapas iniciales del proyecto, particularmente en la fase de visión o planeación; pero en los proyectos agiles tiene una característica natural, definida en el manifiesto ágil y asociada a los requisitos cambiantes entre un *sprint* y otro. Por eso el rol del arquitecto en proyectos ágiles es más intensivo. Usando estas metodologías, muchos cambios solicitados entre los *spints* pueden impactar la arquitectura y requerir nuevas definiciones, lo que la convierte en un rol con mayor interacción en la participación a lo largo del proyecto. El trabajo del arquitecto no solamente aparece al inicio sino también durante la ejecución del proyecto, particularmente después de la planeación del *sprint.*

En la Figura 5 se explica cómo se organiza el *architect owner* y otros miembros del equipo entorno a los requisitos definidos, y donde debe colaborar con los otros miembros con el objetivo de evolucionar la arquitectura de la solución. Generalmente, los subsistemas tienen definidos sus propios *architect owners*, como responsables de visionar la arquitectura al inicio del proyecto y de coordinar la toma de decisiones arquitectónicas con el equipo y con el apoyo de asesores de otros dominios.

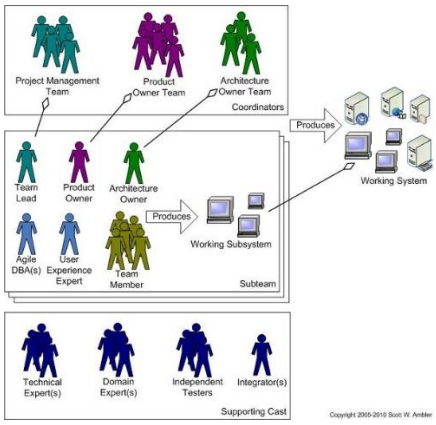


Figura 5: Estructura organizacional en proyectos ágiles [13]

Díaz et al [26] presentan otro enfoque (Figura 6) a través de un modelo adaptado de la metodología SCRUM, en el que el *product owner* define las características del producto y que son listadas en las historias de usuario, conocidas como *product backlog*.

Posteriormente, estas historias son priorizadas de acuerdo con las necesidades del negocio y se asignan a las iteraciones o *sprints*. Cada una de las iteraciones comienza con una sesión de planeación, donde el *product owner* y el equipo planean lo que se debe hacer. En este modelo adaptado de SCRUM, el arquitecto ágil realiza su labor en conjunto con el equipo y de forma evolutiva e incremental. Este enfoque es flexible y los cambios se permiten a través de mecanismos de extensión. El modelo determina que el trabajo del arquitecto se enfoca en la toma de decisiones de diseño, dependencias, limitaciones y renuncias. Estas variables deben ser balanceadas con las prioridades del negocio. Durante la sesión de planeación del *sprint* se define los modelos de arquitectura que serán implementados. En la reunión de revisión del *sprint* el *product owner* evalúa el producto potencialmente entregable y el nivel de adherencia o cumplimiento de la arquitectura.

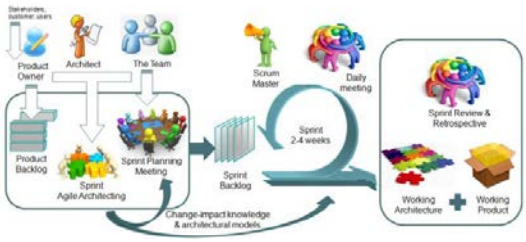


Figura 6: Modelo de SCRUM adptado con arquitectura ágil [26]

3 Aproximación a algunas metodologías ágiles de Arquitectura Empresarial

Las metodologías ágiles fueron concebidas para procesos de desarrollo de software y enfatizan en un desarrollo rápido y flexible. La definición de las arquitecturas, en alguno de sus campos de acción (aplicaciones, información, tecnológica, de integración o de seguridad) es una tarea importante en la planeación y ejecución de los proyectos ágiles. La combinación de metodologías se hace necesaria para lograr la agilidad deseada. Unir los beneficios de estas metodologías con las tradicionales es una opción que ha tomado fuerza en la transición a nuevos modelos de gestión tecnológica.

3.1 Arquitectura de dos velocidades

La habilidad para desarrollar y entregar nuevos productos de forma oportuna se ha convertido en un factor competitivo para las empresas. Uno de los conceptos que está surgiendo para administrar la complejidad y agilizar los ciclos de entrega de productos es la Arquitectura de dos velocidades, o *Two-speed architecture*. Este tipo de arquitectura requiere una ejecución rápida de cara al cliente, que se conjuga con una arquitectura tradicional centrada en el legacy [11]. Para lograrlo se requiere componentes modulares y altamente desacoplados, para permitir el despliegue rápido de nuevos programas o servicios y evitando en mayor medida el largo trabajo de integración. Adicionalmente, se debe definir y construir microservicios (pequeñas funcionalidades desacopladas), que estarán disponibles en los canales de interacción con el cliente y con un tiempo de implementación reducido, generalmente días, a diferencia de los servicios rutinariamente concebidos. Todo esto debe estar acompañado de cero tiempos de indisponibilidad, que afectan la confianza y experiencia del cliente, además de analítica en tiempo real, alta capacidad de configuración y reusabilidad de los componentes. Sin embargo, crear tal capacidad empresarial no necesariamente es fácil y requiere de la conformación de nuevas herramientas y la modificación de procesos de gobierno de TI [27].

Estas capacidades pueden ser naturales para las empresas modernas, pero las establecidas y con larga permanencia en el mercado pueden encontrar dificultad a la hora

de desarrollar las capacidades digitales, debido a que todavía tienen procesos de negocio soportados en sistemas *legacy* en el *back end*. Estas empresas tienen una alta presión para renovar sus arquitecturas a través de proyectos de transformación de corta duración. Se hace necesario entonces crear un modelo de gobierno que facilite la interacción entre TI y el negocio de manera clara y efectiva, para garantizar la transformación a la arquitectura digital. El modelo de arquitectura de dos velocidades también se enfrenta a retos, principalmente en la transformación de sistemas legados, que generalmente tienen mayor complejidad a la hora de evolucionarlos haciendo que los proyectos de transformación tengan mayor costo y duración.

3.2 *Lean architecture*

El concepto *lean architecture* nace de la aplicación de los principios del sistema de producción de Toyota y mantiene los mandatos de la metodología *lean* en cuanto a deshacerse de lo que no genera valor intrínseco. Esta arquitectura se relaciona con un concepto de arquitectura ligera combinada con la consistencia y la reducción del desperdicio. El modelo ofrece el concepto de que la arquitectura ágil debe ser definida de acuerdo con la experiencia del equipo, lo que implica que el sistema puede ser mejorado y evolucionado día a día en la medida en que los requisitos se documentan y el equipo mejor la comprensión y claridad de los mismos [16]. Se trata de un enfoque incremental dirigido por la experiencia y el conocimiento del equipo y que honra el valor de la adaptación. A medida que el equipo madura en el entendimiento de los requerimientos, la arquitectura madura a su ritmo.

El modelo *lean architecture* también adopta la perspectiva de que la documentación innecesaria es un factor que genera alto costo, y en la mayoría de casos es innecesaria y termina archivada y sin uso. También enfatiza en la comunicación a través de medios eficientes y en la discusión cara a cara. Esta metodología critica y evita la excesiva planeación con enfoques a largo plazo, centrándose en lo que es importante ahora e inmediato para tomar decisiones y producir productos de forma inmediata, proporcionando valor verificable y comprobable a los usuarios del negocio.

3.3 Scaled Agile Framework SAF

Esta metodología es una práctica asociada al modelo *lean* y está enfocada principalmente en proveer un conjunto de patrones que puedan ser usados para construir nuevas soluciones software, y evolucionar las existentes de forma rápida. Históricamente, las metodologías SCRUM, XP y otras, se han enfocado en el equipo y en el proceso de desarrollo. Mientras que SAF presenta una vista ejecutiva y unificada que permite subir de nivel para análisis y tendencias o para desagregar en mayor detalle para la toma de decisiones [14]. Generalmente, un equipo de trabajo bajo la metodología SAF se compone de entre 8 y 10 miembros, que tienen una visión *end-to-end* de los requerimientos. También existen proyectos que pueden contener varios equipos con tamaños diversos y con integrantes de distintos perfiles que tienen el objetivo de sincronizar y alinear los *sprint's* y las entregas, de forma tal que la labor de codificación se puede realizar de forma paralela. Estos equipos se reúnen al comienzo de cada ciclo de salidas o liberaciones para definir los incrementos de cada *sprint.* En SAF, para asegurar que el desarrollo se hace bien se sugiere un conjunto de prácticas que están dirigidas más hacia la prevención que a las pruebas posteriores. El modelo comienza con un diseño de arquitectura ágil considerado como emergente o ascendente, y que evoluciona de forma colaborativa. Esto parte de uno de los principios del manifiesto ágil, en el que el mejor diseño de arquitectura nace de equipos que son autónomos [15]. Uno de principios bajo el modelo SAF es

construir la arquitectura más simple posible, que permita el funcionamiento del sistema con los requerimientos definidos ya incorporados. Esto soporta uno los principios del manifiesto ágil que aduce que son bienvenidos los requisitos cambiantes incluso cuando se ha iniciado a labor de codificación [15].

4 Conclusiones

A pesar que las metodologías ágiles de desarrollo de software han recibido atención durante los últimos años, ha habido percepciones divergentes sobre la relación entre métodos ágiles y arquitectura. Diversos autores y estudios han criticado el rol de la arquitectura en los proyectos de desarrollo de software, por el retraso que generan y la posibilidad de crear una arquitectura a medida que el proyecto evoluciona. Sin embargo, si en las etapas previas no se cumplen las condiciones de diseño establecidas por la arquitectura, las correcciones en las etapas posteriores son costosas [25].

La mayoría de *framework* o metodologías tradicionales de Arquitectura se pueden usar en ambientes ágiles, porque ofrecen un enfoque incremental e iterativo. Pero las nuevas metodologías enfocadas en agilidad ofrecen métodos más claros para su desarrollo [12], porque están enfocadas en un diseño incremental y colaborativo de la Arquitectura que debe proporcionar los requisitos mínimos para soportar los requerimientos del negocio, y posteriormente evolucionar a un estado deseado de la arquitectura. El uso de las metodologías agiles en el trabajo de la Arquitectura se ha desplegado de forma masiva por sus ventajas en términos de acortar los tiempos de entrega de soluciones que soportan la evolución de las empresas. Sin embargo, es necesario modificar los procesos de gobierno de TI para que soporten su uso.

Referencias

[1] Hitpass, B. (2012). Business Process Management (BPM). Fundamentos y conceptos de implementación. Santiago: Editorial BHH Ltda.

[2] Josey, A. et al. (2013). TOGAF versión 9.1 - Guía de Bolsillo. Zaltbommel: Van haren Publishing.

[3] Lledó, P. (2014). Gestión Lean y ágil de proyectos. Lean Project Management. USA: PabloLledó Ediciones.

[4] Matsumura, M. et al. (2009). Adopción de SOA para Dummies®. New York: Wiley Publishing Inc.

[5] OCDE (2013). Perspectivas económicas de América Latina 2014 - Logística y competitividad para el desarrollo. México: OCE – CEPAL.

[6] Ross, J.; Weill, P. & Robertson, V. (2006). Enterprise Architecture as strategy: Creating a foundation for business execution. USA: Harvard Business School Press.

[7] Sáez, V. et al. (2003). Temas básicos de innovación tecnológica en las empresas. Madrid: Universidad Politécnica de Madrid.

[8] Zachman, J. (1987). A framework for Information Systems. The IBM Systems Journal 26(3), pp. 454-470.

[9] González, S. (2013). Enterprise Architecture and Organizational Transformation. Proceedings Delivering Business Value through Enterprise Transformation, pp. 12-19. San José, Costa Rica.

[10] Glaiel, F. (2012). Agile project dynamics: A Strategic project management approach to the study of large-scale software development using system dynamics. Massachusetts: MIT Sloan School of Management.

[11] Bossert, O., Ip, C. & Laartz, J. (2014). A two-speed IT architecture for the digital enterprise. USA: McKinsey Quarterly Report.

[12] Coplien, J. (2012). The architecture owner role: How architects fit in on agile teams. Online [Mar. 2015].

[13] Heusser, M. (2015). Introducing the scaled agile framework. CIO Magazin. Online [Mar. 2015].

[14] Cunningham, W. (2001). Manifesto for agile software development. Online [Feb. 2015].

[15] Coplien, J. & Bjørnvig, G. (2010). Lean Architecture for agile software development. West Sussex: Wiley.

[16] Dybå, T. & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. Information and Software Technology 50(9-10), pp 833–859.

[17] Hochstein, L. & Lindvall, M. (2005). Combating architectural degeneration: A survey. Information and Software Technology 47(10), pp 643-656.

[18] Beck, K. & Andres, C. (2004). Extreme Programming Explained: Embrace Change. Boston: Wesley.

[19] Breivold, H. et al. (2010). What does research say about agile and architecture? Proceedings Fifth International Conference on Software Engineering Advances, pp 32-37. Nice, France.

[20] Baba, M., & Abrahamsson, P. (2008). Architecture-Centric Methods and Agile Approaches. Lecture Notes in Business Information Processing 9, pp 242-243.
[21] Duka, D. (2013). Adoption of agile methodology in software development. Proceedings International Convention on Information & Communication Technology Electronics & Microelectronics, pp. 426-430. Opatija, Croatia.
[22] Falessi, D. et al. (2010). Peaceful Coexistence: Agile Developer Perspectives on Software Architecture. IEEE Software 27(2), pp. 23-25.
[23] Booch, G. (2009). The Defenestration of Superfluous Architectural Accoutrements. IEEE Software 26(4), pp. 7-8.
[24] Kruchten, P. (2009). Software Architecture and Agile Software Development - An Oxymoron? Workshop Keynote software architecture challenges in the 21st century. Los Angeles.
[25] Díaz, J. et al. (2013). Change-Impact driven Agile Architecting. Proceedings 46th Hawaii International Conference on System Sciences, pp 4780-4789.
[26] Golden, B. (2015). What Gartner's Bimodal IT Model Means to Enterprise CIOs. CIO. Online [Feb. 2015].

Caracterización y calibración de un sensor de humedad de bajo costo para optimizar su desempeño dentro de una red WSN en escenarios *indoor*

Luis Noreña T.
Diego Cortés
Fundación Universitaria de San Gil – Colombia

1 Introducción

Medir con precisión la humedad del suelo en la agricultura del arroz de secano en el Casanare es importante para lograr eficiencia porcentual en la distribución y uso del agua en los sistemas de riego que lo nutren. Tradicionalmente, la frecuencia de riego se ha basado por estimación empírica, donde el personal técnico es quien define el momento y cantidad de agua a utilizar, lo que conlleva excesos o posible déficit de agua. Por otro lado, la aplicación de manera correcta y oportuna de un volumen adecuado conlleva a producciones uniformes de arroz durante el desarrollo fenológico del cultivo, y a bajos costos de operación de los equipos de riego. Las medidas se evalúan con un sensor comercialmente validado y a través de mediciones gravimétricas directas. De esta forma se determina la desviación que puede tener estos con respecto a los que arroja el dispositivo de bajo costo, y de esta manera identificar las limitaciones que se deben tener en cuenta para su uso. Además, se evaluará su costo incluyendo los demás elementos requeridos para un correcto funcionamiento, tales como el sistema de almacenamiento de datos y transmisión y el sistema de alimentación eléctrica. Esta situación problemática plantea la necesidad de pensar en otra metodología de medición de la humedad volumétrica, y disponer de sensores de humedad distribuidos de acuerdo con la variabilidad espacial de los suelos de cada terreno, de tal manera que la información censada pueda ser recogida y enviada a un centro de control, para su posterior análisis y toma de decisiones. Con este objetivo, el sistema que se desarrolla en este proyecto incluye, además de hacer uso eficiente de sensores de humedad volumétrica de bajo costo (tipo HL69), la calibración para que esa información sea confiable y se pueda enviar haciendo uso de la tecnología inalámbrica Zigbee.

El sensor de bajo costo se calibra tomando un sensor de altas prestaciones como referencia, método que ha sido abordado en diferentes investigaciones [9, 10]. También se hará uso de muestras de suelo de diferente textura y con humedad volumétrica conocida previamente. Finalmente, para evaluar la calidad de servicio de la transmisión inalámbrica de los datos será necesario tomar medidas distribuidas en zonas extensas de cultivo, y enviar esos datos a un centro de control situado a varios km de distancia. Los resultados obtenidos se utilizarán para la futura implementación de un mapa de riego, haciendo uso de herramientas informáticas tal como se plantea en la agricultura de precisión [4].

2 Contexto de investigación

La primera fase se orienta a evaluar el desempeño del sensor de humedad volumétrica de bajo costo, para su posterior acondicionamiento e integración dentro de un sistema de transmisión de datos inalámbricos. La segunda fase será la implementación de una red de sensores inalámbricos distribuidos en configuración malla, para el

monitoreo del riego. El despliegue de esta red se efectuará inicialmente en un ambiente de laboratorio, con el fin de analizar su comportamiento y poder efectuar los ajustes pertinentes al momento de configurar las técnicas de enrutamiento de cada estación de datos. Estas técnicas permitirán lograr mayor cobertura de la estructura y evitar desvanecimiento de enlaces o pérdida de nodos en la red. Con lo anterior se pretende lograr un aporte en el campo de la investigación y aplicación de las *WSN* en configuración de red malla [2].

El montaje de esta red se realizará con dispositivos indicados para un futuro despliegue en el cultivo, de tal manera que permitan la reconfiguración de los parámetros básicos y faciliten la adquisición y transmisión de las variables ambientales deseadas en puntos estratégicos del terreno. Además, se contará con una interfaz gráfica de visualización ubicada en el nodo principal, que permitirá conocer y/o almacenar las variables de interés de la red y realizará labores de administración.

En la calibración del sensor se evaluará la humedad volumétrica de diferentes tipos de suelos, cuyas propiedades físicas sean conocidas previamente, y se variará el contenido de agua a cada una de ellas que tendrá que ser detectado por el sensor [3]. Estas medidas variarán por efectos de la temperatura del suelo y los contenidos de sales que tengan, por lo que inicialmente se trabajará en suelos no salinos y las medidas de humedad se tomarán en horas fijas para contrarrestar imprecisiones en la medida por variación de la temperatura.

El valor de la constante dieléctrica del agua posibilita la estimación de su concentración en el suelo mediante la medida de la capacidad eléctrica entre dos electrodos enterrados [2]. Sin embargo, la salinidad presente en muchos tipos de suelo introduce una conductancia en paralelo que hace menos precisa la medida de la humedad en sistemas de baja frecuencia, basados exclusivamente en la estimación del módulo de la impedancia presentada. Para solucionarlo se trabaja con frecuencias mayores a 70 *MHz*. Con el sistema que aquí se describe se determinará el grado de precisión de la sonda de bajo costo en el registro de la humedad del suelo, calibrándola con la humedad gravimétrica y una sonda comercialmente validada compatible con la plataforma Arduino. Que a su vez hará parte de un sistema de trasmisión de datos eficiente que se podrá utilizar como información para generar un mapa de riego en tiempo real.

3 Metodología

3.1 Caracterización del sensor

Durante el período de pruebas preliminares no se contaba con valores de referencia. Para el análisis en estas instancias uno de los principales problemas era el no conocer el nivel inicial de humedad. Para lograr tener un valor de referencia y como parte de los objetivos del proyecto, se adquirió un sensor comercial modelo EC-5 Small y un sensor de humedad de bajo costo HL-69 que debe ser calibrado. En la Tabla 1 se detallan los parámetros del sensor. El sensor EC-5 y de acuerdo con las especificaciones del fabricante, permite obtener una precisión de alrededor del 5% con calibraciones genéricas en cualquier tipo de suelo, y del 2% con calibraciones específicas. La conexión a la sonda EC-5 se realiza mediante un *plug* de 3.5mm.

3.2 Calibración del sensor

Las Sondas de Humedad del Suelo son sensores capacitivos del tipo FDR (Frequency Domain Reflectrometry), que miden la constante dieléctrica o permitividad del suelo para

calcular su contenido de humedad. La fracción volumétrica del suelo ocupada por agua tiene una amplia influencia en la permitividad dieléctrica del suelo, porque su valor dieléctrico (80) es mucho mayor que el de los otros constituyentes del mismo. A partir de estos valores las sondas de humedad miden la variación de la capacitancia del suelo y a partir de ella el valor de la constante dieléctrica. La capacidad del suelo altera la resonancia de frecuencia del circuito, relacionándola con el cambio de contenido del agua (ver Tabla 1). Estos datos son alterados por el contenido de sales y la densidad aparente del suelo. La medida de las sales se soluciona trabajando con frecuencias mayores y la densidad aparente, previamente conocida, será la variable común durante la calibración.

Tabla 1: Parámetros básicos del sensor

accuracy	Mineral soil	±3% VWC, most mineral soils, up to 8 dS/m ±1-2% VWC with soil specific calibration
accuracy	RoockWool	±3% VWC, 0.5 to 8 dS/m
accuracy	Potting Soil	±3% VWC, 3 to 14 dS/m
resolution	0.1% VWC (mineral soil)	0.25% VWC (rockwool)
Range	calibration dependant	up to 0-100% VWC with polynomial equation
dimensions	3 dimensions	8.9 x 1.8 x 0.7 cm
Cable leghnt	Sensors come standard with 5 m cable	Custom cable lengths available. Maximum cable length of 40 m.
Measurement time	Time	10 ms
Power	2.5 - 3.6 V DC @ 10 mA	Output proportional to input voltage. 2.5 V and 3 V excitations supported with calibration equations
Output.		1 erg/(G·cm^3) = 1 emu/cm^3 $\rightarrow 4\pi \times 10^{-4}$ T
Temperature	Celsius	-40°C to +50°C

Para este ensayo se tendrán tres depósitos de acrílico con las siguientes medidas: 1mx1mx0.6m (Figura 1), en sus paredes tendrán ranuras para introducir los sensores, en su interior se ubicarán suelos en lo posible de texturas arcillosa, arenosa y limosa. Estos pedones serán tomados en campo sin alterar su estructura y se les determinarán las siguientes propiedades físicas:

a. Capacidad de campo
b. Punto de marchitez permanente
c. Densidad aparente
d. Conductividad eléctrica
e. Textura

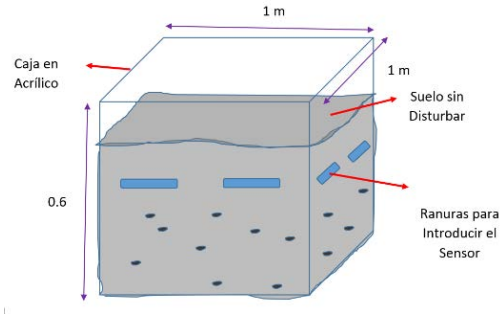


Figura 1: Recipiente con suelo (Autores)

Las propiedades a, b y c serán precisadas en campo y d y e en laboratorio. Posteriormente y antes de iniciar la toma de datos en los recipientes con suelo, se

verificará las lecturas que tengan tanto el sensor comercial como el de bajo costo; para esto se emplea líquido anticongelante con una dieléctrica similar a la de los suelos, y a partir de aquí se observará la variabilidad de las lecturas del sensor de bajo costo.

3.3 Acondicionamiento de la señal del sensor

Las sondas y los instrumentos FDR utilizan un oscilador de corriente alterna para formar un circuito eléctrico en conjunto con el suelo. Después de insertar las sondas, que pueden ser puntas paralelas o anillos de metal en el suelo, el oscilador produce valores de frecuencia de acuerdo al contenido de humedad del suelo [3]. Para traducir esos valores de frecuencia a valores de voltaje que pueda leer el microcontrolador, se hace necesario acondicionar la señal, tal como se describe en la Figura 2.

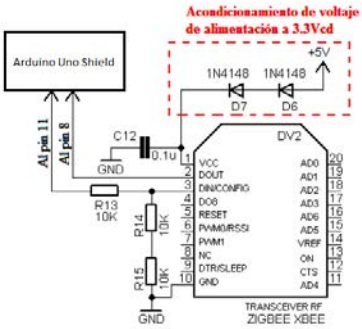


Figura 2: Acondicionamiento del sensor y conexión con el Transceiver (Autores)

Estas muestras de suelo serán depositadas en dichos recipientes, que tendrán aberturas de 1 cm en su parte inferior y que servirán como drenaje para el exceso de agua. Posteriormente se les dejará secar hasta que su humedad sea constante y coincida con su punto de marchitez permanente, una vez se logre esto se le agregara un Volumen de agua conocido y se determinará su humedad volumétrica de acuerdo con la siguiente relación:

$$Hv = Hg * da \qquad (1)$$

Donde Hv = Humedad Volumétrica y da = Densidad Aparente. Este seguimiento se realizará de dos maneras: con la sonda Decagon EC-5 y la sonda de bajo costo, y medidas directas a través del método de estufa, que serán las que defina el patrón de humedecimiento y la variabilidad en las lecturas. Estos datos deberán ser similares a los registrados por el sensor. Si no es así, se ajustarán las lecturas de este con los valores tomados en laboratorio y se creará un modelo matemático de la relación de ajuste, que permita acercar los datos observados a lo real. Este procedimiento se realiza para los tres depósitos y se grafica de acuerdo con la Figura 3.

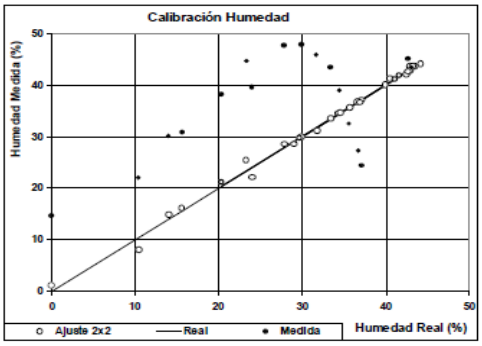


Figura 3: Corrección de datos Humedad [2]

3.4 Configuración de la red

Para la transmisión de datos se utiliza tecnología inalámbrica bajo el protocolo ZigBee. Se seleccionó porque se adapta mejor a los requisitos de las WSNs, y porque el estándar fue concebido para ser utilizado en aplicaciones con un perfil similar al de las redes de sensores [1]. En ZigBee, la tasa de datos máxima es casi un tercio de la tasa de datos de Bluetooth, sin embargo, permite la creación de redes más flexibles con topologías tipo estrella, árbol y malla, en las que se pueden conectar teóricamente miles de nodos. Además, el consumo de los módulos ZigBee es más bajo que el de los usados en otras tecnologías, tales como *bluetooth*, que en este caso se trata de corrientes en niveles microamperios.

La selección del módulo a utilizar en este proyecto depende, en gran medida, de las necesidades de la aplicación a desarrollar, en este caso soluciones en el sector agrícola [7]. Después de realizar una completa revisión sobre los módulos de transmisión disponibles en el mercado, se determinó que el más adecuado para el desarrollo del trabajo era el XBee Pro serie 2, porque tiene la capacidad de gestionar una red en configuración malla y cumple con las características hardware necesarias para la integración de todos los elementos que componen la aplicación: sensores analógicos, fuentes de poder, antenas y módulos Arduino.

Zigbee utiliza Direct Sequence Spread Spectrum DSSS para modular la información, antes de enviarla a la capa física, lo que le brinda mayor inmunidad al ruido [1] (Figura 4). Básicamente, cada bit de información a transmitir es modulado en 4 señales diferentes, este proceso hace que el total de la información a transmitir ocupe un ancho de banda más grande, pero que utilice una densidad de potencia espectral inferior para cada señal. Esto provoca menos interferencia en las bandas de frecuencia y mejora la relación señal-ruido (SNR) en el receptor, debido a que el mensaje que está siendo enviado por el transmisor es más fácil de detectar y decodificar.

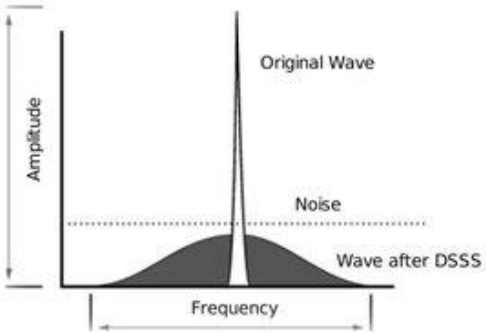


Figura 4: Uso del Direct Sequence Spread Spectrum DSSS

Existen diferentes modulaciones DSSS en función de los límites físicos del hardware del circuito y el número de símbolos que se puede procesar en un momento dado: desplazamiento de fase binaria (BPSK), Offset de cuadratura Phase Shift Keying (O-QPSK) y Secuencia paralela Spread Spectrum (PSSS), logrando comunicación con anchos de banda desde 20Kb a 250Kb/s, que permiten enviar información con facilidad desde el sensor y con una buena calidad de servicio, una característica distintiva de las NGN [10].

Zigbee es inherentemente una red malla. Los *routers* y el coordinador *descubren* la ruta hacia el destinatario del mensaje, y si no hay comunicación directa lo entrega al *router* más cercano, que a su vez repetirá esta tarea hasta que el mensaje llegue al destinatario

final [5]. Los *end-devices* siempre entregan los mensajes a su coordinador (*router* o coordinador de la red Zigbee), que almacena los mensajes hasta que éstos despierten y lo contacten, momento en el cual se los entrega. Todo lo relativo al sincronismo y bajo consumo queda manejado por la red Zigbee y solamente se debe tener en cuenta que la aplicación soporte las demoras introducidas por el tiempo de tránsito dentro de la red y el tiempo que un *end-device* pasa en bajo consumo (con el receptor apagado). En otras iniciativas se ha optado por otro tipo de redes inalámbricas [8]. En la Figura 5 se describe el caso más simple, en el que una serie de aplicaciones remotas reportan información a un punto central, sin que haya comunicación en el sentido inverso, y este punto central coincide con el coordinador de la red Zigbee.

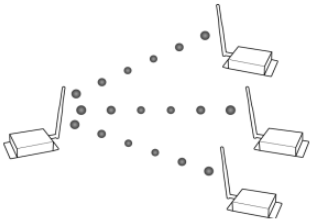


Figura 5: Comunicación en un sentido, hacia el coordinador, en el centro de red [6]

4 Resultados

Se espera que los coeficientes de correlación sean mayores a r > 0.8, de esta manera se infiere que existe una correlación alta entre lo medido y lo que hay en terreno. En caso en que no corresponda se busca la causa de la diferencia, pero en lo posible se tratará de buscar una relación entre las medidas del sensor de bajo costo y el Decagon EC-5. Además, se espera que los costos asociados sean inferiores a los sensores comercialmente validados, y que la transmisión de datos sea confiable y posible.

En el sistema conectado al módulo central se recibirán los mensajes que envíen los remotos, y para determinar quién lo envía hay que insertar algún tipo de identificador dentro del mismo, o utilizar el modo API, aunque este tema se dejará para otra investigación. A fin de mantener la integridad de un mensaje, todos los bytes que lo conforman deberán ser enviados uno a continuación del otro, sin una demora mayor a lo que indica el comando ATRO y con una longitud menor a 100 bytes.

Otro experimento es la verificación de la comunicación entre los dos módulos Xbee en condiciones *indoor*. La Figura 6 es un ejemplo de cómo, por medio del software X-CTU, se puede visualizar en tiempo real el comportamiento de los módulos al separar el del receptor. Se pueden tomar medidas de recepción de paquetes y el nivel RSSI del paquete a diferentes distancias, para este caso a 30 m, 60 m y 100 m de distancia. Para la prueba se utiliza dos dispositivos Xbee conectadas a dos computadores portátiles.

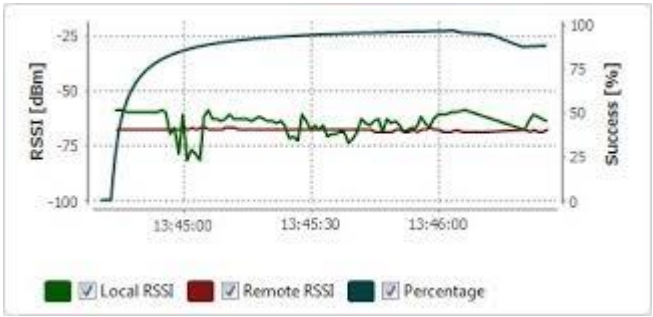


Figura 6: RSSI (Autores)

Otros resultados experimentales se basan en los paquetes perdidos con valores entre 0 y 1.000 y la medida del factor LQI dentro de los valores típicos entre -95dBm y -18dBm, según el protocolo IEEE 802.15.4. A continuación se presentan los resultados de tres envíos de secuencias de 32 bits con 5, 10 y 20 bytes de tamaño de carga útil por paquete, y una aplicación de Zigbee con monitoreo de red de variables ambientales. En la Figura 7 se muestran las medidas LQI y de paquetes perdidos, respectivamente, para 5 bytes de carga por paquete.

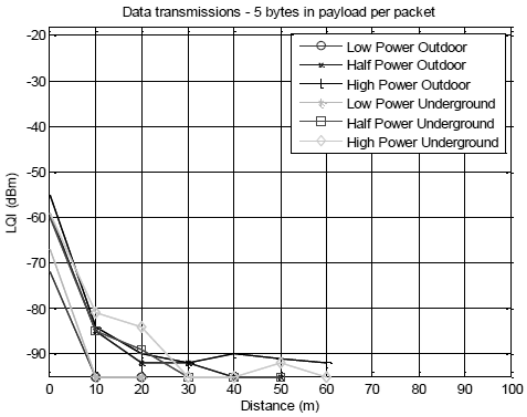


Figura 7: Comportamiento de la toma de datos bajo el criterio del LQI dentro de los valores típicos -95dBm (Autores)

5 Conclusiones

Se espera que el sensor de bajo costo sea una alternativa técnicamente confiable, de fácil uso e interpretación que les permita a los agricultores adquirir tecnología de punta en la gestión del riego, y esto contribuya a disminuir los gastos de funcionamiento durante el desarrollo del cultivo. Además, la transmisión de los datos a un módulo central permitirá un mayor criterio para la toma de decisiones.

El criterio de calidad de servicio LQI se puede utilizar en una red ZigBee como mecanismo de ruteo en una red malla. Así se elegirían las rutas con LQI más alto, lo que se traduce en agilidad de convergencia de la información en tiempo real. Pero existen otros factores a tener en cuenta en el ruteo, tal como el consumo de energía de las baterías. Los nodos que intervengan con mayor frecuencia en el paso de mensajes agotarán sus baterías antes, y este es un indicador de la calidad de los paquetes recibidos por el receptor. Puede usarse la intensidad de señal de recepción o la relación señal ruido. Cuanta más alta sea la última se considera que habrá más garantía de que el mensaje llegue a destino. El análisis previo del comportamiento, tanto del sensor, del formateo de sus datos y de su integración dentro de una red permite obtener información en tiempo real de datos fiables y sin la intervención de agentes externos. Esta experiencia abre nuevas oportunidades en la aplicación de las redes de sensores inalámbricos en áreas relacionadas con el sector agrícola.

References

[1] Adams, T. (2005). An introduction to IEEE 802.15.4. Proceedings IEEE Aerospace conference. Big Sky, USA.

[2] Buelvas, D. & Hincapie, R. (2011). Implementación de redes inalámbricas de sensores WSN (Wireless Sensor Networks) para el monitoreo estructural. Universidad Pontificia Bolivariana.

[3] Martin, C. (2010.). Métodos para Medir la humedad del suelo para la programación del riego ¿Cuándo? The University of Arizona.

[4] Keshtgari, M. & Deljoo, A. (2012). A Wireless sensor network solution for precision agriculture based on ZigBee technology. Journal Wireless Sensor Network 4(1), pp. 25-30.
[5] Iacono, L. et al. (2010). Estudio de plataformas de hardware empleadas en redes de sensores inalámbricas. Memorias XVI Congreso Argentino de Ciencias de la Computación, pp. 784-793. Buenos Aires, Argentina.
[6] Caprile. R. (2008). Utilización de XBee ZB para comunicación de aplicaciones. Reporte Técnico CTC-059. Cika Electrónica.
[7] Caprile. R. (2008). Utilización de módulos XBee 802.15.4 con coordinador. Reporte Técnico CTC-057. Cika Electrónica.
[8] Shaikh, Z. et al. (2013). Crop irrigation control using Wireless sensor and actuator network (WSAN). Proceedings International Conference on Information and Emerging Technologies, pp. 1-5. Karachi, Pakistan.
[9] Szilvia, B. & Kovács, G. (2009). NGN access network and its rural concerns. Proceedings Joint International Conference, pp. 1-6. Prague, Czech Republic.
[10] Wang, X.; Ma, L. & Yang, H. (2011). Online water monitoring system based on ZigBee and GPRS. Procedia Engineering 15, pp. 2680-2684.
[11] Zhi, C. & Lu, C. (2005). Humidity sensors - A review of materials and mechanisms. Sensor Letters 3, pp. 274-295.

Interfaz para navegación autónoma de discapacitados visuales en salas de exposiciones museográficas

Mauricio Arias C.
Anderson Arango
Stiven Agudelo
Dany Urrego
Jakeline Serrano
Instituto Tecnológico Metropolitano – Colombia

1 Introducción

El Consejo Internacional de Museos adjudica a los museos la responsabilidad de conservar y cuidar los patrimonios naturales y culturales, así como investigar y divulgar los acervos que custodia [1]. El museo se concibe entonces como un espacio de comunicación [2, 3] y es a través de sus exposiciones donde se da el encuentro entre el objeto y el visitante [4]. Por tanto, debe llevar la información a una dimensión tangible, que pueda interpretarse y que sea participativa e incluyente de todo público [5, 6]. El Museo de Ciencias Naturales de La Salle, un proyecto cultural del Instituto Tecnológico Metropolitano ITM, de Medellín Colombia, se ha convertido en un eje de producción y apropiación del conocimiento por medio de la exposición de sus colecciones. Sin embargo, ha identificado la necesidad de adaptar su espacio para que sea realmente incluyente con los discapacitados visuales, de manera que puedan acceder a las colecciones con mayor autonomía y disfrutarlas, deleitarse, conocer y aprender acerca de los especímenes expuestos. Debido a la inversión que esto conlleva, son pocos los museos que a nivel mundial han puesto en práctica planes de accesibilidad e inclusión social como este [7]. Se han desarrollado diversos sistemas de navegación para usuarios con discapacidad visual, tanto para interiores como para exteriores, y las técnicas utilizadas así como las interfaces desarrolladas se describen en la siguiente sección. Satisfacer esta necesidad identificada requiere soluciones simples y replicables, que no afecten el normal desempeño del usuario.

Este trabajo presenta el diseño y la evaluación de una interfaz que facilita la navegación autónoma de personas con discapacidad visual en espacios interiores, en particular en salas de exposiciones museográficas. La interfaz está constituida por señales podotáctiles, un dispositivo ponible y un audífono de conducción ósea. Las señales podotáctiles le indican a los usuarios la trayectoria a seguir y los puntos de parada, simplificando así el problema de guiado. El dispositivo ponible es un cinturón con sensores y actuadores para evitar obstáculos, complementado con una cámara para detectar los códigos de identificación de los especímenes ubicados en los puntos de parada. La información correspondiente a cada espécimen identificado se transmite por medio de un audífono de conducción ósea. La interfaz desarrollada no aísla al usuario del entorno ni le impone una alta carga cognitiva.

En la siguiente sección se resume el estado del arte en técnicas e interfaces que permiten la navegación de personas sin conocimiento del entorno en el cual se desplazan. En la sección 3 se presenta el proceso de diseño de la interfaz bajo una metodología de diseño de producto de ingeniería. La evaluación de la interfaz y sus resultados se presentan en la sección 4 y los resultados y trabajos futuros se discuten en la sección 5.

2 Trabajos relacionados

Algunos dispositivos previamente desarrollados permiten detectar y evitar obstáculos, otros guían al usuario hasta destinos específicos proporcionándoles posiciones y orientaciones durante el recorrido. Todo esto implica usar diversas técnicas, tales como estimación de posición, uso de balizas y uso de sensores [8]. La técnica de estimación de posición (*dead-reckoning*) permite estimar la posición actual del usuario con base en una localización previamente estimada o conocida. Esto se logra interpretando señales procedentes de sensores que porta el usuario, tales como podómetros, brújulas, acelerómetros, magnetómetros y giróscopos. Esta técnica es útil para la navegación en interiores. Nakamura, Aono y Tadokoro [9] presentan aplicaciones para personas con discapacidad visual y Ojeda y Borenstein [10] hacen lo mismo con aplicaciones depuradas para personas con visión normal. A pesar de la sencillez de su implementación, estas técnicas presentan como desventaja la acumulación de errores en posicionamiento a medida que el usuario se desplaza, aunque en [10] se presentan errores de menos del 2% de las distancias recorridas.

El uso de balizas requiere la intervención del espacio de navegación con identificadores, tales como códigos QR detectados por una cámara [11], emisores infrarrojos [12], o de emisores de ultrasonido o de luz visible [13, 14]. Un entorno con estos identificadores se puede complementar con etiquetas RFID [15, 11] y nodos de redes inalámbricas dedicadas a la localización de los usuarios. Algunos de estos dispositivos pueden generar errores debido a la interferencia de señales y otros requieren estar en línea de vista para desempeñarse de manera efectiva, pero necesitan una intervención del espacio que demanda mayores tiempos y costos. También se reporta el uso de sensores RGB-D, tales como el Kinect de Microsoft, para apoyar la navegación de personas ciegas por medio de la detección de obstáculos en tiempo real [16], aunque la baja portabilidad del Kinect lo hace incómodo para el usuario. Adicionalmente, se presentan dificultades en la obtención de información de superficies brillantes.

En cuanto a las interfaces para seguir rutas, se pueden clasificar como acústicas, táctiles y combinaciones de ellas [17]. NavBelt [18] es una interfaz acústica que utiliza un dispositivo ponible tipo cinturón con sensores de ultrasonido y un computador portátil, para detectar y evitar obstáculos. Las instrucciones para evadir obstáculos se entregan como comandos de voz por medio de auriculares o enviando imágenes panorámicas acústicas por medio de sonido estéreo. También se envían señales de distancia y dirección por medio de sonidos binaurales para guiar a las personas [19]. Las desventajas comunes de los dispositivos acústicos tienen que ver con una alta carga cognitiva y el aislamiento del sonido ambiente al usar auriculares, lo cual a su vez genera inseguridad en los discapacitados visuales. Los dispositivos ponibles vibrotáctiles para navegación están constituidos por matrices de vibradores, ubicados dentro de una prenda de vestir que se pone sobre la espalda [20, 21], para describir rutas por medio de direcciones y paradas. También se encuentran como cinturones vibrotáctiles [22, 17], que permiten entregarle al usuario información direccional múltiple, por medio del control de los intervalos de vibración y la cantidad de vibradores usados. Debido a que los usuarios reciben instrucciones direccionales hápticas, no hay impacto negativo en su capacidad de escuchar y percibir el entorno.

3 Metodología

Se parte de un proceso de diseño para el que se adaptó una metodología de diseño de producto [23]. Como se muestra en la Figura 1, se parte de la comprensión clara del

problema a partir de una revisión de la literatura relacionada y la identificación de las necesidades del usuario. Con esta información se realizó el diseño conceptual de un dispositivo que satisficiera la necesidad del usuario y posteriormente se procedió a diseñar el dispositivo en detalle (cada uno de sus sub-sistemas). El dispositivo diseñado se desarrolló como prototipo y se adaptó al cuerpo de personas con discapacidad visual, como un dispositivo ponible, para evaluar su desempeño por medio de datos cuantitativos y cualitativos obtenidos a través de experimentos.

	FASE 1: Desarrollo del concepto	FASE 2: Diseño de sistema	FASE 3: Diseño de detalle	FASE 4: Pruebas y refinamiento
Diseño	Identificación de necesidades del usuario Desarrollo de conceptos de diseño industrial	Definición de subsistemas e interfaces principales Refinación del diseño industrial	Definición de la geometría de las piezas Selección de materiales	Prueba de desempeño Implementación de cambios de diseño
Manufactura	Evaluación factibilidad de producción	Definición del esquema final de ensamble	Definición de proceso de producción de piezas	Refinación de procesos de fabricación y ensamble

Figura 1: Adaptación de la metodología de diseño a partir del proceso genérico para desarrollo de productos impulsados por tecnología [23]

3.1 Desarrollo del concepto de diseño

Identificación de necesidades del usuario

Partiendo de la necesidad de brindarles autonomía a las personas con discapacidad visual para recorrer las exposiciones y acceder al conocimiento inherente a los especímenes, se procedió a recopilar y analizar información (revisión de la literatura) acerca de dispositivos similares existentes. El análisis se complementó con entrevistas a dos discapacitados visuales –ambos hombres adultos con ceguera total– y con experiencia en desplazamientos con bastón en interiores y exteriores. Las respuestas permitieron establecer que para los ciegos (en ciudades como Medellín) es común reconocer las señales podotáctiles y desplazarse sobre ellas en exteriores para llegar a sitios de interés, apoyándose siempre en el uso del bastón. Por otro lado, en interiores siempre usan bastón para desplazarse, pero no tienen apoyos adicionales que les permita desplazarse con seguridad y confianza. En opinión de los entrevistados, las características ideales de una interfaz de navegación útil deben ser: pequeña, liviana, segura, con baja complejidad tecnológica (en relación con el uso), que no *robotice* (que no afecte la naturalidad de los movimientos), que no bloquee el sonido ambiente ni genere muchas señales al mismo tiempo (para evitar confusiones) y que se pueda aprender a usar rápidamente. Toda la información recopilada se tabuló en una matriz de necesidades-métricas para orientar el diseño (Figura 2).

Desarrollo de concepto de diseño industrial

Se descompuso el problema en sub-problemas o sub-funciones que cada elemento debe cumplir para ejecutar la función general de la interfaz (Figura 3). La descomposición del problema facilitó el análisis y el equipo de diseño generó conceptos de solución, proponiendo dispositivos físicos que pudieran satisfacer la necesidad de cada sub-función. Los dispositivos fueron evaluados por medio de matrices de selección, con criterios que depuraron la solución óptima.

	Necesidad	Métrica (Unidad)	1 Desviación con respecto a la ruta [sí, no]	2 Llegada a punto de interés [sí, no]	3 Exceso de señales [sí, no]	4 Bloque de sonido ambiente [sí, no]	5 Detección de obstáculo [sí, no]	6 Peso del dispositivo [kg]	7 Volumen del dispositivo [m³]	8 Tiempo de entrenamiento [h]
1	Desplazamiento de manera autónoma sobre una ruta predefinida.		■							
2	Llegar a los puntos de interés de la ruta		■	■						
3	Baja carga cognitiva asociada al funcionamiento del dispositivo				■		■			
4	No debe existir interferencia del sonido ambiente					■				
5	Se debe brindar percepción de seguridad por medio de la detección de obstáculos				■		■			
6	Bajo peso del dispositivo ponible							■		
7	El dispositivo ponible no debe ser voluminoso							■	■	
8	El tiempo de entrenamiento debe ser mínimo.				■					■

Figura 2: Matriz de necesidades-métricas, obtenida a partir de la identificación de necesidades del usuario

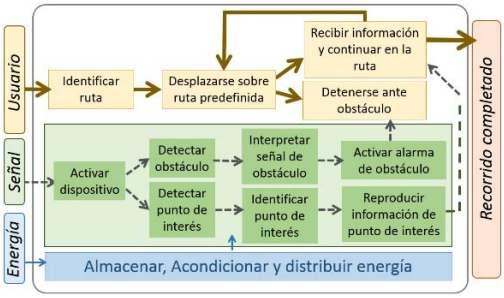


Figura 3: Diagrama funcional para el diseño

En la Figura 4 se presenta la solución conceptual –genérica– y en la Tabla 1 el listado de elementos seleccionados para integrar el sistema (interfaz).

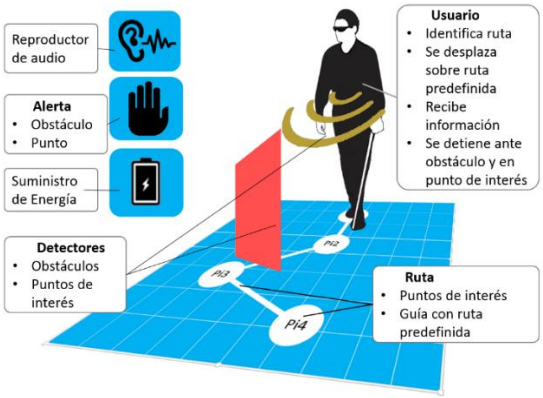


Figura 4: Concepto solución a partir del diagrama funcional

Tabla 1: Resultados de matrices de selección

Sub-problema	Solución	Elemento
Identificar ruta.	Posicionar usuario en inicio de ruta.	Usuario
Desplazarse sobre ruta predefinida.	Agregar señales podotáctiles al piso del espacio museográfico. Agregar marcas sobre puntos de interés para identificar cada uno.	Guías podotáctiles. Códigos QR
Recibir información, continuar en ruta.	Agregar audífonos que entreguen información, sin bloquear sonido ambiente.	Audífonos de conducción ósea.
Detenerse ante obstáculo.	Enviar señal al usuario que indique obstáculo y que debe detenerse. Con baja carga cognitiva.	Usar motor de vibración.
Activar alarma de obstáculo.	Enviar señal desde microcontrolador a efector final.	µControlador
Interpretar señal de obstáculo.	Recibir señal procedente de transductor en microcontrolador.	µControlador
Detectar obstáculo.	Usar transductor, para detectar objetos distancia aprox. 60 cm.	Sensor de presencia.
Detectar punto de interés.	Usar cámara para identificar marcas en puntos de interés.	Cámara
Identificar punto de interés.	La información adquirida desde cada marca, debe ser procesada para identificar su correspondencia con un espécimen.	Procesador
Reproducir información de punto de interés.	Se debe reproducir la información de audio correspondiente a cada punto de interés.	Reproductor de audio.
Almacenar, acondicionar y distribuir energía.	La energía necesaria para completar un recorrido, se almacena, acondiciona y distribuye a dispositivos electrónicos.	Batería Reguladores de voltaje
Activar dispositivo	Se usa un interruptor On/Off	Interruptor On/Off

3.2 Diseño del sistema

Sistema podotáctil

Además de usar un bastón, las personas con discapacidad visual utilizan señales podotáctiles para desplazarse en exteriores: guiadas o advertencia. Las baldosas

podotáctiles de guiado tienen líneas rectas en relieve, en la dirección de desplazamiento de las personas, que se instalan en el piso de forma paralela a paredes y aceras. Se complementan con las baldosas de advertencia, que son promontorios en forma de taches redondeados y que le indican al usuario cuando detenerse (en este trabajo se les llama *marcas* podotáctiles). Si el usuario camina sobre estas señales recibirá información táctil por medio de sus pies. El subsistema podotáctil, que hace parte de la solución propuesta, está constituido por guías para el desplazamiento, por marcas de advertencia para detenerse en puntos de interés y por códigos QR que identifican cada punto de parada.

Dispositivo ponible

Es un dispositivo que se adapta al usuario como un cinturón. Está constituido por un sensor de presencia que envía una señal a un microcontrolador cuando se presentan obstáculos sobre las guías en el recorrido; un actuador que genera vibraciones sobre el abdomen para alertar sobre el obstáculo; una cámara para identificar los códigos QR que están sobre las marcas indicando puntos de interés; una tarjeta MP3 para reproducir la información acerca de los especímenes que están en los puntos de interés; y un microcontrolador que recibe señales desde los sensores y envía otras a los actuadores.

Sub-sistema acústico

Constituido por audífonos de conducción ósea que se ubican cerca del canal auditivo, no sobre él. Su elección tuvo que ver con aspectos tales como no bloquear los canales auditivos, no generar confusiones y generar seguridad en los usuarios.

Sub-sistema de procesamiento

Se agregó un sub-sistema de procesamiento de imágenes basado en *mini-laptop* con sistema operativo Windows, para minimizar el tiempo de detección e identificación de los códigos QR. La señal de salida se envía al microcontrolador del dispositivo ponible para reproducir los audios. Incluir el laptop redujo tiempos de procesamiento, solucionó el problema de suministro de potencia y facilitó la interconexión de dispositivos, pero incrementó el tamaño y el peso de la interfaz, por lo cual se adaptó a los usuarios por medio de un morral en su espalda mientras se realizaba la evaluación. En la Figura 5 se muestran los sub-sistemas que componen la interfaz.

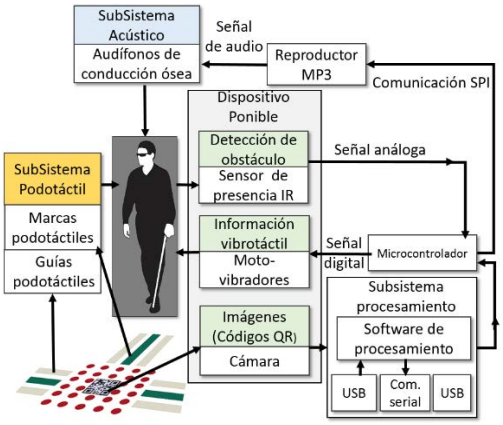


Figura 5: Arquitectura del sistema

Se utilizó software CAD para circuitos electrónicos para seleccionar los dispositivos electrónicos y realizar el diseño del detalle del dispositivo ponible. El diseño del software por medio de un IDE para programación de microcontroladores (Arduino) permitió programar las tarjetas seleccionadas. Se realizaron diferentes pruebas hasta optimizar el

desempeño del dispositivo ponible, previa experimentación con discapacitados visuales en un entorno museográfico. Las pruebas de la totalidad del sistema se presentan en la siguiente sección.

4 Evaluación de la interfaz

Cuatro personas participaron en la evaluación del sistema (dos hombres y dos mujeres), con edad promedio de 22 años y que se denominarán A, S, J y M. Los hombres son mayores de edad con estudios profesionales y las mujeres son menores de edad con nivel de formación de educación básica secundaria, y todos con ceguera total y con el conocimiento necesario para desplazarse en entornos exteriores utilizando un bastón. En el periodo de experimentación ninguno tenía discapacidades adicionales y todos presentaban desempeño perfecto de sus demás sentidos. Sin embargo, los dos hombres tenían mayor experiencia y autonomía para desplazarse en entornos exteriores e interiores, básicamente por experiencia previa en el desplazamiento con bastón y a la diversidad de actividades que realizan en exteriores. La autonomía de desplazamiento en las mujeres jóvenes era menor. En adelante, a todos se les denomina *participantes*.

Las señales podotáctiles se instalaron en el piso de la Sala 2 del Museo de Ciencias Naturales de La Salle, donde se encuentran exhibidos los especímenes de los páramos colombianos (exposición: *Paraísos perdidos de Colombia- páramos*). El conjunto total de guías y marcas podotáctiles constituyen un recorrido completo. Las guías constituyen tramos que están delimitados por marcas podotáctiles, como se muestra en la Figura 6.

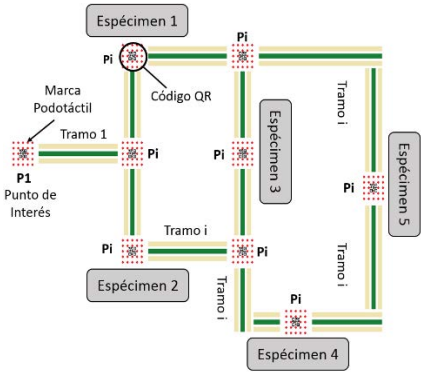


Figura 6: Distribución de guías y marcas podotáctiles

Se adaptó la interfaz como un cinturón a cada participante (Figura 7), garantizando que la distancia de la cámara al piso no superara los 1.40m y que su inclinación con respecto a la horizontal fuera de -60° aproximadamente.

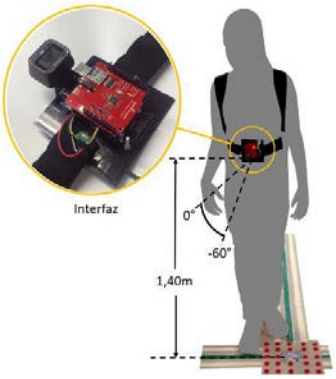


Figura 7: Adaptación de interfaz como dispositivo ponible

Los participantes fueron instruidos brevemente con respecto al uso del dispositivo y de las guías, así como de las características del espacio. La evaluación consistió en realizar tres recorridos completos, diez tramos por recorrido (sin repetir tramos) en direcciones diferentes y con puntos de inicio diferentes, para evitar que memorizaran el recorrido. Un tramo estaba acotado entre un punto de interés inicial (*Pin*) y un punto de interés final (*Pfn*). Se adquirieron datos para 120 tramos en total (30 tramos por participante). Para cada recorrido se ubicaban inicialmente en un punto de interés aleatorio denominado *Pin*, y a partir de ese punto debían desplazarse –sobre las guías podotáctiles que constituían un tramo– hasta el punto de interés siguiente *Pfn*. Al llegar a él el participante debía detenerse y esperar a que se reprodujera un audio por medio de los auriculares de conducción ósea, con la información correspondiente a las características físicas, comportamiento y sonidos del espécimen dispuesto en el punto.

Tabla 2: Relación entre ítems y sub-problemas

Ítems evaluados	Sub-problema
Llegada al punto de interés Pfn	Identificación de la ruta y desplazamiento sobre ruta predefinida
Detección de código QR del punto Pfn	Identificación de punto de interés
Detección de obstáculo en tramo Pin-Pfn	Identificación de obstáculo y generación de alarma en la ruta
Reproducción de audio correspondiente a punto Pfn	Recepción de información correspondiente a punto de interés

A los participantes se les permitía tocar los especímenes para complementar los audios con el tacto. Los ítems evaluados tienen relación directa con los sub-problemas identificados en la etapa de diseño (Tabla 3). Todos los ítems se evaluaron utilizando valores booleanos para indicar su nivel de logro: cumplido: 1, no-cumplido: 0.

Tabla 3: Cálculos de tasa de éxito por ítem y participante

Datos	Participantes				Prom. TE (x Ítem)	Desv. Estd.
	M	A	S	J		
Llegada a Pfn	1,00	0,97	0,9	0,93	0,95	0,04
Detección de obstáculo en Tn	0,97	0,97	0,93	0,86	0,93	0,05
Detección de Código en Pfn	0,93	0,93	0,86	0,8	0,88	0,06
Reproducción de audio Pfn	0,93	0,93	0,86	0,8	0,88	0,06
Promedio TE (x Participante)	0,96	0,95	0,89	0,85		
Desv. Estd.	0,03	0,02	0,03	0,06		

Luego de la etapa de evaluación se realizó la de encuesta (Tabla 4), con el objetivo de conocer la percepción que tuvieron los participantes con respecto a la interfaz, su comodidad y utilidad.

Tabla 4: Encuesta final para los participantes

#	Pregunta
P1	¿Requirió de información adicional para usar la interfaz? Explique
P2	¿Considera que usar señales podotáctiles ha sido un acierto? Explique.
P3	¿Cuál fue su percepción de seguridad al desplazarse sin bastón?
P4	¿Sintió incomodidad al usar el dispositivo?Explique

P5	¿Qué opina acerca del peso?
P6	¿Sintió confusion o saturación, generada por la información que le entregaba la interfaz (táctil, auditiva, vibratoria)?
P7	¿Aprendió algo acerca de los espécimenes expuestos?

5 Resultados

Se calculó la tasa de éxito para cada ítem de evaluación y para cada participante. En la Tabla 3 se aprecia que la mayor tasa de éxito por ítem corresponde a la *llegada al punto de interés* (punto final del tramo) y la más baja es compartida –con iguales valores– por los ítems *Detección de código* y *Reproducción de audio* al llegar al punto final del tramo. Esto último tiene lógica debido a que el dispositivo de navegación fue programado de tal manera que la reproducción de audio dependiera de la previa detección del código QR. También indica que el software del dispositivo se desempeña según lo esperado. Sin embargo, el hecho de no detectar igual número de códigos para igual número de llegadas al punto de interés tiene que ver con aspectos técnicos y humanos.

La luz tenía una mayor incidencia en algunos puntos de la sala 2, generando reflejos que no eran filtrados físicamente previa llegada al sensor de la cámara. Factores humanos, como la velocidad de avance, reducía drásticamente el tiempo disponible para que el sensor visual captara en un sólo cuadro la imagen correspondiente al código QR del punto de llegada. Adicionalmente, la orientación del participante al aproximarse al punto *Pfn,* desplazándose sobre las guías, no debe superar un ángulo de 45 grados (±45°) con respecto a la línea central de la guía podotáctil, pues de otra manera no se podrá garantizar que el sensor visual detecte el código QR. En el caso particular del participante J (hombre adulto y experimentado, pero con las tasas de éxito más bajas), se observó que su técnica de desplazamiento sin bastón sobre señales podotáctiles implicaba *palpar* con el pie las guías, manteniendo un ángulo de ataque a las mismas de entre 30° y 45° (aproximado, no medido), hasta llegar a las marcas del punto de interés. Esta característica en su desplazamiento también justifica que la detección de obstáculo sobre las guías tuviera baja tasa de éxito. En general, como la llegada al punto depende de la habilidad y experiencia de los participantes para desplazarse sobre las guías, la tasa de éxito es alta. La detección de obstáculos depende en mayor medida del desempeño de hardware y el software del dispositivo, por lo cual la tasa de éxito fue mayor al 90%. La detección de código fue afectada en mayor medida por los reflejos de la luz en la sala, la velocidad y dirección de desplazamiento sobre las guías de los usuarios, y su tasa de éxito fue del 88%. Además, dado que de esta detección dependía la reproducción de audio, las tasas de ambos ítems son iguales.

5.1 Resultados cualitativos

Los participantes opinan que la información entregada antes de iniciar los desplazamientos y los entrenamientos previos sobre algunos de los tramos, fue suficiente para comprender el método de navegación a través de la exposición. En cuanto a la percepción de seguridad, los hombres adultos (A y J) se sintieron seguros al caminar sin bastón debido a su experiencia con señales podotáctiles, sin embargo, en el caso de J, su seguridad le animó a avanzar rápidamente, por lo que perdió la ruta y generó fallas en el desempeño de la interfaz. Por otro lado, M y S, siendo menores y con menos experiencia en señales, tuvieron un sentimiento de inseguridad que fue superado después de las instrucciones iniciales y el entrenamiento básico. M se desplazó con total seguridad, pero más lento que J, mientras que S tenía tendencia a perder la ruta por inseguridad en su desplazamiento. Los participantes se manifestaron cómodos los audífonos de conducción

ósea, que les permitían recibir información sin bloquear el sonido ambiente, lo cual aumentó la percepción de seguridad. De igual manera consideraron cómodo el dispositivo ponible, pero les pareció poco práctico el morral en que se alojaba el *laptop*. No hubo confusiones generadas por saturación de información entregada por la interfaz. A lo sumo se recibía dos tipos de información a la vez: *táctil + vibrotáctil*, o *táctil + audio*.

6 Conclusiones y trabajos futuros

Los problemas de adquisición de la imagen asociados a la iluminación y el tiempo de detección del código QR serán corregidos, puesto que no detectar el código afecta directamente la reproducción de los audios en los puntos de interés. Con el mismo objetivo, los usuarios serán instruidos para que mantengan una dirección de desplazamiento con mínima oscilación alrededor de la línea central de las guías podotáctiles. Además, se sugerirá una velocidad baja de avance para garantizar que el código QR pueda ser captado, para no salir de la ruta predefinida por las señales y para detenerse a tiempo cuando el dispositivo vibrotáctil indique la presencia de un obstáculo. Además, se rediseñará el dispositivo ponible para eliminar la conexión con el *laptop*, de esta manera se convertirá en un computador ponible.

Después de mejorar estos aspectos, se realizará experimentos con un grupo de individuos que satisfaga un tamaño de muestra con alto nivel de confianza estadística. Dicho grupo deberá ser heterogéneo en edades, nivel de experiencia y habilidades de desplazamiento con bastón en interiores y exteriores.

Por medio de un nuevo proyecto se desarrollarán dos versiones con tecnologías diferentes a las utilizadas en el actual dispositivo. Con el primero se pretende simplificar la complejidad tecnológica del dispositivo, para reducir costos de producción y permitirle a otros museos tener acceso a un dispositivo que da solución a una necesidad real y que tiene impacto directo e inmediato en la población con discapacidad visual. En el segundo se utilizarán tecnologías de uso común (*smartphones*), para realizar la actividad de navegación autónoma sin necesidad de fabricar un dispositivo completo.

Las diferentes versiones de dispositivos serán registradas, socializadas y entregadas a los diferentes museos nacionales, en respuesta a una necesidad identificada desde el Museo de Ciencias Naturales de La Salle, un proyecto del ITM.

Referencias

[1] ICOM (2006). Código de deontología del ICOM para los museos. Consejo Internacional de Museos (ICOM).

[2] Serrat, N. & Font, E. (2005). Técnicas expositivas básicas. En Santacana, J. & Serrat, N. (Eds.), Museografía didáctica, pp. 253-301. Barcelona: Ariel.

[3] Hernández, F. (1994). Manual de museología. Madrid: Sintesis.

[4] Amato, P. (2004). Proyectar un museo: Nociones fundamentales. Cuadernos IIla 22. España: Iila.

[5] Gallego, S. & Núñez, A. (2014). Museos para todos: evaluación de una guía audiodescriptiva para personas con discapacidad visual en el museo de ciencias. Revista española de discapacidad 2(2), pp. 145-167.

[6] Silverman, L. (2007). Los museos en una nueva era: Los visitantes y la construcción de significado. Online [Jul. 2014].

[7] Jiménez, C.; Seibel, C. & Soler, S. (2012). Museos para todos: La traducción e interpretación para entornos multimodales como herramienta de accesibilidad universal. Monti 4, 349-383.

[8] Fallah, N. et al. (2012). The user as a sensor: navigating users with visual impairments in indoor spaces using tactile landmarks. Proceedings SIGCHI Conference on Human Factors in Computing Systems, pp. 425-432. Austin, USA.

[9] Nakamura, K.; Aono, Y. & Tadokoro, Y. (1997). A walking navigation system for the blind. Systems and computers in Japan 28(13), 36-45.

[10] Ojeda, L. & Borenstein, J. (2007). Personal dead-reckoning system for GPS-denied environments. Proceedings International Workshop on Safety, Security and Rescue Robotics, pp. 1-6. Rome, Italy.
[11] Ojeda, L. & Borenstein, J. (2007). Personal dead-reckoning system for GPS-denied environments. Proceedings International Workshop on Safety, Security and Rescue Robotics, pp. 1-6. Rome, Italy.
[12] Jain, D. (2014). Path-guided indoor navigation for the visually impaired using minimal building retrofitting. Proceedings 16th international ACM conference on Computers & accessibility, pp. 225-232. Rochester, USA.
[13] Xie, B. et al. (2014). LIPS: A light intensity based positioning system for indoor environments. ArXiv ID: 1403.2331.
[14] Nakajima, M., & Haruyama, S. (2013). New indoor navigation system for visually impaired people using visible light communication. EURASIP Journal on Wireless Communications and Networking 2013(1), pp. 1-10.
[15] Chumkamon, S.; Tuvaphanthaphiphat, P. & Keeratiwintakorn, P. (2008). A blind navigation system using RFID for indoor environments. Proceedings 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, pp. 765-768. Krabi, Thailand.
[16] Filipe, V. et al. (2012). Blind navigation support system based on Microsoft Kinect. Procedia Computer Science 14, pp. 94-101.
[17] Manduchi, R. et al. (2015). Vibrotactile Guidance for Wayfinding of Blind Walkers. IEEE Transactions on Haptics 8(3), pp. 306-317.
[18] Shoval, S.; Borenstein, J. & Koren, Y. (1998). Auditory guidance with the navbelt-a computerized travel aid for the blind. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 28(3), pp. 459-467.
[19] Loomis, J. et al. (2005). Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance. Journal of visual impairment & blindness 99(4), 219-232.
[20] Ertan, S. et al. (1998). A wearable haptic navigation guidance system. Proceedings Second International Symposium on Wearable Computers, pp. 164-165. Pittsburgh, USA.
[21] Tan, H. et al. (2003). A haptic back display for attentional and directional cueing. Haptics-e 3(1), pp. 1-20.
[22] Tsukada, K. & Yasumura, M. (2004). Activebelt: Belt-type wearable tactile display for directional navigation. Lecture Notes in Computer Science 3205, pp. 384-399.
[23] Ulrich, K. et al. (2004). Diseño y desarrollo de productos: enfoque multidisciplinario. México: McGraw-Hill.

Optimización del enrutamiento de datos en una red de sensores inalámbricos basado en un algoritmo multi-objetivo

Ángela Rodríguez
Armando Ordóñez
Luis Eraso
Fundación Universitaria de Popayán – Colombia

1 Introduction

Among the wide spectrum of elements involved in the Internet of things (IoT), wireless sensor networks (WSNs) play a crucial role due to the fact that they collect data which will make possible decision making processes [1]. However, WSN sensors have limited energy sources which cannot be recharged once networks are deployed; therefore, one of the key challenges is energy saving [2]. The highest energy consumption is produced during data transmission phase due to the operation of radiofrequency modules [3, 4], hence optimization of data transmission can make a substantial contribution to battery saving by optimizing data gathering paths. Gathering paths are directly associated to WSN topology which determines diverse factors such as: distance between sender and receiver sensor nodes, packets length, etc. Therefore optimizing data gathering implies optimization of the different elements associated to WSN topology, which in turn may be seen as a multi objective optimization problem.

Previous approaches [3-7] have incorporated bio inspired techniques in routing optimization. In a preliminary approach of the present work, the simple genetic algorithm was used to incorporate characteristics of a routing WSN tree as variables of the fitness function [8]. However, any of the lasts works considered multi objective optimization.

After the proposal presented in [8], it was identified the need of a technique that optimizes several objectives in parallel and usually contradictory; that way MOR4WSN was created as a robust platform for optimization of WSN routing based on a multi-objective evolutionary technique known as NSGA-II (Fast Elitist Non Dominated Sorting Genetic Algorithm) [9]. In [10] one can find an explanation of why NSGA-II is an adequate meta-heuristic to solve WSN routing; also, [10] details why each component of NSGA-II must be adapted to form MOR4WSN according to WSN conditions.

MOR4WSN uses two objective functions for creating optimal WSN routes directly related with hierarchical topologies [11]. Besides, the current way how NSGA-II creates its initial population required a hard modification to become MOR4WSN [12], as well as the crossover operator performance [13]. Subsequently MOR4WSN uses a binary index for refining the search [14]. As a multi-objective approach, MOR4WSN gives not a unique solution to a problem but a set of them known as Pareto Set [15, 14]. The present paper describes the design and implementation of the mutation operator in MOR4WSN, which also needed to be tuned in order to preserve valid routing trees in all generations during MOR4WSN execution.

The rest of this paper is organized as follows: Section 2 presents the conceptual background as well as WSN modeling using genetic algorithms. Section 3 presents the algorithm design for mutation operator adapting. Section 4 shows experimental evaluation of MOR4WSN in WSN networks, and Section 5 details conclusions and future work.

2 WSN modeling in genetic algorithms

1.1 Wireless sensor networks terminology

WSNs are composed of sensors known as *source nodes* whose function is to gather information from environment and to communicate with other nodes through wireless technologies in order to send data to a central point known as *base station* where information analysis can be performed. As a WSN lacks of physical infrastructure, different topologies can be adopted according to domain needs. The most common topology for high scale network deployments is tree based or hierarchical. Figure 1 shows a directed acyclic graph representing a WSN rooted with the base station while the rest of the nodes are data sources and the directed edges indicate its next hop node for data transmission.

Figure 1: WSN with hierarchical topology

Next, terminology associated with hierarchical topologies and general WSN is presented:

- *Source node:* node that picks up data from environment and sends it to the base station.
- *Leaf node*: node that does not have children in the tree, that is, the one that does not receive data from other nodes to be transmitted to the base station.
- *Tree depth*: maximum amount of hops between nodes that must be done by a source node (a leaf node) for reaching a base station.
- *Routing node*: receives data from other nodes, and retransmit it to other node adding its own data. *WSN routing*: needed mechanisms to carry out data transmission from source to destination [16].
- *Network longevity*: time period counted from WSN start-up until the moment in which one node or a percentage of nodes is unable of sending data to base station. Network longevity is directly affected by energy consumption, which in turn is associated to data routing [4, 5].

1.2 Genetic algorithms

Genetic algorithm starts from a set of randomly generated possible solutions (*initial population*) represented as *chromosomes* (or *individuals*) that consist of a set of gens, all of them with the same length, what means all chromosomes having equal quantity of gens. Gens at chromosomes keep data and value through indices and alleles that represent any possible problem solution information unit as shown in Figure 2. Genetic algorithm uses a *fitness function* to evaluate each individual along all *generations* assigning a rank value that determines the proximity of each one to the optimal solution. In order to get closer to the optimal solution, diverse genetic operators are applied to the initial population: *selection* choose a percentage of the best ranked chromosomes from the current population to be parents of the next generation and operates them by pairs applying the process of *crossover* which allows to create offspring (*children chromosomes*) from the combination of the parent's genetic information. *Mutation* genetically alters individual information in order to improve criteria to be optimized [17].

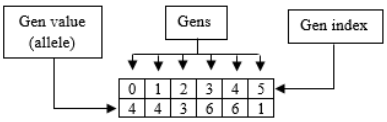


Figure 2: Chromosome with 6 gens

1.3 Multi-objective genetic algorithms (MOGA)

In contrast to simple (mono-objective) genetic algorithms, multi-objective genetic algorithms (MOGA) aim to solve problems that require optimizing several objectives at the same time, which may be opposed to each other. Given its nature, multi-objective algorithms give not a unique solution but a set of them, which are all optimized solutions having all objectives into account; this concept is called Pareto Dominance and is illustrated in Figure 3. Here functions f_1 and f_2 represent problem objectives. Regarding the Pareto Front or Non-Dominated Front, a solution *a* 'dominates' *b* provided that: *a* is not worse than *b* in all objectives, and *a* is strictly better than *b* in at least one objective [18]. That way, the solutions that are not dominated by any other else conform the Non Dominated Front. In Figure 3 these non-dominated solutions are represented with the points marked as 3 and 5. Selecting a unique optimal solution from Pareto front is not a trivial task, hence recent approaches have presented mechanism for this selection based on binary index [14].

Most of the existing multi-objective evolutionary algorithms have a computational complexity of $O(mN^3)$ [18], where m is the number of objectives and N is the population size. In order to get a lower complexity, NSGA II (Fast Elitist Non Dominated Sorting Genetic Algorithm) was created to refine search spaces by applying two concepts: elitism and diversity preservation. With this modification, NSGA II reduces complexity to $O(mN^2)$ [9, 18].

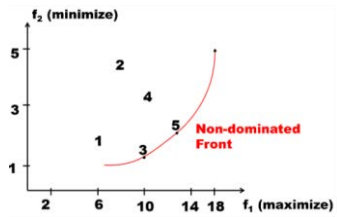


Figure 3: Pareto Front and Dominance [9]

1.4 WSN modeling in genetic algorithms

In MOR4WSN hierarchical WSNs are modelled in chromosomes as treated in [3-7]. Let N be the total of source nodes in a WSN (Figure 4a), each node is labelled with an integer from 0 to N-1 and base station is always marked as N. Chromosome in Figure 4b represents such routing scheme where the index corresponds to the identifier of a source node and the allele indicates the next-hop node of such index. i.e. in Figure 4b the gen value in index 0 is 2 indicating that node 0 selects node 2 to transmit its data, node 2 transmits to node 5, the value in 5 is 6 indicating that the next node is the base station; therefore the full gathering path from node 0 is expressed as the path 0 → 2 → 5 → 6. In MOR4WSN one important condition of chromosomes to be a valid routing solution, is that the next-hop node of a source node has to be within its communication range (neighbor nodes). Neighbor nodes are known during WSN deployment; MOR4WSN uses a mechanism (a neighbor's matrix) to assure at least two neighbor nodes per source node, and that way to avoid isolated nodes [12].

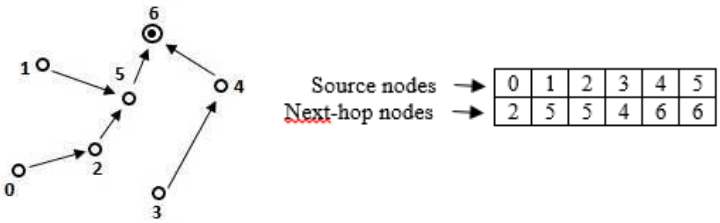


Figure 4: (a) Directed routing tree in a WSN; (b) Chromosome representing the directed tree

3 Mutation in MOR4WSN

Like the biological concept, mutation operator in NSGA-II is in charge of altering the genetic information of one or some gens (according to the mutation probability value p_m) at a given chromosome of the current population in order to inject genetic diversity to children. Alleles to be mutated are randomly chosen; mutation is done during all evolution process.

In MOR4WSN, mutate an individual means to switch the next-hop node of a randomly chosen source node In Figure 5 one can realize that when applying mutation to an index value (an allele) a cycled graph can appear as the new routing topology and also, as it happens with the crossover process [13], existence of the base station as allele is not assured in the new routing structure after applying mutation. Besides, to get a valid routing topology is essential that the new allele is in the list of neighbors of the altered index.

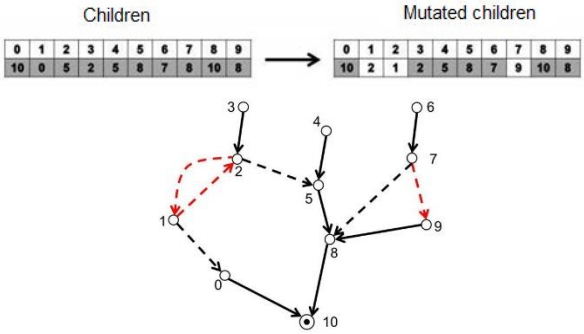


Figure 5: Mutation in NSGA-II

Thus, as was done for lasts operator adaptations [12, 13], MOR4WSN needs to preserve tree topologies along generations, and conditions mentioned in last paragraph must be complied. Mutation adaptation is as follows. The main function (steps 1 and 2) depends of the recursive subroutine *compute_descendants* (steps 3 to 5) and the subroutine *is_leaf* (step 6).

1. Let *i* be an index to be altered, let *z* be the amount of index to be altered, let *ind* be the individual.
2. For *i*=0; *i*<*z*; *i*++ do

Associate to *i* its neighboring list.

Mark all *i*-descendants
If *i* - next hop node is not an *i*-neighbor and is not part of *i*-descendants
Yes → ind[i] == BS
No → *i* will have a different treatment 'cause one have to assure that *i* remains at any allele value.
Stop-for.

Compute-descendants (x, ind, desc) of a node x in the tree.

3. Let *x* be the node, *ind* be the individual and *desc* be the descendants set of *x*.
4. *desc* ← Ø
5. For (*i*=0; *i*<=N; *i*++) do
   If ind[*i*] == *x*
   Yes → desc ← {i}
   Compute-descendants (*i, ind, desc*)
   Stop-for.

Is_leaf: is the node *x* a leaf in the tree?

6. For (*i*=0; *i*<=N; *i*++) do
   If ind[*i*] == *x*
   Yes → return 0
   Continue
   return 1
   Stop-for.

4 Evaluation

Evaluation compares WSN lifecycle (in terms of sensing and data gathering rounds) using MOR4WSN as routing scheme with other routing technique for hierarchical topologies Tree Routing (TR, which is used by ZigBee). Table 1 lists simulation parameters values; homogeneous networks are assumed, that means all nodes having the same initial energy level and its covering range being the same. Results of comparative tests are shown in Figs. 6 and 7 where can be seen that inclusion of multi objective alternative for selecting tree topology increases network longevity even up to 23% compared with an standard technique. In both evaluated cases, albeit fluctuations in curves showing the number of rounds, MOR4WSN exceeds in 85% of measures the longevity of the standard Tree Routing whose only criteria for selecting paths is the number of hops from source node to base station [19]. Equally, Figures 6 and 7 shows that advantages in life cycle of MOR4WSN are independent of the number of nodes in the network.

Table 1: Simulation parameters

Parameter	Value	Unity / observation
Population size	N	N = number of nodes
Generations	2N	
Crossing probability	0.5	----
Mutation probability	0.5	----
Initial node energy	1000	Unities of energy timer
Packet size	128	Bytes.
Covering range of nodes	2%	Percent of total deployment area
Minimal energy threshold	10%	Percent of initial energy.

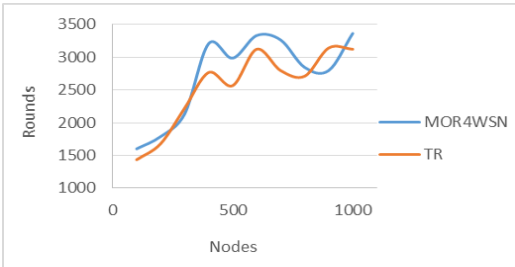


Figure 6: WSN life cycle until the first node dies

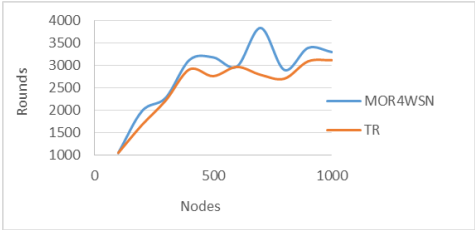


Figure 7: WSN life cycle until 10% of nodes die

5 Conclusions

MOR4WSN is an evolutionary approach for optimization of WSN topologies that aims to reduce the level of energy consumption during data gathering rounds, and that way get a longer lifecycle for the network. This approach involves the development of objective functions, initial population, and diverse operators such as crossover, as well as a binary index based approach for refining solutions search. The present paper describes the design and implementation of the mutation operator in MOR4WSN, which needed to be tuned in order to preserve valid routing trees in all generations during MOR4WSN execution. The algorithm presented here contributes to preserve the trees during the whole execution of MOR4WSN and in fact according to simulation tests results, this evolutionary multi-objective proposal extends the lifetime of the network regarding standard protocols.

Acknowledgements

The authors acknowledge to GIT group of the University of Cauca in Colombia for the academic support as well as to the I3A institute of the University of Castile-La Mancha in Spain. Financial support is acknowledged to University of Cauca and to the Administrative Department of Science Technology and Innovation of Colombia – Colciencias.

References

[1] Atzori, L.; Iera, A. & Morabito, G. (2010). The internet of things: A survey. Computer networks 54(15), pp. 2787-2805.

[2] Whitmore, A.; Agarwal, A. & Xu, L. (2014). The Internet of Things - A survey of topics and trends. Information Systems Frontiers 17(2), pp. 261-274.

[3] Bari, A. et al. (2009). A genetic algorithm based approach for energy efficient routing in two-tiered sensor networks. Ad Hoc Networks 7(4), pp. 665-676.

[4] Chakraborty, A. et al. (2011). A Genetic algorithm inspired routing protocol for wireless sensor networks. International Journal of Computational Intelligence Theory and Practice 6(1), pp. 1-8.

[5] Gupta, S. et al. (2013). GAR: an energy efficient GA-based routing for wireless sensor networks. Lecture Notes in Computer Science 7753, pp. 267-277.

[6] Islam, O.; Hussain, S. & Zhang, H. (2007). Genetic algorithm for data aggregation trees in wireless sensor networks. Proceedings 3rd IET International Conference on Intelligent Environments, pp. 312–316. Ulm, Germany.

[7] Apetroaei, I. et al. (2011). Genetic algorithms applied in routing protocols for wireless sensor networks. Proceedings 10th Roedunet International Conference, pp. 1-6. Iasi, Romania.

[8] Vivas, A. & Muñoz, J. (2013). Adaptación del algoritmo genético para la creación de árboles de enrutamiento en una red de sensores inalámbrica. Lámpsakos 10, pp. 34-42.

[9] Deb, K. et al. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. Lecture notes in computer science 1917, pp. 849-858.

[10] Rodríguez A. & Corrales, J. (2016). Adaptación de una metaheurística evolutiva para generar árboles enrutadores en una red de sensores inalámbricos del contexto de la agricultura de precisión. Revista Ingenierías Universidad de Medellín 30. *In press*.

[11] Rodríguez, A.; Ordóñez, A. & Falcarin, P. (2015). Energy optimization in wireless sensor networks based on genetic algorithms. IEEE Technically Co-Sponsored SAI Intelligent Systems Conference 2015. London, UK.

[12] Rodríguez, A.; Ordóñez, A. & Ordóñez, H. (2015). An adaptation of the NSGA-II algorithm for energy efficient routing in sensor networks. Proceedings 14th Mexican International Conference on Artificial Intelligence. Cuernavaca-Mexico.
[13] Rodríguez, A. et al. (2015). Adapting NSGA-II for hierarchical sensor networks in the IoT. In Complex Adaptive Systems Conference. San Jose, USA.
[14] Rodríguez, A.; Ordóñez, A. & Ordonez, H. (2015, May). Energy consumption optimization for sensor networks in the IoT. Proceedings IEEE Colombian Conference on Communications and Computing, pp. 1-6. Popayán, Colombia.
[15] Cabezas, I. & Trujillo, M. (2012). A method for reducing the cardinality of the Pareto front. Lecture Notes in Computer Science 7441, pp. 829-836.
[16] Cui, X. & Liu, Z. (2009). BCEE: a balanced-clustering, energy-efficient hierarchical routing protocol in wireless sensor networks. Proceedings IEEE International Conference on Network Infrastructure and Digital Content, pp. 26-30. Beijin, China.
[17] León, A. (2009). Diseño e implementación en hardware de un algoritmo bio-inspirado. Tesis de Maestría. Instituto Politécnico Nacional. México.
[18] Deb, K. et al. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2), pp. 182-197.
[19] Ortiz, A. (2011). Técnicas de enrutamiento inteligente para redes de sensores inalámbricos. Tesis de Maestría. Universidad de Castilla-La Mancha. España.

Análisis y diseño de un prototipo suelo generador de energía eléctrica utilizando materiales piezoeléctricos

Dionel Castro S.
Emilsen Arias C.
Fundación Universitaria de San Gil – Colombia

1 Introducción

La generación de energía verde y eficiente utilizando materiales piezoeléctricos [1] se ha convertido en un reto [2, 8]. En este proyecto se pretende mostrar una alternativa que no signifique riesgos para la naturaleza y sea parte de una solución para preservar el planeta. En los últimos años se ha despertado el interés por el aprovechamiento de la energía cinética y potencial presente en el tráfico vehicular [5, 9], en parte por la necesidad de contar con dispositivos energéticamente autónomos, tales como lámparas para alumbrado público, semáforos, sensores de trafico, cámaras de vigilancia, entre otros [10]. Uno de los inconvenientes para este aprovechamiento energético es obtener dispositivos que aprovechen al máximo la configuración de los mismos y sirvan como parte activa de fuentes de alimentación.

Este estudio tiene como tema central los materiales piezoeléctricos, capaces de responder a las deformaciones físicas mediante la generación de señales eléctricas, un fenómeno llamado la *piezoelectricidad* [11]. En este caso, el material será sometido a la presión ejercida por el peso de las personas que se desplazan en motocicletas sobre el piso generador. El enfoque se relaciona con aplicaciones de tráfico vial, donde se utilizan como generadores eléctricos. Un asunto importante para realizar este proyecto, es disponer del modelo eléctrico equivalente del material piezoeléctrico utilizado, que responda correctamente a la excitación provocada por el tráfico de motocicletas como fuente de energía mecánica generadora.

Este trabajo tiene como objetivo diseñar e implementar un prototipo que genere energía con base en materiales piezoeléctricos. A partir de datos de caracterización determinados con el prototipo generador, se plantea un modelo eléctrico del sistema que permita definir la configuración óptima de elementos, la forma en que se deben conectar y los elementos de circuito destinados a conseguir la máxima eficiencia en el almacenamiento de la energía.

2 Diseño y construcción del prototipo

El cable piezoeléctrico es el componente principal del prototipo que, como se aprecia en la Figura 1 y se describe en [12], se trata de un cable coaxial en forma de espiral calibre 20AWG con capacitancia nominal 980pF/m. En el prototipo el cable se encuentra dispuesto en forma de zigzag uniendo dos estructuras de madera, una fija y otra móvil, como se observa en la Figura 2.

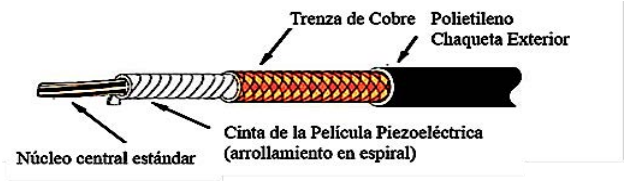


Figura 1: Cable Piezoeléctrico (METROLOG Controles de Medicao Ltda.)

Tal como lo explica el fenómeno piezoeléctrico, el movimiento de estas estructuras hace que el cable, bajo un esfuerzo mecánico [13], se estire y contraiga provocando una presión y descompresión que genera señales eléctricas con amplitudes cercanas a los 11Vpp, según resultados de pruebas realizadas (Tabla 2). La construcción del prototipo permitirá obtener, a partir de tráfico rodado, la forma y magnitud de la respuesta eléctrica del material piezoeléctrico. De esta manera se dispondrá de valores iniciales que permitirán obtener el modelo eléctrico, que posteriormente alimentará los procesos de análisis y simulación de los diferentes tipos de configuración de los elementos semiconductores, y así determinar la más adecuada.

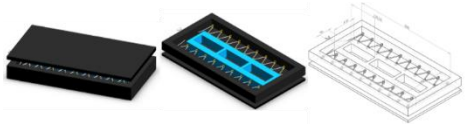


Figura 2: Diseño de la estructura mecánica del prototipo suelo generador

En la Figura 3 se observa en forma de zigzag el cable instalado dentro del prototipo construido con dos rampas adicionales que facilitan el paso de los vehículos, que serán la fuente de exitación mecánica.

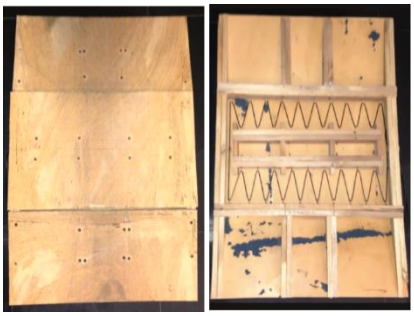
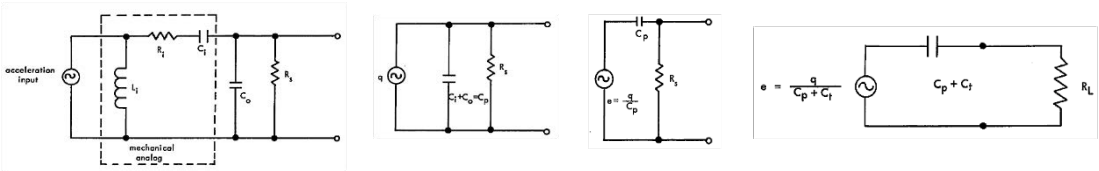


Figura 3: Prototipo suelo-generador

3 Modelo eléctrico del prototipo suelo-generador

El modelo que refleja el comportamiento eléctrico del prototipo se muestra en la Figura 4(c), que también representa el material piezoeléctrico utilizado sometido a excitación mecánica, donde el resistor R_s es la resistencia interna del material, normalmente excede los 20GΩ, por lo que puede ser ignorada. El modelo representativo se conforma con la asociación serie de un generador de tensión e con la capacidad equivalente del material C_p.



(a) (b) (c) (d)

Figura 4: Circuito eléctrico equivalente para un material piezoeléctrico: (a) Modelo completo, (b) Modelo de carga eléctrica, (c) Modelo eléctrico simplificado, (d) Modelo eléctrico simplificado con capacitancia externa y resistencia de carga

4 Series de Fourier y lugar geográfico para pruebas y ensayos

Las series de Fourier [14] permiten representar una función periódica como sumatorias de senoides para obtener el espectro de la serie en frecuencia. Cualquier

función periódica práctica se puede representar como suma de senoides. Esta representación, junto con el teorema de superposición, permite encontrar la respuesta de circuitos a entradas periódicas arbitrarias utilizando técnicas fasoriales. Una función periódica es la que se repite cada T segundos, en otras palabras, satisface una función $f(t)$.

$$f(t) = f(t + nT) \qquad (1)$$

Donde n es un entero y T es el periodo de la función.

De acuerdo con el teorema de Fourier, toda función periódica práctica de frecuencia ω_0 se puede expresar como una suma infinita de funciones seno o coseno, que son múltiplos enteros de ω_0.Por lo tanto, $f(t)$ se puede representar como:

$$f(t) = a_0 + a_1 cos w_0 t + b_1 sen w_0 t + a_2 cos 2w_0 t + b_2 sen 2w_0 t + a_3 cos 3w_0 t + b_3 sen 3w_0 t + \cdots \qquad (2)$$

Expresado de otra forma:

$$f(t) = \underbrace{a_0}_{\text{dc}} + \underbrace{\sum_{n=1}^{\infty}(A_0 \cos \omega_0 t + b_n sen \omega_0 t)}_{\text{ac}} \qquad (3)$$

Donde $\omega_0 = 2\pi/T$ es la frecuencia fundamental en radianes por segundo. Las senoides $sen\ n\omega_0 t$ o $cos\ n\omega_0 t$ son las armónicas n-ésimas de $f(t)$, que será impar si n es impar y par si n es par. Las contantes a_n y b_n son los coeficientes de Fourier. Según [14], las series de Fourier permiten representar una función de período T como sumatorias de senoides, tal como se resume en (4):

$$f(t) = a_0 + \sum_{n=1}^{\infty} A_n \cos(\omega_n t - \theta_n) \qquad (4)$$

Donde:

$$\omega_n = 2n\pi/T = 2n\pi f \qquad (5)$$

$$A_n = \sqrt{{a_n}^2 + {b_n}^2} \qquad (6)$$

$$\theta_n = tan^{-1}\left({b_n}/{a_n}\right) \qquad (7)$$

$$a_0 = \frac{1}{T}\int_{-T/2}^{T/2} f(t)dt \qquad (8)$$

$$a_n = \frac{2}{T}\int_{-T/2}^{T/2} f(t) \cos\left(\frac{2n\pi}{T}t\right) dt \qquad (9)$$

$$b_n = \frac{2}{T}\int_{-T/2}^{T/2} f(t) \sin\left(\frac{2n\pi}{T}t\right) dt \qquad (10)$$

Por lo tanto, la excitación mecánica del suelo-generador debe corresponder a un tráfico continuo de vehículos para cumplir la condición de periodicidad que exigen (2), (3) y (4), lo que permite aplicar el desarrollo la señal eléctrica generada por el cable en series de Fourier [12].

Luego del diseño y construcción del prototipo, se requiere seleccionar el lugar de pruebas para el desarrollo del modelo eléctrico, es por esto que como escenario de ensayos y ubicación del prototipo se elige la calzada de ingreso al parqueadero de motocicletas, en el campus de UNISANGIL, Yopal. Una de las fortalezas de esta ubicación es el flujo vehicular, cuyo comportamiento periódico ($T = 2.3s$) es regular y totalmente predecible en determinas horas del día (6:00am – 8:00am, 11:00am - 12:00m, 6:00pm –

7pm y 9:00pm – 11:00pm), situación que representa ventajas para la obtención del modelo eléctrico.

A pesar que el prototipo construido ejerce algo de fatiga sobre el material piezoeléctrico, es importante resaltar que su construcción facilita la transmisión de la presión de rodadura ejercida por las distintas motocicletas que por allí transitan (tráfico real), y así se obtienen grandes picos en las señales eléctricas generadas. El acceso al parqueadero permite el ingreso o salida de una motocicleta a la vez, situación que provoca un tráfico vehicular periódico. Es en estas condiciones se quiere simular el circuito de generación de energía asociado a la configuración de los elementos utilizados.

4.1 Lugar de emplazamiento

El prototipo para pruebas se ubica sobre la calzada de ingreso, quedando expuesto a posibles alteraciones o daños del mismo. La idea es quitar las rampas de subida de las motocicletas y enterrar bajo la calzada los materiales piezoeléctricos, logrando disponer de este escenario como el de probable uso y obteniendo menor fatiga de dichos materiales, la no-alteración de la movilidad de los usuarios, y de ser así, que no sea facilmente manipulable o alterable de forma intencional o casual.

5 Modelado eléctrico de cable piezoeléctrico bajo estímulo de tráfico rodado

Al tener la estructura mecánica de suelo-generador en la ubicación dispuesta, la adquisición y registro de datos se realiza con un osciloscopio RIGOL DS1052A, una sonda de pruebas RP2200 de multiplicidad 10× (10:1, 500 MHz BW) e impedancia de entrada 1MΩ ±2%, y un computador MacBook Pro (2.3 Ghz, procesador Intel Core i5, memoria de 4GB, Disco duro de 500GB) con el software ULTRASCOPE para DS1000E Series. A partir de datos de caracterización del tráfico rodado se plantea un modelo eléctrico del sistema, que ayude a definir la configuración ideal de elementos, la forma en que deben conectarse y los elementos de circuito destinados a conseguir la máxima eficiencia en el almacenamiento de la energía.

En la Figura 5 se presenta la forma de onda registrada en las pruebas efectuadas. En ella se resalta el intervalo temporal T_2, que representa el período de repetición de las señales seleccionadas que simulan el paso repetitivo de una motocicleta, T_1 es el intervalo temporal entre los estímulos recogidos debidos a cada eje de la motocicleta y A_m representa el valor de tensión de pico máximo.

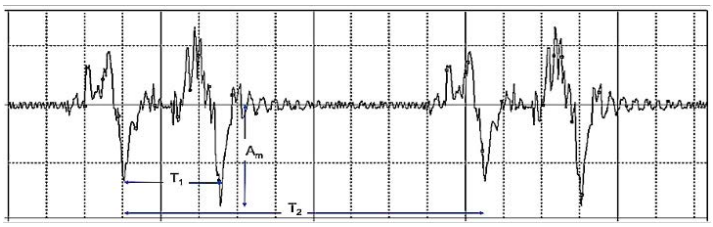


Figura 5: Señal eléctrica, registrada de un cable piezoeléctrico, por la presión ejercida de dos motocicletas.

En la Tabla 1 se presenta el resumen de 30 ensayos realizados sobre el prototipo: tiempos característicos T_2 y tiempos de las señales generadas, utilizados para determinar la frecuencia fundamental (frecuencia de paso de motocicletas de diferentes marcas y cilindrajes) y el modelo eléctrico. En la Tabla 2 se presenta el resumen de 11 ensayos efectuados: tiempos característicos T_1 y amplitud máxima A_m de las señales generadas debido al paso de motocicletas de diferentes marcas y cilindrajes, utilizados para

determinar el modelo matemático y eléctrico. Estas muestras han sido tomadas para una motocicleta por ensayo.

Tabla 1: Tiempos característicos T_2 de 30 señales generadas

No. de ensayo	T_2 (s)	No. de ensayo	T_2 (s)	No. de ensayo	T_2 (s)
1	2,409	11	1,962	21	2,391
2	1,981	12	1,896	22	2,034
3	2,589	13	2,603	23	1,860
4	2,252	14	2,304	24	2,543
5	2,427	15	2,379	25	2,209
6	2,331	16	2,407	26	2,425
7	2,570	17	2,296	27	2,398
8	1,997	18	2,421	28	1,902
9	2,003	19	2,178	29	1,883
10	2,185	20	2,312	30	2,452

Tabla 2: Tiempos característicos T_1, y amplitud máxima A_m

Ensayo	T_1 (s)	A_m (V)	Moto	Distancia ejes (m)	V (Km/h)
1	0,42000	6,32	Honda Cb110	1,257	10,77
2	0,62000	5,12	Yamaha Crypton 115	1,235	7,17
3	0,63000	4,32	Auteco Activ 110	1,227	7,01
4	0,56000	4,00	Honda Cb110	1,257	8,08
5	0,48000	5,20	Suzuki Vivax 115	1,23	9,23
6	0,50000	5,36	Yamaha Crypton 115	1,235	8,89
7	0,44000	5,04	Yamaha Bws 125	1,29	10,55
8	0,53000	6,00	Auteco Activ 110	1,227	8,33
9	0,72000	4,96	Pulsar 135 Ls	1,325	6,63
10	0,58000	5,12	Suzuki Vivax 115	1,23	7,63
11	0,56000	4,80	Akt Flex Ultra 125	1,27	8,16

6 Metodología para la obtención del modelo eléctrico equivalente

Una vez obtenidas las respuestas eléctricas, se plantea la síntesis del modelo eléctrico equivalente del material piezoeléctrico. El modelo que refleja el comportamiento eléctrico de un material piezoeléctrico sometido a excitación mecánica es el formado por la asociación serie de un generador de tensión con la capacidad equivalente del material, o alternativamente su equivalente Norton, mediante fuente de corriente en paralelo con dicha capacidad equivalente.

La forma de onda de la respuesta eléctrica del cable piezoeléctrico, sometido a estímulo equivalente a tráfico rodado, se obtiene al aplicar entre sus dos terminales una sonda de medida conectada a un canal de un osciloscopio. Este método representa una carga conectada al cable de 1 MΩ, somo se detalló previamente. Partiendo de la medida registrada se debe encontrar el modelo interno del cable piezoeléctrico, que actuando como generador incorpora una impedancia serie de tipo capacitivo. Este trabajo consiste en obtener el modelo eléctrico del cable que al aplicarle una carga del mismo valor produzca exactamente la misma forma de onda registrada en el osciloscopio, y permita experimentar mediante simulaciones cuál será su comportamiento cuando el piezoeléctrico se conecte a circuitos de acumulación de carga eléctrica.

En este método se asume que la excitación mecánica periódica proporcionada por las pruebas corresponderá a un tráfico continuo de motocicletas, así se cumplirá la condición de periodicidad en la señal que permite aplicar el desarrollo en serie de Fourier a la señal eléctrica medida y obtener su modelo lineal equivalente, a partir del cual y teniendo en

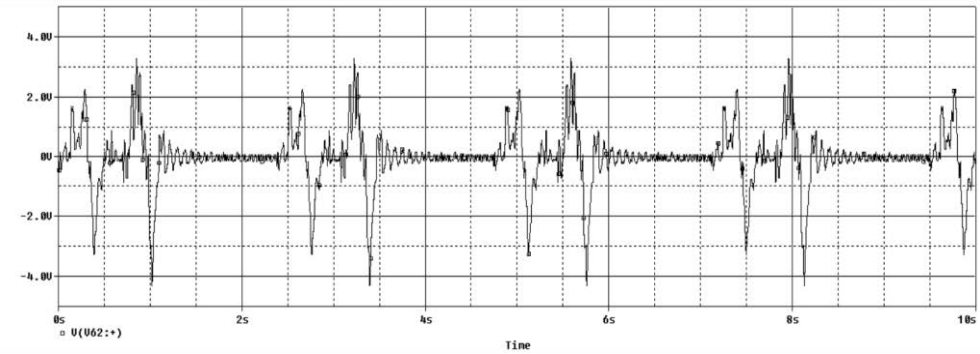
cuenta la capacidad serie y la impedancia de carga con la que se ha registrado la medida, se podrá calcular el equivalente del generador eléctrico buscado. Por lo tanto, para los ensayos efectuados se realizó el siguiente procedimiento [10]:

- Extracción de los valores de amplitud y tiempo de cada ensayo en formato texto con los valores correspondientes a un período eléctrico T_2.
- Obtención de las componentes de Fourier de la señal registrada de cada ensayo, hasta el armónico necesario, usando un generador de tensión $VPWL\_F\_RE\_FOREVER$ con datos texto, mediante laherramienta incorporada en el programa de simulación electrónica PSpice, sobre la frecuencia central dada por $f1 = (1/T_2)$.
- Comprobación que el número de componentes de Fourier [14] elegido reproduce la señal eléctrica original de partida. Para ello se construye la asociación serie de los generadores sinusoidales de tensión mediante el desarrollo de Fourier y se representa junto con el generador $VPWL\_F\_RE\_FOREVER$, cargando ambos en condiciones próximas a circuito abierto (1000 $G\Omega$ en el simulador).
- Obtención del modelo real mediante la asociación serie de los generadores internos equivalentes y el condensador equivalente serie del material piezoeléctrico.
- Validación del modelo eléctrico real con carga de 1 MΩ, equivalente a la sonda del osciloscopio utilizada en la adquisición de las respuestas eléctricas reales, por comparación con la gráfica disponible de los datos de partida procedentes de los ensayos.

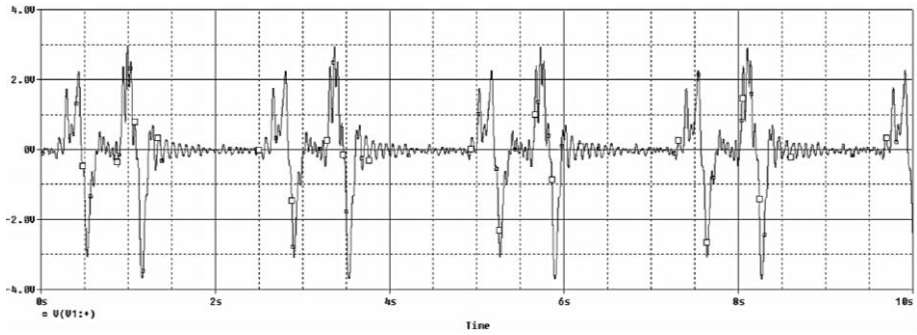
La realización de este proceso aplicado a los ensayos conduce a la utilización de 60 generadores sinusoidales en serie con la capacidad del material, consecuencia del desarrollo en serie de Fourier calculado para cada ensayo (Tabla 3). Se respeta el número completo de posiciones decimales que entrega la herramienta de cálculo del simulador, para evitar la generación de un ruido eléctrico acumulativo en el tiempo que se produce. En la Figura 6 se ilustra el método.

Tabla 3: Resumen del análisis de Fourier entregado por la herramienta pspice a/d

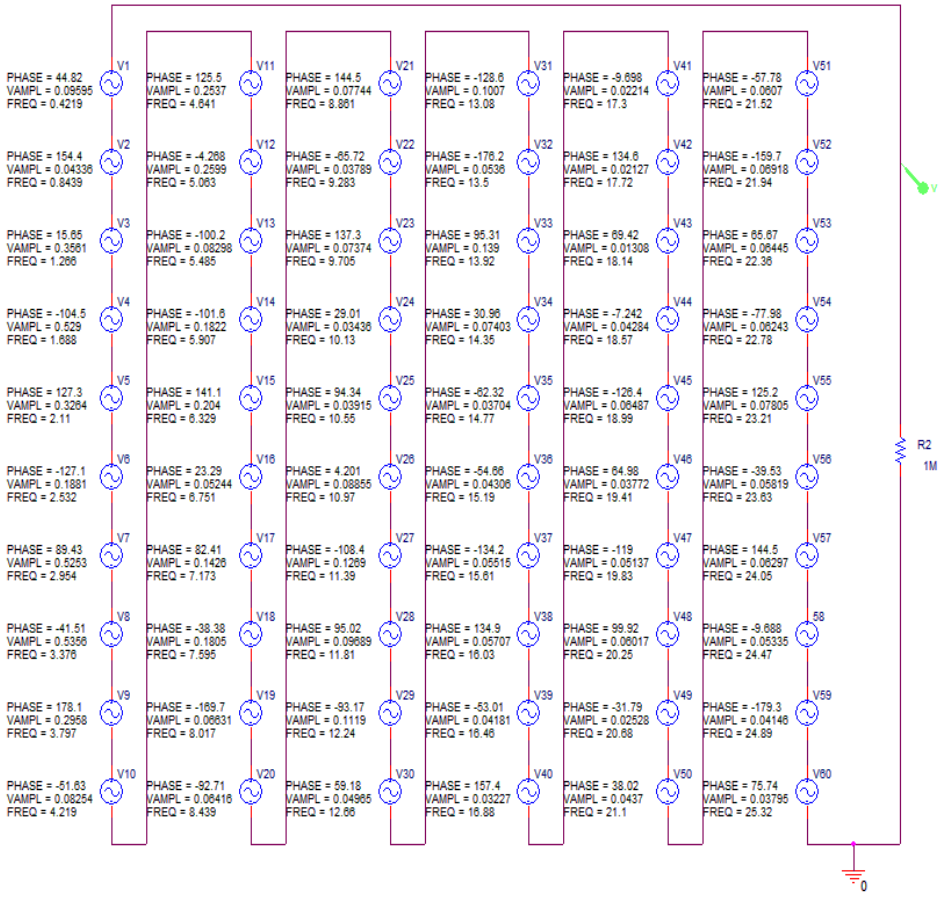
HARMONIC No.	FREQUENCY (HZ)	FOURIER COMPONENT	PHASE (DEG)
1	0,4219	0,0960	44,8200
2	0,8439	0,0434	154,4000
3	1,2660	0,3561	15,6500
4	1,6880	0,5290	-104,5000
5	2,1100	0,3264	127,3000
6	2,5320	0,1881	-127,1000
7	2,9540	0,5253	89,4300
8	3,3760	0,5356	-41,5100
9	3,7970	0,2958	178,1000
10	4,2190	0,0825	-51,6300



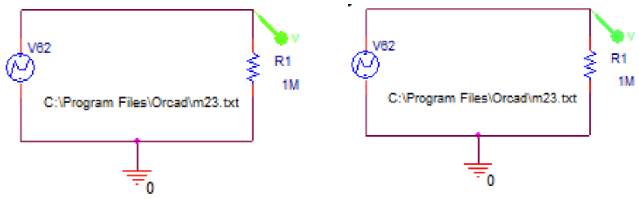
(a)



(b)



(c)



(d)

Figura 6: Validación mediante simulación del modelo matemático y eléctrico del material piezoeléctrico [15] de la Tabla 3. (a) Señal generada por el cable, (b) Señal obtenida por el modelo eléctrico. (c) Circuito simulado de la señal eléctrica correspondiente al modelo eléctrico. (d) Circuitos simulado que genera señal real

En la Figura 6 se presenta el resultado final del procedimiento al comparar la señal introducida al simulador correspondiente a los ensayos y la producida por el modelo equivalente final, representado mediante varios generadores $Vsin$ en serie sobre una carga de 1MΩ, que reproduce el efecto de la sonda de prueba del osciloscopio utilizado para su adquisición. La comparación de ambas señales presenta una alta coincidencia, tanto en forma de onda como en valores instantáneos, comprobando que con el desarrollo hasta el armónico 60 se puede reproducir (en las mismas condiciones de carga de los ensayos) la forma de onda de la señal original, tolerando un error en las oscilaciones mecánicas que provoca el funcionamiento del prototipo y que no contribuye al estudio sobre acumulación de carga que se describe en este trabajo.

7 Uso del modelo para evaluar la optimización del aprovechamiento energético

A continuación se describe la utilización de los modelos lineales equivalentes obtenidos para cuantificar y optimizar el método de acumulación de la carga liberada. Se toma este estudio planteando que la disposición de los cables de tipo piezoeléctrico que actuarán como generadores de carga se hallarían de forma consecutiva.

El estudio compara diferentes topologías rectificadoras según el tipo de diodo semiconductor empleado, con el objetivo de almacenar la mayor carga posible de la generada en el compuesto piezoeléctrico. Para realizar la compración se realizan las correspondientes simulaciones electrónicas [15] con el simulador PSpice, una herramienta de reconocido prestigio dada la calidad de los modelos funcionales de los diferentes semiconductores que implementa y de su motor de cálculo de resolución de las ecuaciones circuitales.

Para recopilar la información se emplean valores constantes, tanto de la capacidad como de la resistencia terminal de carga. La modificación del valor de la capacidad solamente contribuye al tiempo que necesitará el circuito acumulador para alcanzar su valor de régimen permanente. Fijando el valor de la capacidad se puede utilizar un tiempo de simulación fijo, facilitando la agrupación de los resultados gráficos. La contribución de un cable actuando como generador se resuelve utilizando estructuras rectificadoras basadas en puente de diodos. En la Figura 7 se presenta la evolución de la tensión en el condensador de acumulación de carga, agrupando los resultados en función de la tecnología de los semiconductores para un cable.

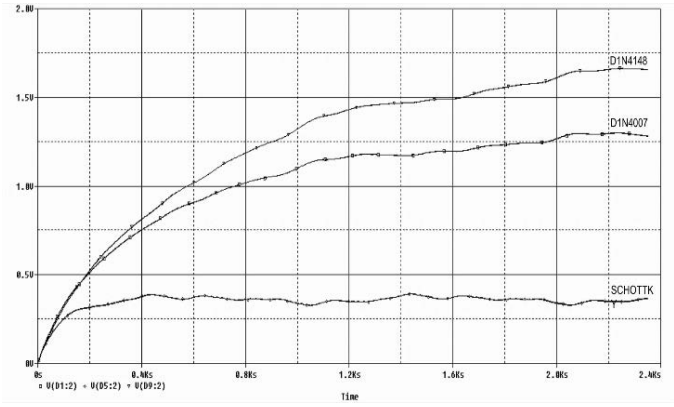


Figura 7: Voltaje en el capacitor en configuración puente-rectificador para un cable según tipo de diodo utilizado

Se comparan los resultados para diodos rectificadores de silicio de propósito general (tipo 1N4007), de señal (tipo 1N4148) y de tipo Schottky (tipo BAT17). Los mejores

resultados se obtienen con diodos de señal y diodos de propósito general, como se observa en la Figura 8. En razón de la diferencia de tensiones logradas entre las tecnologías, se decide utilizar en los siguientes análisis el tipo de diodo 1N4148.

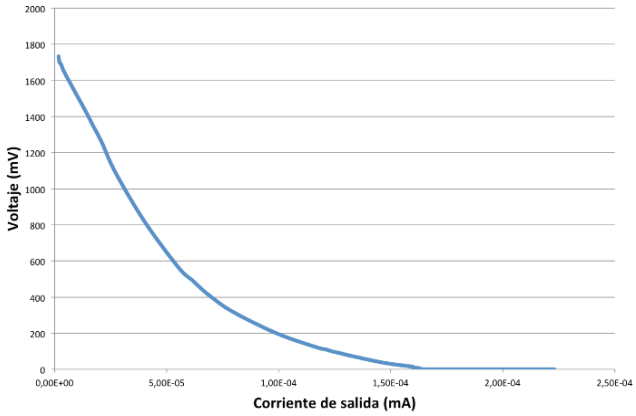


Figura 8: Tensión disponible según corriente consumida

8 Curvas de regulación y parámetros eléctricos del prototipo

A continuación se presenta el comportamiento de la tensión disponible en función del consumo de corriente que se solicite, así como el valor de potencia eléctrica disponible para su aplicación en un sistema autoalimentado. Para esto se evalúan las sucesivas simulaciones modificando la resistencia terminal de carga, para un circuito elemental con topología rectificadora en puente. Se obtuvo resultados para condición de circuito abierto y el detalle en torno al punto de máxima potencia con valores de resistencia terminal desde 1 GΩ hasta 100 MΩ (con decremento de valor constante de 200 MΩ), desde 100MΩ hasta 1MΩ (con decremento de valor constante de 20M Ω), desde 1MΩ hasta 100k Ω (con decrementos de valor constante de 100k Ω), desde 100k Ω hasta 10k Ω (con decremento de valor constante de 10k Ω) y para valores de 1 kΩ y 100 Ω.

En la Figura 8 se observa la curva de regulación que refleja la variación de la tensión continua disponible en el condensador acumulador, para diferentes consumos de corriente dados por las resistencias de carga utilizadas. En la Figura 9 se presenta la potencia que es capaz de entregar la estructura según la corriente consumida por la carga. En la Figura 10 se observa el comportamiento del consumo de potencia con respecto a la resistencia de carga.

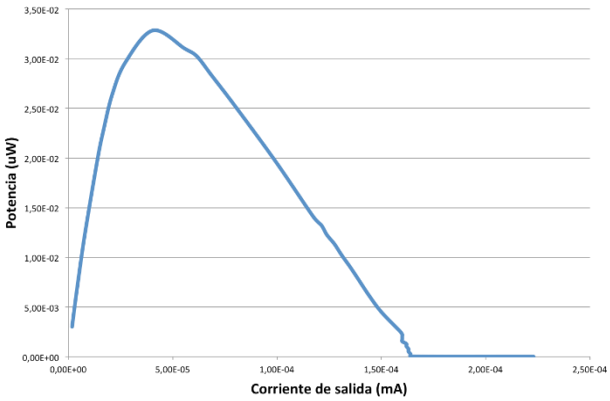


Figura 9: Potencia disponible según corriente consumida por la carga

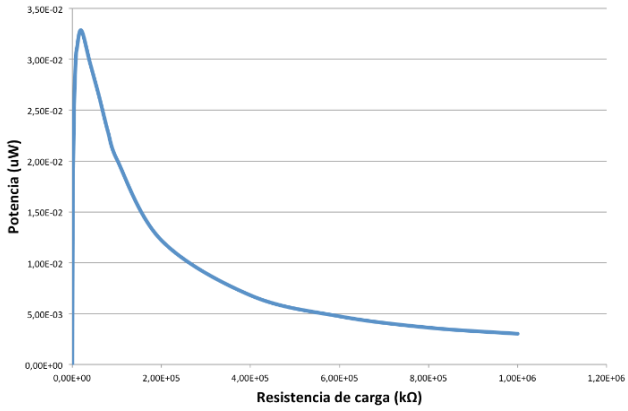


Figura 10: Potencia disponible según resistencia de carga

9 Análisis de resultados

El análisis realizado para el prototipo suelo-generador muestra que el punto optimo de funcionamiento se obtiene con un consumo de 40.5 nA en régimen continuo, para 811mV de tensión, liberando en estas condiciones una potencia de 33.8 nW, correspondiente a la simulación con R=20MΩ.

Estos resultados verifican a su vez el principio de máxima transferencia de potencia, que se presenta cuando la resistencia de carga (20MΩ) es igual a la resistencia equivalente de salida del circuito que se ha modelado, conformado por el cable piezoeléctrico, la estructura rectificadora y el capacitor de acumulación de carga.

Los valores bajos de potencia máxima utilizable son consecuencia del problema práctico enfrentado, porque se trabaja con una fuente de energía en forma de tensión eléctrica continua y con una alta resistencia equivalente de salida (20MΩ), que depende de la frecuencia del trafico y las amplitudes pico a pico que pueda generar.

10 Aplicaciones del prototipo suelo generador

Si bien los análisis realizados sobre los parámetros eléctricos a partir de las curvas de regulación del prototipo no fueron los mejores en cuanto a potencia generada, los últimos desarrollos en la electrónica permiten su aprovechamiento. Este es el caso del convertidor elevador dc-dc LTC3108 [16], que es ideal para la amplificación y gestión de energía proveniente de fuentes de tensión extremadamente bajas, tales como TEG (generadores termoeléctricos), pequeñas celdas solares y generadores piezoeléctricos. Este convertidor funciona con voltajes de entrada de 20 mV. En la Figura 11 se observa una aplicación del circuito LTC3108 y el prototipo suelo generador.

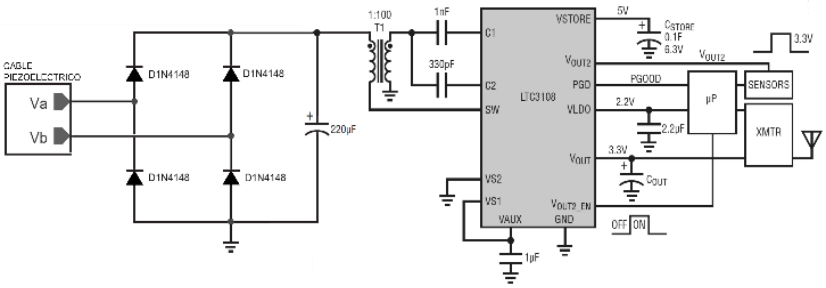


Figura 11: Diagrama de conexiones para una aplicación de adquisición y transmisión inalámbrica de datos desde un lugar remoto utilizando el LTC3108 y el prototipo suelo generador

11 Conclusiones

El prototipo de pruebas construido cuenta con un diseño y una estructura que facilitan la transmisión de la presión de rodadura [5, 9] ejercida por las distintas motocicletas (trafico real), obteniendo altos picos en las señales eléctricas generadas.

Tener un modelo eléctrico del cable piezoeléctrico facilitará futuros estudios, porque permite trabajar con la cantidad de cables que soporte el simulador, y de esta forma se ahorran costos significativos por construcción de prototipos adicionales.

La mejor configuración electrónica fue la entregada por los diodos 1N4148. Cabe resaltar que las pruebas se realizaron sobre un único cable piezoeléctrico, por lo que si se hicieran pruebas adicionales con varios cables en diferentes configuraciones, las pequeñas pérdidas presentadas en este análisis se verían multiplicadas por la cantidad de cables a utilizarse.

En el diseño electrónico del prototipo ayudó la herramienta de simulación PSpice [15], debido a que se redujeron los tiempos de ejecución y los costos del proyecto, razones que hacen aplicable este tipo de herramientas en los actuales procesos de diseño.

Durante las simulaciones se observó que los modelos y la cantidad de componentes electrónicos afectan los tiempos de simulación, haciendo necesario equipos de cómputo con buenas características en hardware y software.

Referencias

[1] Piezo materiales (2007). Piezoelectric Materials. Online [May 2014].

[2] Cook, K. et al. (2008). Piezoelectric energy harvesting: A green and clean alternative for sustained power production. Technology & Society 28(6), pp. 496-509.

[3] Berger, M. (2010). Nanotechnology recycles environmental energy waste into hydrogen fuel. Online [Jun 2014].

[4] Beeby, S. et al. (2007). A micro electromagnetic generator for vibration energy harvesting. Journal of Micromechanics and Microengineering 17, pp. 1257-1265.

[5] Chapa, J. (2008). Energy-generating floors to power Tokyo subways. Online [May 2014].

[6] Zhu, M.; Worthington, E. & Njuguna, J. (2009). Analyses of power output of piezoelectric energy-harvesting devices directly connected to a load resistor using a coupled piezoelectric-circuit finite element method. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 56(7), pp. 1309–1317.

[7] Platt, S. et al. (2005). The use of piezoelectric ceramics for electric power generation within orthopedic implants. IEEE-ASME Transactions on Mechatronics 10(4), pp. 455-461.

[8] Shenck, N. & Paradiso, J. (2001). Energy scavenging with shoe-mounted piezoelectrics. IEEE Micro 21(3), pp. 30-42.

[9] Vázquez, M.; Jiménez, F. & Frutos, J. (2011). Banco de ensayos para materiales piezoeléctricos en aplicaciones viales. Boletín de la Sociedad Española de Cerámica y Vidrio 50(2), pp. 65-72.

[10] Vázquez, M.; Jiménez, J. & Frutos, J. (). Modeling piezoelectric harvesting materials in road traffic applications. Proceedings 2nd International Conference on Mathematical, pp. 1-6. Puerto de la Cruz, España.

[11] Amato, I. (1989). Education: Piezo: Tough plastic with a sensitive side. Los Angeles Times, December 04.

[12] Measurement Specialties (2012). Piezo Sensor - Piezo Spiral Wrapped Coaxial Cable. Online [Feb. 2014].

[13] Gonzáles, A. (2009). Materiales Piezoelectricos y Efecto Piezoeléctrico. Cátedra de Tecnología Electrónica, Universidad tecnológica Nacional. Online [Mar 2014].

[14] Alexander, C. & Sadiku, M. (2006). Fundamentos de circuitos eléctricos. México: McGraw-Hill.

[15] Hoyo, J. (2000). Notas Tutorial Cadence OrCAD rel. 9.2. Mexico: Sedinet.

[16] Linear Technology (2013). LTC3108-Ultralow Voltage step-up converter and power manager. Online [May 2014].

Redes inalámbricas de sensores aplicadas al monitoreo de integridad estructural

Gustavo Meneses
Universidad de San Buenaventura Medellín – Colombia

1 Introducción

Las obras de infraestructura tales como puentes, edificios, túneles, carreteras, entre otras, son elementos importantes para el desarrollo rural y urbano en todos los países, y por aspectos estratégicos en cuanto a la seguridad de las personas, el comercio, la prestación de servicios públicos, las redes de transporte, y otros tantos. El uso permanente en el tiempo, los esfuerzos ante cargas estáticas y dinámicas, el desgaste natural de los materiales y la influencia de los cambios en el entorno, las condiciones climáticas y los fenómenos naturales influyen en su vida útil, llevándolas en ocasiones a situaciones extremas de colapsos y fallos que pueden tener consecuencias trágicas. Adicionalmente, errores en el diseño pueden tener también malos resultados [1].

En la última década dos campos de aplicación de la ingeniería, relacionados con la telemetría y la instrumentación electrónica, han tenido un desarrollo considerable: las redes inalámbricas de sensores y el monitoreo de integridad estructural (Structural Health Monitoring SHM) [2]. Las primeras se benefician de diferentes estándares de comunicación, principalmente de la familia IEEE 802.XX y de los avances tecnológicos en áreas como los sistemas microelectromecánicos (Microelectromechanical Systems MEMS), la recolección de energía (Energy Harvesting-Scavenging) y los microcontroladores de altas prestaciones de procesamiento con bajo consumo de energía [3]. Paralelamente, los sistemas de monitoreo de integridad estructural aprovechan las oportunidades que ofrece la gama de sensores y técnicas de análisis y procesamiento de señales disponibles, para entregar información que permita determinar acciones preventivas o correctivas derivadas del estado físico de las infraestructuras [2, 4].

2 Redes inalámbricas de sensores

Estas redes aprovechan las características de los llamados sensores inteligentes, y potencian sus posibilidades a partir de la integración de prestaciones de trabajo en red. Las comunicaciones inalámbricas aumentan la flexibilidad y la escalabilidad de los grupos de sensores que se despliegan sobre un área determinada para captar variables de interés. En su estructura genérica un nodo sensor está constituido por un bloque de cómputo o unidad central de procesamiento (CPU), una etapa de energía/alimentación, un elemento sensor, elementos de memoria RAM/ROM/EEPROM y un transceptor. Adicionalmente, se pueden tener bloques de captación de energía, hojas de datos electrónicas o sistemas micro-electro-mecánicos [3], como se observa en la Figura 1.

2.1 Estándares de comunicación para redes inalámbricas de sensores

En su forma tradicional las redes inalámbricas de sensores se basan principalmente en estándares de comunicación de la familia IEEE 802.XX, y se encuentran otros formulados específicamente para tal fin, como el IEEE 802.15.4, que define todo lo relacionado con las redes inalámbricas de área personal de bajo tráfico (Low Rate Wireless Personal Area Networks LR-WPAN). Tal vez Zigbee sea la versión más conocida para la implementación práctica del estándar IEEE 802.15.4, sin embargo, existen otras

como Xbee de Digi y MiWi de Microchip [5]. La operación en la banda de 2.4 GHz es la más común, con un rango aproximado de un centenar de metros y una tasa de bits de 250 kbps sobre 15 canales posibles.

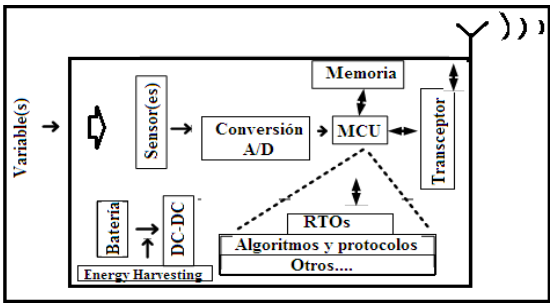


Figura 1: Nodo sensor inalámbrico

2.2 El protocolo MiWi de Microchip

Las principales características del protocolo MiWi son la facilidad de implementación sobre microcontroladores de 8, 16 o 32 bits y su aspecto liviano, en comparación con Zigbee (Tabla 1) en cuanto al consumo de memoria de programa [5].

Tabla 1. Memoria de programa MiWi vs Zigbee [6]

		MiWi	Zigbee
Memoria de programa (Kbyte)	Coordinador	<16kbyte	37-96 kbyte
	Router	<16kbyte	30-64 kbyte
	Dispositivo Terminal	2-8 kbyte	18-40 kbyte

Además de los transceptores, las comunicaciones bajo MiWi se sustentan sobre dos bases funcionales: una formulación llamada MiApp y otra denominada MiMAC, que rigen los aspectos de la capa de aplicación y la capa física del protocolo respectivamente [7, 8]. Adicionalmente, el protocolo MiWi es *royaltie-free* y a diferencia de Zigbee el número máximo de nodos de las redes que soporta es menor, 1024 nodos, que suele ser suficiente para la mayoría de aplicaciones.

A su vez, el protocolo MiWi ofrece variantes MiWi P2P y MiWi Pro, que poseen características definidas que varían en cuanto al número máximo de nodos, las topologías cubiertas y aspectos como el enrutamiento, entre otros. MiWi P2P [9] cubre aplicaciones que requieran redes con pocos nodos y baja complejidad, menos de 127 nodos, mientras que MiWi pro puede albergar más de 8.000 nodos. En el caso específico del protocolo MiWi P2P, elegido para la arquitectura propuesta en este trabajo, opera en topologías estrella y *peer-to-peer* (Figura 2) con redes de un salto, de modo que el rango está determinado por el alcance del radio transceptor. Al igual que en IEEE 802.15.4 se trabaja con dispositivos de dos categorías: de funciones completas (Full-Function Devices FFD) y de funciones reducidas (Reduced Function Devices RFD), que se reflejan en roles de coordinadores de red y dispositivos terminales [5, 9].

3 Monitoreo de integridad estructural

La integridad estructural de las obras civiles u otras obras de infraestructura basadas en concreto, acero y demás materiales, es un ítem importante para ingenieros civiles, arquitectos, geólogos y otros profesionales. La información captada por redes de sensores cableadas e inalámbricas provee elementos de análisis y decisión que pueden marcar la diferencia sobre la seguridad y la vida de muchas personas. Adicionalmente, el monitoreo

continuo permite optimizar las labores de mantenimiento e incluso realizar predicciones sobre la respuesta de las obras ante eventos que supongan un riesgo y que pueda limitar su operación y vida útil [10].

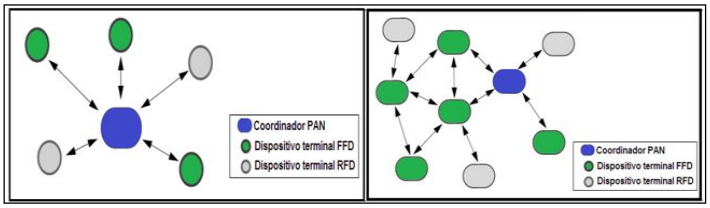


Figura 2: Topologías estrella y *peer-to-peer* bajo el protocolo MiWi P2P

Actualmente se encuentran implementaciones de monitoreo estructural en diferentes ciudades de Europa, Estados Unidos y Japón, que se ocupan de monitorear el comportamiento de los parámetros estructurales de puentes, túneles, carreteras, edificios y otras obras [11]. Igualmente, se encuentran soluciones de monitoreo de integridad estructural en edificios y monumentos históricos, lo que evidencia el dinamismo y crecimiento en este campo de aplicación, que se deriva de especialidades como la telemetría y la instrumentación electrónica y geológica.

3.1 Adquisición de datos en aplicaciones de monitoreo de integridad estructural

Los datos recolectados por los sensores desplegados en el campo usualmente deben ser sometidos a procesamiento y análisis en dispositivos como computadores, o almacenados en bases de datos para su posterior procesamiento. Dependiendo de la complejidad de la estructura monitoreada, el número de sensores y de variables bajo supervisión, se puede tener escenarios de diferente complejidad que plantean retos especiales en cuanto al proceso de adquisición de datos. Si bien es posible lo que se denomina procesamiento en el mismo nodo (*on-node processing*), usualmente la información de mayor valor se deriva de técnicas como la fusión de datos y de sensores, que requieren un nivel de procesamiento, almacenamiento y transmisión superior al que ofrece un nodo de red.

3.2 Sensores y variables

En la Tabla 2 se muestra algunas de las variables captadas y algunos de los sensores comúnmente utilizados en aplicaciones de monitoreo de integridad estructural.

Tabla 2: Sensores y variables de las aplicaciones de monitoreo de integridad estructural [12]

Sensores	Variables
Inclinómetros	Inclinación
Acelerómetros	Desplazamiento
Galgas extensométricas	Socavación
Sensores de humedad	Fracturas
Sensores de temperatura	Ruido
Piezómetros	Vibración
	Filtraciones

La utilización de acelerómetros basados en diferentes principios y tecnologías como sistemas micro-electro-mecánicos y fibras ópticas [13] es recurrente. Además, la ubicación, bien sea en ambientes interiores, exteriores, enterrados o incrustados sobre muros u otro tipo de superficies también determina el tipo de sensor a utilizar. Por eso es que para una misma variable se pueden encontrar diferentes sensores, que se emplean dependiendo del entorno de la aplicación. Otros aspectos, tales como el tipo de respuesta,

la resolución e incluso el costo, pueden determinar la elección de uno u otro sensor. La densidad de sensores por unidad de área es un aspecto importante a tener en cuenta, porque está ligada directamente con la complejidad que puede representar el manejo de las comunicaciones. Es importante recordar que el estado de algunas variables, especialmente las que revisten complejidad, se determina a partir de los datos recolectados por varios tipos de sensores, por lo que la relación no siempre es uno a uno.

3.3 Comunicaciones

Las redes tradicionales de sensores de tipo cableado no están excluidas de las aplicaciones de monitoreo de integridad estructural, el uso de formatos de transmisión tradicionales y buses de campo es totalmente válido. Sin embargo, en muchas situaciones y por cuestiones de flexibilidad, escalabilidad y costos, es mejor adoptar esquemas inalámbricos de comunicación. Este trabajo se centra en una propuesta de arquitectura mixta para cubrir las comunicaciones en aplicaciones de monitoreo de integridad estructural en entornos urbanos. Se propone un despliegue de red inalámbrica de sensores en el campo, soportada por nodos sensores y un coordinador de red que, a su vez, se apoya en *gateways* que proveen conectividad hacia redes de infraestructura cableadas y no-cableadas basadas en los diferentes estándares de la familia IEEE 802.XX, y cubriendo el rango de las redes de área local y metropolitano. Se contempla el uso de redundancia de comunicaciones apoyado en comunicaciones móviles soportadas en infraestructuras tradicionales como GSM.

4 La arquitectura propuesta

Se presenta una propuesta de arquitectura de comunicaciones para aplicaciones de monitoreo de integridad estructural en entornos urbanos en Colombia, que se basa en principios de operación similares a los de la plataforma Dexternet [14], e igualmente en las modificaciones y adaptaciones realizadas para la plataforma EnViBo [15]. Con esta ultima plataforma se realizaron pruebas de desempeño que mostraron buenos resultados, y a partir de la experiencia se considera que este funcionamiento puede ser replicado en diferentes tipos de aplicaciones de monitoreo con redes inalámbricas de sensores.

En el nivel inferior de la arquitectura propuesta se encuentra un conjunto de nodos sensores, agrupados en una red inalámbrica que adopta el estándar IEEE 802.15.4, específicamente el protocolo MiWi P2P, para cubrir la transmisión de los datos captados por los sensores desplegados sobre las obras de infraestructura monitoreadas. Luego se tiene un nivel intermedio que tiene al coordinador de red. En este nivel es posible visualizar los datos en una interfaz que se ejecuta en un computador y que recibe los datos recolectados por el coordinador vía serial/USB, bien sea de manera cableada o a través de RFCOMM utilizando un *gateway* Bluetooth. Adicionalmente, existe la posibilidad de tener en el coordinador una salida vía WiFi hacia otras redes o dispositivos. En el nivel superior se tienen opciones de conectividad a redes que operan sobre infraestructuras consolidadas mediante módulos Ethernet, que permiten interacción en Itranets e Internet, o a través de módulos GSM con redes de telefonía móvil. La arquitectura propuesta se observa en la Figura 3.

Mediante esta arquitectura se puede tener visualización de los datos recolectados por los nodos sensores en computadores, teléfonos inteligentes (*smarpthones*) o *tablets*. La utilización de aplicaciones desarrolladas para sistemas operativos, tales como Android o de lenguajes de programación para páginas web, permite esta funcionalidad. Igualmente, es posible alimentar bases de datos y aprovechar funcionalidades como correos

electrónicos y mensajes de texto para enviar alertas o mensajes programados periódicamente. La integración de las diferentes capas permite el tránsito de los datos por redes de área personal, redes de área local y redes de área amplia y metropolitana.

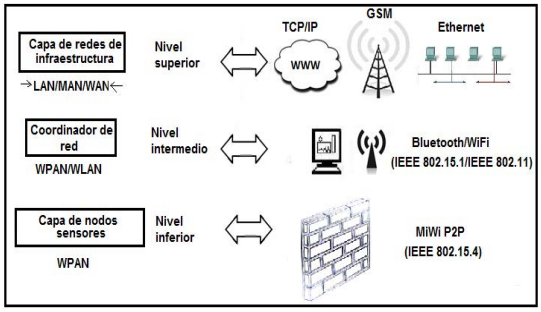


Figura 3: Arquitectura de tres niveles para integración de redes inalámbricas de sensores en aplicaciones de monitoreo de integridad estructural.

Como se describe más adelante, este tipo de arquitectura podría adaptarse, por ejemplo, a las obras consideradas en estado crítico o de riesgo potencial en la ciudad de Medellín, y articularse con sistemas gubernamentales de alerta temprana y gestión de esfuerzos y recursos de prevención y control, como el Sistema de Alerta Temprana de Medellín y del Valle de Aburra SIATA [16]. De esta forma podrían coordinarse mejor las acciones para manejar situaciones que puedan poner en peligro la vida e integridad de las personas que habitan en su área de influencia.

4.1 Coordinador de red y nodos sensores

La Figura 4 muestra los bloques básicos que tiene un nodo en el nivel inferior de la arquitectura propuesta. Los bloques punteados son opcionales y se pueden agregar dependiendo de las características y el diseño de la aplicación de monitoreo de integridad estructural. Tanto en los nodos sensores como en el coordinador de red existe la opción de implementar almacenamiento local de datos, por medio de EEPROM o medios extraíbles, y de implementar recolección de energía con módulos solares fotovoltaicos. En los nodos se puede agregar un módulo Bluetooth de bajo consumo de energía, como *gateway* para establecer conectividad con dispositivos que soporten el estándar IEEE 802.15.1 (*smartphones*, *tablets* o computadores). También puede adicionarse un módulo WiFi, que es recomendao para el coordinador de red por cuestiones de tamaño, consumo de energía y capacidades de procesamiento, necesarias para garantizar la coexistencia de varios estándares de comunicación en una misma unidad de red.

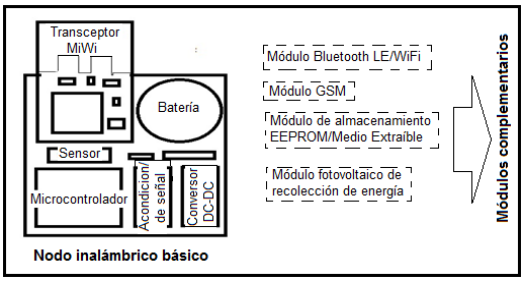


Figura 4: Nodo inalámbrico básico, módulos principales y complementarios

4.2 Conectividad hacia otras redes

Mediante la adición de módulos Ethernet o GSM se pueden proveer posibilidades de conexión hacia redes de telefonía móvil y redes basadas en TCP/IP. Esta conectividad

permite extender el rango de alcance para la transmisión de los datos, pasando de decenas y centenares de metros al orden de kilómetros. La redundancia de comunicaciones es una característica deseable e incluso necesaria en los sistemas de telemetría modernos, puesto que incrementa la fiabilidad de las soluciones implementadas y multiplica las posibilidades de visualización y recepción de notificaciones, incluso se puede tener opciones de telemando cuando sea necesario.

El rango de alcance, el alto grado de ubicuidad en ambientes urbanos y el grado de madurez y estabilidad de servicios como internet y la telefonía móvil, facilitan la disponibilidad de tránsito de datos de medición, alarmas y notificaciones por estas redes, sin que represente un incremento significativo en los costos de la solución implementada. En todos los casos no es necesario llegar al rango de las redes amplias y metropolitanas, porque en algunas situaciones la cobertura a nivel de área personal y local es suficiente para cubrir los requerimientos de monitoreo.

5 Prototipos de nodo sensor y coordinador para una aplicación de monitoreo de integridad estructural

Se presentan los diseños de dos prototipos de nodos: un nodo sensor genérico y un nodo coordinador de red, que utilizan el protocolo MiWi P2P como base de sus comunicaciones. Los datos aportados por los nodos sensores, ubicados sobre los muros u otras superficies de las obras de infraestructura, sirven para determinar desplazamientos que pueden indicar fallas estructurales debidas a hundimientos, socavación de las fundaciones, movimientos de tierra, sismos, entre otros. Con el establecimiento de valores de referencia con rangos de variabilidad preestablecidos como normales, se puede establecer alarmas o puntos de alerta para notificar al personal encargado: ingenieros, arquitectos, organismos de atención y prevención de desastres, o similares.

5.1 Nodo sensor basado en inclinómetro biaxial

En la Figura 5 se muestra un prototipo desarrollado para monitorear inclinación en los ejes *x,y*, que se basa en un inclinómetro biaxial SCA121T-D03 con un rango de medición de +/- 90°. El nodo sensor utiliza un microcontrolador de 8 bits (PIC18F4620) [17] y un transceptor MiWi MRF24J40MA. La unidad se alimenta con baterías recargables de litio-ion de 3.7 voltios, que alimentan todo el circuito (Figura 6). Periódicamente se transmiten las lecturas de la inclinación en los dos ejes al nodo coordinador, etiquetando los datos para facilitar su posterior procesamiento en una interfaz de visualización desarrollada como un instrumento virtual de Labview.

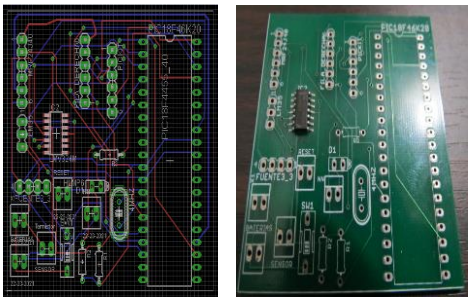


Figura 5: Diseño (izquierda) y tarjeta de circuito impreso de nodo sensor genérico (derecha)

Utilizando el modo *sleep* del microcontrolador y cesando la actividad del transceptor MiWi, en los periodos que no se realizan transmisiones el nodo se pone en modo de consumo reducido.

Figura 6: Prototipo de nodo sensor inalámbrico basado en inclinómetro biaxial para fijación en muro

5.2 Prototipo de nodo coordinador de red de área personal

El coordinador de la red de área personal se implementó utilizando un microcontrolador de 16 bits (PIC24FJ128GA010) [18] y un transceptor MiWi MRF24J40MB como elementos centrales sobre una tarjeta de desarrollo Explorer 16 [19]. Como se observa en la Figura 7 se ha integrado un *gateway* Bluetooth y una tarjeta SD al nodo coordinador, que permiten hacer almacenamiento local de datos. La transferencia de los datos hacia la interfaz de visualización, que se ejecuta con un computador con Labview, puede hacerse vía serial/USB o RFComm con Bluetooth. También es posible transferir datos vía RFComm a consolas de puertos seriales virtuales que se ejecuten en tabletas o teléfonos inteligentes, o de manera más elaborada en aplicaciones desarrolladas para dichos dispositivos. El *firmware* que se ejecuta en los dos prototipos mostrados se basa en los recursos y herramientas de Microchip denominados MPLAB IDE, Application Libraries y MiWi Stack, que han sido utilizados, adaptados, modificados e intervenidos según los requerimientos puntuales de la aplicación que se desea ilustrar.

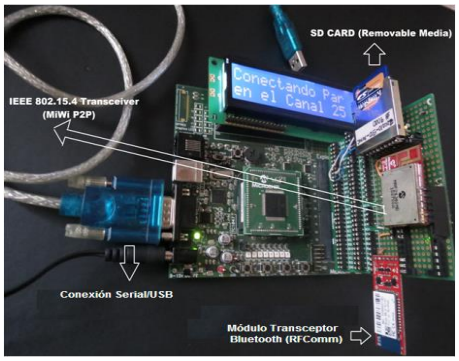


Figura 7: Prototipo de nodo coordinador de red implementado sobre la tarjeta de desarrollo Explorer16 de Microchip [17]

La utilización de *gateways* permite integrar las comunicaciones de la aplicación de monitoreo en el nivel intermedio de la arquitectura con otras redes de área personal bajo el estándar IEEE 802.15.1, y con redes WLAN bajo el estándar IEEE 802.11. Esta integración permite el reenvío de los datos recogidos por los sensores desplegados *in situ* y su visualización a través de dispositivos que incorporan este tipo de recepción/transmisión bajo estos estándares. Desde la capa intermedia y a través de los dispositivos antes mencionados, se puede impulsar el tránsito de los datos por otras redes a través del enlace, por ejemplo, de servicios como Internet, dado que en algunos casos estará habilitada esa opción de navegación.

Adicionalmente, si se incorpora un dispositivo como un modem GSM, se puede reenviar datos por redes de telefonía móvil utilizando servicios como el de mensajes de texto (SMS). Este tipo de interfaces permite acceder a la capa superior de la arquitectura con posibilidades de cobertura a nivel metropolitano. De manera complementaria está la opción de acceder a puertos RJ45 habilitados para TCP/IP, que permite igualmente la retransmisión de datos a destinos remotos en redes locales cableadas o hacia direcciones externas a través de servidores.

En la Figura 8 se describen tres opciones de módulos que pueden ser integrados al nodo coordinador, como *gateways* en el nivel intermedio de la arquitectura. Todos presentan características que un momento dado pueden ser consideradas como ventajas o desventajas, al considerar aspectos como el alcance y la facilidad de integración al *firmware* que se ejecuta en dicho nodo.

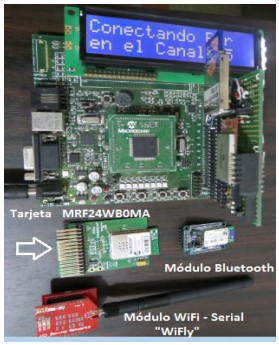


Figura 8: Prototipo de nodo coordinador de red y opciones de módulos para integración de gateway Bluetooth o WiFi

La Tabla 3 muestra un comparativo entre las tres opciones de *gateway*. Se muestran dos módulos WiFi, uno controlado vía SPI y otro serial, y un módulo Bluetooth.

Tabla 3: Comparación de prestaciones de diferentes opciones de dispositivos *gateways*

Dispositivo	Ventajas	Desventajas
Módulo Bluetooth	Fácil integración	Corto alcance
Tarjeta MRF24WB0MA	Mayor Alcance	Difícil Integración
Módulo WiFly RN-171	Mayor Alcance	Dificultad de integración media

El módulo MRF24WB0MA que alberga el transceptor IEEE 802.11 presenta una complejidad especial de integración, debida a la cantidad de archivos involucrados en el proyecto de MPLAB que lo hace operativo. En la práctica esto resulta en la integración de dos proyectos en uno, abarcando una amplia cantidad de archivos con extensión c y h, además de un número sobsiderable de modificaciones y adaptaciones en el código.

Como se mencionó antes, también está la opción del módulo Ethernet ENC28J60, pero tiene la desventaja de requerir infraestructura cableada, que no siempre está disponible en todos los sitios durante el despliegue de las redes de medición. En las zonas en las que no es posible tener cobertura WiFi o Ethernet para enlaces en el nivel intermedio o superior de la arquitectura, se puede considerar la opción de incluir radio-módems que operen en bandas ISM, como la de 900MHz, para extender el rango de cobertura de las comunicaciones en la aplicación de monitoreo de integridad estructural.

5.3 Interfaz gráfica de usuario

Después de recibir las lecturas de los ángulos de inclinación desde el nodo sensor, se procede a su visualización, registro, reporte, almacenamiento, procesamiento,

retransmisión y análisis. Para esto se puede utilizar Labview, debido a que posee las herramientas que permiten realizar estas tareas. La Figura 9 se muestra la ubicación del sensor y en la 10 el registro visual de una prueba de medición con el inclinómetro de muro, en la que se introdujeron vibraciones y cambios de ángulo en sentido horario y anti-horario, en periodos espaciados de cinco segundos durante una sesión de un minuto.

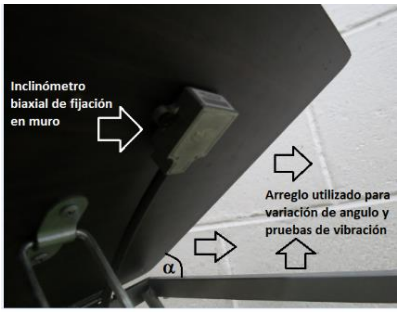


Figura 9: Disposición de elementos para la realización de pruebas de variación angular y de detección de perturbaciones debidas a vibración

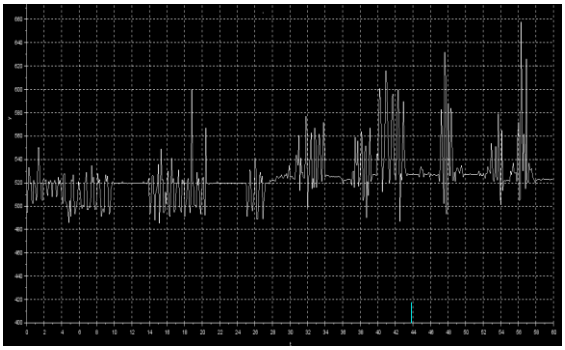


Figura 10: Registro de las variaciones de inclinómetro biaxial generado desde un *waveform chart* de Labview

5.4 La plataforma Ubidots

Ubidots es una plataforma que nace en el entorno del internet de las cosas, apalancada como una *start-up* y concebida desde la región antioqueña. Actualmente, esta plataforma permite monitorear a distancia diferentes objetos, fenómenos y variables utilizando tarjetas de desarrollo Microchip y tarjetas *open hardware* Arduino y Raspberry Pi, por mencionar algunas [20]. Labview también se puede integrar a Ubidots mediante instrumentos virtuales, que se apoyan en las funciones *get* y *post* para recuperar y publicar datos desde una ubicación en la web, como se observa en la Figura 11.

A partir del conocimiento y la experticia de personas especializadas en el análisis y seguimiento de los parámetros que revelan la condición de las diferentes obras de infraestructura, se pueden adoptar diferentes esquemas de procesamiento de datos y de señales, que arrojaran información de valor para emitir juicios sobre la integridad de una obra especifica. Existen técnicas avanzadas y herramientas de modelación, predicción e inferencia que pueden hacer uso del volumen de datos que entrega una red de sensores desplegada sobre una obra. En la Figura 12 se observa el contexto completo de desarrollo de una aplicación de monitoreo de integridad estructural, que integra una red inalámbrica de sensores desplegada sobre una obra bajo monitoreo, con un instrumento virtual desarrollado en Labview con la plataforma Ubidots y el Sistema de Alerta Temprana SIATA.

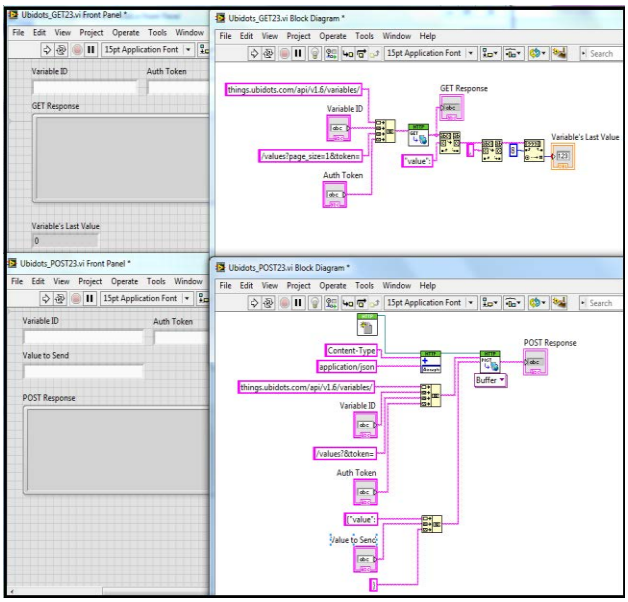


Figura 11: Instrumentos virtuales para la captura y publicación de variables en páginas web a través de las funciones *get* y *post*

Figura 12: Escenario global de una aplicación de monitoreo de integridad estructural en el área del Valle de Aburra

La suma de recursos y capacidades planteada en este trabajo puede resultar en una sinergia interesante, que optimizaría los resultados en la prevención, control y atención de desastres relacionados con las obras de infraestructura de Medellín y del Valle de Aburra.

6 Análisis de resultados

Con base en las pruebas realizadas con dos nodos, un nodo sensor y un coordinador de red de área personal, y bajo el protocolo MiWi P2P, se realizaron ajustes sobre un instrumento virtual (VI) que sirve para registrar las variaciones de inclinación provocadas. Algunas oscilaciones derivadas de aspectos de sincronización entre la transferencia de datos entre el nodo coordinador y el VI de Labview pueden ser observadas en la Figura 13 y han sido corregidas. Adicionalmente, se han realizado pruebas de operación continua del nodo sensor, en las que se obtuvieron funcionamientos por periodos de más de un día al alimentar el nodo con un par de baterías de Li-ion de 3.7 voltios conectadas en serie. En un día se pueden transmitir entre 1.200 y 1.400 paquetes con transmisiones periódicas separadas por aproximadamente un minuto y teniendo en cuenta algunos eventos de desconexión-reconexión de los nodos y también algunos paquetes pérdidos o defectuosos.

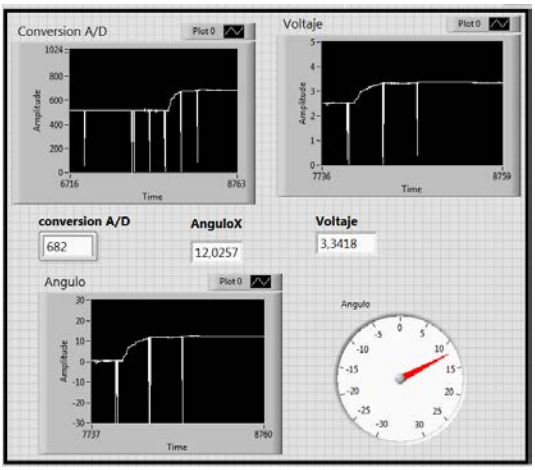


Figura 13: Registro de variaciones de inclinación (angulares) sobre el eje X de un inclinómetro biaxial en un instrumento virtual de Labview

En pruebas complementarias realizadas con baterías recargables tradiciones de plomo-ácido de 12 voltios, se alcanzaron periodos de transmisión continua por más de una semana de operación y 10.000 paquetes transmitidos. Dichas pruebas se han realizado en interiores, pero se espera variabilidad en los resultados cuando se efectúen pruebas en ambientes exteriores debida a la influencia de variables como la temperatura sobre el desempeño de las baterías.

Las baterías recargables de NiMH también se probaron y arrojaron resultados comparables a los obtenidos con las de de Li-ion, la variabilidad observada depende del estado de carga de las baterías al comenzar las pruebas, lo cual se considera normal al igual que la influencia de la cantidad de ciclos de carga-descarga que han afrontado. Más detalles acerca de cómo se han realizado estas pruebas y la descripción de algunos resultados adicionales obtenidos se pueden consultar en [21].

Actualmente se lleva a cabo un proyecto piloto para comenzar tareas de monitoreo de integridad sobre la infraestructura de una empresa de la ciudad de Medellín. En este proyecto se intregarán redes inalámbricas de sensores en el rango de área personal utilizando MiWi P2P; igualmente se adoptará un esquema de comunicaciones similar al indicado por la plataforma propuesta en este trabajo. Los detalles de esta experiencia se compartirán una vez concluya el contrato de jecusión y se alcancen los objetivos establecidos en el mismo.

7 Conclusiones

Luego de aproximadamente una década de desarrollo continuo a gran escala, las redes inalámbricas de sensores, los estándares de comunicación y las tecnologías que las soportan, han alcanzado un grado de madurez que las hacen confiables y viables desde el punto de vista económico para integralas en aplicaciones de monitoreo de obras de infraestructura.

El monitoreo de integridad estructural se vislumbra como un campo de proyección que viene creciendo en diferentes partes del mundo y que muestra su importancia y utilidad en escenarios ligados a la vivienda, el transporte, los servicios públicos, la industria y la producción. Sin embargo, en el medio colombiano el desarrollo se puede calificar como incipiente, por lo que existe aún un vasto campo para el desarrollo de aplicaciones.

Las actuales soluciones de instrumentación y telemetría basadas en redes inalámbricas de sensores plantean ciertas ventajas en aplicaciones de monitoreo de integridad estructural, frente a la contraparte cableada debido a características que pueden favorecer una mayor autonomía, flexibilidad y escalabilidad.

Los dos prototipos presentados en este trabajo y la arquitectura propuesta pretenden abrir un camino que puede ser abordado desde diferentes ángulos y por diferentes profesionales, de acuerdo con sus necesidades y campos de acción. Es en todo caso una mirada inicial sobre una confluencia de dos especialidades de la ingeniería, que pueden reunirse en aplicaciones que generen valor agregado a los enfoques y soluciones existentes.

En este momento se trabaja sobre los prototipos mostrados con el fin de mejorarlos y para llegar a versiones más robustas, que cumplan con los requerimientos de trabajo a la intemperie y sobre periodos extendidos de tiempo, garantizando autonomía y fiabilidad acordes con las exigencias de un monitoreo continuo sobre obras de infraestructura, tales como muros de contención, columnas, juntas, cimientos y fundaciones en puentes y edificios.

Técnicas modernas como la fusión de datos y de sensores contribuyen a un análisis dinámico mas completo de la obra de infraestructura monitoreada. Igualmente, el análisis de modos de vibración y la incorporación de técnicas tipo radar proporcionan información acerca de la integridad de las obras, que pueden ser más completa que la entregada por los sensores que miden solamente un punto específico.

Referencias

[1] Universidad de Los Andes (2014). Concepto técnico en relación a las causas más probables del colapso del edificio space. Contrato No. 4600051633 de 2013, Alcaldía de Medellín.

[2] Yun, C. & Min, J. (2011). Smart sensing, monitoring, and damage detection for civil infrastructures. KSCE Journal of Civil Engineering 15(1), pp. 1-14.

[3] Fernández, R. et al. (2014). Redes inalámbricas de sensores: Teoría y aplicación práctica. España: Universidad de la Rioja.

[4] Marulanda, J.; Thomson, P. & Marulanda, J. (2000). Monitoreo de salud estructural. Ingeniería y Competitividad 2(2), pp. 40-46.

[5] Flowers, D. & Yang, Y. (2010). Microchip MiWi™ Wireless Networking Protocol Stack. Technical Report AN1066. Microchip Technology Inc.

[6] Iacono, L. (2011). Wireless Sensor Network Protocols. Online [Jun. 2015].

[7] Yang, Y. (2009). Microchip Wireless (MiWi™) Application Programming Interface-MiApp. Technical Report AN1284. Microchip Technology.

[8] Yang, Y. (2015). Microchip Wireless (MiWi™) Media Access Controller - MiMAC. Techical Report AN1283XX. Microchip Technology.

[9] Yang, Y. (2010). Microchip MiWi P2P Wireless Protocol. Techical Report AN1204. Microchip Technology

[10] Fraser, M. et al. (2010). Sensor network for structural health monitoring of a highway bridge. Journal of Computing In Civil Engineering 24(1), pp. 11-24.

[11] Lueker, M. et al (2010). Bridge scour monitoring technologies: development of evaluation and selection protocols for application on river bridges in Minnesota. Report No. Mn/DOT 2010-14. Minnesota Department of Transportation.

[12] Transportation Research Board (2008). Use of inclinometers for geotechnical instrumentation on transportation projects: State of the practice. Transportation Research Circular, Number E-C129.

[13] Tam, H. et al. (2008). Utilization of fiber optic Bragg grating sensing systems for health monitoring in railway applications. Proceedings Ninth International Symposium on Precision Engineering Measurement and Instrumentation, pp. 1824-1831. Changsha, China.

[14] Kuryloski, P. et al. (2009). DexterNet: An open platform for heterogeneous body sensor networks and its applications. Proceedings Sixth International Workshop on Wearable and Implantable Body Sensor Networks, pp. 92–97. Berkeley, USA.

[15] Meneses, G. & Lemos, J. (2014). EnViBo: Embedded network for vital sign and Biomedical Signal Monitoring. Proceedings IEEE Colombian Conference on Communications and Computing, pp. 1-6. Bogotá, Colombia.
[16] Bustamante, V.; Ceballos, L. & Castro, J. (2005). Sistema de alerta temprana ambiental: Glosario de términos. Alcaldía de Medellín.
[17] Microchip Technology (2008). PIC18F2525/2620/4525/4620 data sheet enhanced flash micro controllers with 10-Bit A/D and NanoWatt technology. Technical Report DS39626E. Microchip Technology.
[18] Microchip Technology (2009). PIC24FJ128GA010 Family Data Sheet 16-Bit Flash Microcontrollers. Technical Report DS39747F. Microchip Technology.
[19] Microchip Technology (2005). Explorer 16 Development Board User's Guide. Technical Report. DS50001589B. Microchip Technology.
[20] Peláez, A. (). Ubidots: Aplicaciones para el Internet de las Cosas. Online [Jun. 2014].
[21] Meneses, G. (2015). Achieving reduced power consumption of sensor nodes in wireless networks using MiWi protocol. Memorias del VII Congreso Iberoamericano de Telemática 2015, pp. 189-196. Popayán, Colombia.

Diseño de un sensor de permitividad dieléctrica relativa de un medio empleando una antena de microcinta con estructuras metamateriales

María Domínguez V.
Daniel Cataño O.
Erick Reyes V.
Instituto Tecnológico Metropolitano – Colombia

1 Introducción

La medición de la permitividad dieléctrica de los materiales ha despertado interés en los últimos años, porque se requiere en aplicaciones de diferentes sectores productivos, tales como medicina, farmacia, industria de alimentos, química e ingeniería civil, entre otros, y debido a que a través de la medición de este parámetro físico es posible realizar cálculos indirectos de otros parámetros, tales como humedad, temperatura y concentración. En el pasado se exploraron diversas técnicas para llevar a cabo la medición de la permitividad dieléctrica y la permeabilidad magnética de un medio, con el fin de realizar una caracterización de sus propiedades eléctricas y magnéticas en el rango de las microondas. Entre las técnicas más empleadas se destacan: por espacio libre [1], de cavidad resonante [2-5], de cable coaxial con línea abierta [4, 6] y la de línea de transmisión [7], y cada una de ellas presenta ventajas y desventajas según el tipo de material y el rango de frecuencia de estudio. Recientemente se han empleado sensores de microondas para la medición de las propiedades de los materiales, debido a que tienen bajo consumo de energía, son económicos, tienen alta precisión y permiten llevar a cabo un análisis en tiempo real, como se ha mostrado en estudios previos [8, 9].

Muchos de los sensores de microondas diseñados para medir la permitividad dieléctrica de materiales, pueden ofrecer información acerca de propiedades como temperatura, humedad, concentración, composición, densidad y otras, tal como lo han mostrado algunas investigaciones [1, 2, 8, 10]. Entre estas técnicas, la más utilizada para dichas mediciones es la que emplea resonadores, entre otras cosas, porque presentan un buen factor de calidad y porque son altamente sensibles a los cambios del medio que los rodea [11, 12]. Otro de los métodos electromagnéticos explorados consiste en la implementación de antenas de microcinta, que brindan flexibilidad en su diseño y permiten incluso realizar la construcción de sensores con tamaño reducido. Esto constituye una ventaja debido a que no se requieren mucho volumen de las muestras que se desean analizar, por lo que la técnica es favorable desde el punto de vista económico. Otra ventaja es su sencillez, lo que la hace fácil de usar. Usualmente, cuando se ha empleado esta técnica se ha usado línea de microcinta como dispositivo sensor, en la que el material bajo prueba se coloca sobre la microcinta generando una atenuación en la potencia de la antena [13]. Además, la técnica brinda la posibilidad de realizar mediciones espectrales cuando se usa un resonador de microcinta, donde el método de medición se basa en calcular el corrimiento del espectro al cambiar el medio que está en contacto con el sensor [2, 9, 14].

Sarabandi y Li [15] utilizaron un anillo resonador diseñado para operar a una frecuencia de resonancia de 1.25 GHz. Esta antena se empleó para medir la permitividad dieléctrica de diferentes muestras de arena, para evaluar los niveles de humedad que presentaban. Posteriormente, se reportó el análisis de las propiedades dieléctricas de algunos solventes líquidos mediante la implementación de una cavidad resonadora

grabada sobre una guía de onda integrada, demostrando las ventajas de esta tecnología en cuanto a tamaño y precisión [16]. Kobayashi y Nikawa [17] emplearon una cavidad cilíndrica resonante para medir la dependencia entre la temperatura y la permitividad dieléctrica de muestras de alcohol etílico; y Bogosanovich [4] empleó un sensor basado en una antena de microcinta para la caracterización de materiales homogéneos con permitividades relativas inferiores a 10, con una precisión de medida del 5%. En este caso no se tuvo en cuenta todos los modos posiblemente excitados en la antena, porque solamente se realizó el análisis para el modo TM_{01}. Adicionalmente, las estructuras de metamateriales basadas en anillos resonadores simétricos han sido empleadas previamente para determinar la permitividad de telas [11], demostrando que el método es fácil de implementar y que puede ser usado incluso en la caracterización de materiales sólidos. En algunas aplicaciones biológicas también han sido implementados sensores basados en antenas de microcinta, para determinar la variación en la concentración de glucosa y la detección de biomoléculas [8, 12].

En el presente trabajo se emplea por primera vez una antena de microcinta con una estructura metamaterial resonadora con forma rectangular, en combinación con el estudio del ancho de banda de una línea de transmisión grabada sobre el plano tierra. De esta manera se analiza la influencia de dos de las estructuras recomendadas para la medición de la permitividad dieléctrica, cuando se utilizan en conjunto. El análisis se llevó a cabo para medios cuya permitividad relativa se encuentra en el rango de 1 a 10. Adicionalmente, se analizó la influencia del grosor de la línea de transmisión grabada sobre el plano tierra y se comparó con una antena con un plano tierra completo, para evaluar la dependencia de la sensibilidad del sensor con los parámetros geométricos. Para llevar a cabo este análisis y para determinar los parámetros óptimos de diseño de la estructura, se empleó el método de elementos finitos vectorial. Para esto se analizó el comportamiento de tres configuraciones diferentes, cuando se emplean técnicas espectrales para el análisis de la permitividad dieléctrica relativa, es decir, se analizó el corrimiento en el pico de resonancia de la antena cuando se cambia el medio que la rodea.

2 Metodología

Para el estudio llevado a cabo en este trabajo se analizó la variación del ancho de la línea de transmisión grabada sobre el plano tierra de la antena con el fin de hallar las dimensiones optimas, específicamente cuando la antena se emplea como elemento sensor. En la siguiente fase del trabajo se realizó la simulación modelando un cambio de permitividad en el medio que rodea la antena de microcinta. La Figura 1(a) muestra la antena modelada y la Figura 1(b) muestra la antena discretizada por medio de un enmallado triangular, diseñado especialmente para la antena b y que cumple con el criterio de que cada elemento debe tener dimensiones de $\lambda/5$ para obtener resultados validos en el proceso de simulación. Todas las simulaciones se realizaron por medio del método de elementos finitos vectorial y utilizando el software de simulaciones electromagnéticas HFSS de Ansoft.

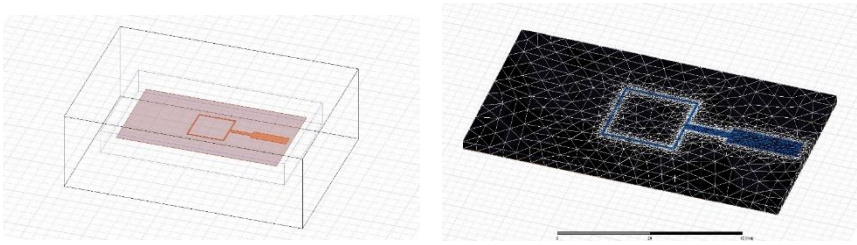


Figura 1: Antena para censado: (a) modelamiento de la antena; (b) discretizado de la antena

3 Diseño del sensor

En la Figura 2 se muestra un esquema 2D de las caras frontal y posterior de la antena con estructura de metamateriales. Se observa que un anillo resonador de cobre ha sido definido sobre un substrato dieléctrico, al que se ha hecho una abertura que posteriormente se encarga de que la estructura tenga un comportamiento metamaterial, y cuya frecuencia de resonancia depende de sus dimensiones [18]. Adicionalmente, sobre este mismo plano (parte superior del substrato) se modela una línea de alimentación con ajuste de impedancia, para garantizar que el dispositivo tenga una impedancia característica de 50 Ω. En la parte posterior de la antena se graba una línea de cobre de grosor Δ y longitud 41mm, para evaluar el comportamiento del sensor al variar el grosor de la línea, y que simultáneamente cumple la función de ser el plano tierra de la antena.

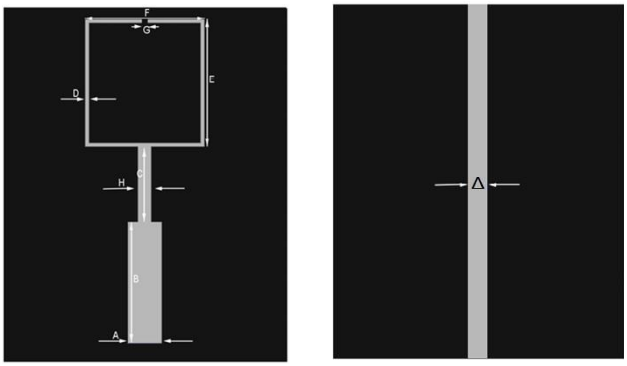


Figura 2: Esquema del sensor: vista frontal y vista posterior

Para el diseño de este sensor se empleó el FR4como substrato, un material típicamente empleado en la fabricación de circuitos impresos. Este substrato presenta un grosor de 1.6 mm y una estructura metálica de 35 µm a ambos lados. El ancho y alto del anillo resonador con estructura rectangular es 9.965 mm y 17 mm respectivamente, y posee un grosor igual al de la abertura de 0.5 mm. El resto de las dimensiones del sensor se resumen en la Tabla 1.

Tabla 1: Resumen de las dimensiones de la antena

Variable	Longitud [mm]	Descripción
A	4.852	Ancho alimentación acople
B	16.202	Largo alimentación acople
C	9.965	Largo alimentación
D	0.500	Grosor SRR
E	17.000	Largo SRR
F	17.000	Ancho SRR
G	0.500	Longitud y ancho de la apertura
H	1.885	Ancho alimentación

Eléctricamente, este tipo de estructuras posee un comportamiento similar al de un circuito resonador LC, que presenta una resonancia a una determinada frecuencia f_m cuando se somete a un campo magnético externo de forma perpendicular al plano de la superficie que va a medir. Induciendo así una corriente sobre el anillo resonador cuya dirección de circulación depende de la dirección del campo, y cuyas frecuencias de resonancia pueden ser determinadas a través de la ecuación (1)

$$f_m = \frac{1}{2\pi\pi\sqrt{(C_g + C_s)L_{eff}}} \qquad (1)$$

Donde L_{eff}, C_g y C_s son la inductancia efectiva, la capacitancia debida al gap y la capacitancia de superficie respectivamente, que dependen básicamente de la geometría y las dimensiones del anillo resonador empleado. Torun et al [12] demuestran que estos parámetros dependen de la permitividad efectiva, es decir, no solamente de la permitividad del material dieléctrico sobre el que se fabrica la antena, sino que además dependen de la permitividad del medio dieléctrico que la circunda. Por lo tanto, a través de la manipulación de la geometría o el medio dieléctrico donde está inmersa la antena es posible manipular su frecuencia de resonancia, y emplearla como sensor de permitividad a través de cambios de medio debidos a concentración, humedad o temperatura.

También se analizaron tres configuraciones en este trabajo, en la que se realizó la variación del ancho Δ del plano tierra para tres valores diferentes de 0.1 mm, 10 mm y 41 mm, con el fin de analizar la sensibilidad del sensor ante las alteraciones realizadas en el plano tierra y por ende alterar la frecuencia de resonancia del metamaterial. Para llevar a cabo el análisis del comportamiento de este sensor, se empleó el software HFSS, que emplea el método vectorial de elementos finitos. A través del software se analizó el comportamiento de los parámetros S de dicha estructura, para determinar la frecuencia de resonancia del sensor cuando es inmerso en diferentes materiales. Para ello se evaluó la respuesta de las pérdidas por retorno (s_{11}) para materiales que poseen permitividad dieléctrica relativa en el rango de 1 a 10.

4 Resultados y discusiones

En la Figura 3 se muestra el corrimiento espectral que se genera cuando la permitividad es alterada de 1 a 10 para la configuración con Δ=0,1 mm. Dando como resultado que dichos corrimientos se dan como consecuencia del cambio en la permitividad efectiva, cuando se altera el medio donde está inmerso el sensor. Se encontró que al incrementar la permitividad del medio, disminuyen significativamente las perdidas por retorno, lo que indica que al incrementarla se irá generando un mayor acople de la radiación entre la línea de transmisión y el resonador. Es de esperar que se presenten cambios en la potencia, lo cual experimentalmente permitiría usar técnicas espectrales o medidores de potencia para la construcción de un sensor basado en esta estructura.

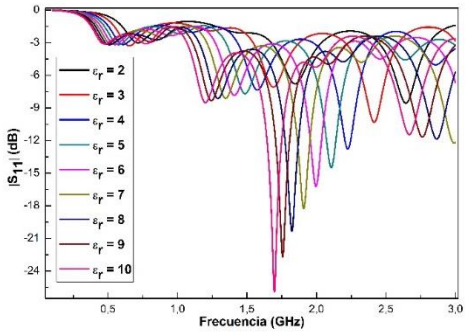


Figura 3: Resultados al emplear el sensor basado en antena de microcinta con estructuras metamateriales con Δ=0,1 mm

En la Figura 4 se resumen los corrimientos en la frecuencia de resonancia, cuando la permitividad relativa varía. Se encontró que presenta un comportamiento asintótico como consecuencia de que el sensor tiende a presentar una saturación en su medida, cuando los cambios inducidos son altos. En la ventana de frecuencia analizada se logró determinar que dependiendo de la frecuencia analizada cambia la sensibilidad, encontrando que para la configuración con Δ=0.1 mm, el sensor puede medir permitividad en el rango 1<ε<3,

con una sensibilidad de 301.14 MHz. Mientras que en el rango de 3< ε<10 disminuye a 95.15 MHz por cada unidad de permitividad relativa. Se puede observar que esta configuración es más atractiva a la hora de determinar permitividades menores a 3, es decir, que podría ser ideal para caracterizar concentraciones de diferentes gases en la atmósfera o en cavidades.

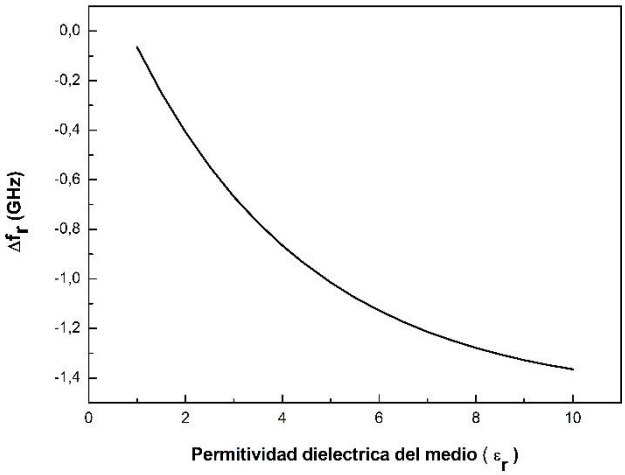


Figura 4: Análisis del corrimiento espectral generado cuando la permitividad del medio analizado varía entre 1 y 10

Posteriormente, se evaluó la influencia del ancho del plano tierra sobre la sensibilidad del sensor. Para esto se analizaron dos configuraciones extras con Δ= 10 mm y 41 mm respectivamente. Los resultados del corrimiento con las tres configuraciones se muestran en la Figura 5. Este análisis se realizó para encontrar la estructura más sensible, que permita tener un sensor para medir pequeñas variaciones en la permitividad relativa de materiales líquidos y gaseosos. Al analizar la configuración con Δ=10 mm y dividirla por tramos, para 1<ε<4.5 se tiene una sensibilidad de 332.01 MHz, mientras que en la región que comprende desde 4.5 a 10 la permitividad decrece hasta 144.6 MHz. Cuando se analizó la configuración con Δ=41 mm la sensibilidad fue aproximadamente de 300 MHz en todo el rango de estudio.

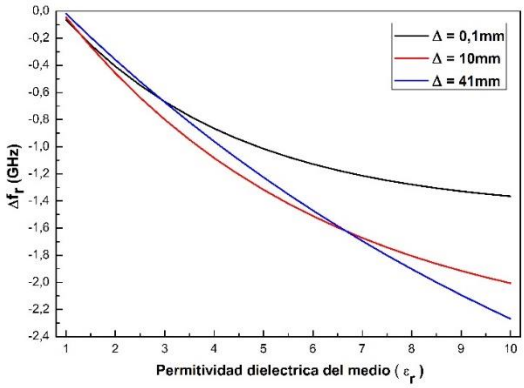


Figura 5: Comparación de resultados del corrimiento espectral obtenidos con las tres configuraciones

Con base en estos resultados se puede concluir que la configuración con mejores es la que posee un plano tierra completo (Δ=41 mm), porque permitiría medir pequeños cambios en la permitividad dieléctrica no solamente en gases sino en materiales sólidos, tales como papel y algunos líquidos con permitividad inferior a 10. Esta configuración

llama la atención porque en comparación con las otras dos configuraciones estudiadas, y a pesar de que su sensibilidad es inferior para rangos de permitividad relativa inferiores a 4.5, permite determinar con buena precisión pequeños cambios en un rango de permitividades más amplio sin llegar a saturarse, como lo hacen las configuraciones con Δ= 0,1 mm y Δ=10 mm.

5 Conclusiones

Se encontró una estructura que funciona como sensor basado en una antena de microcinta con un resonador rectangular, que permite determinar con precisión y sensibilidad la permitividad dieléctrica relativa en un rango de $1<\varepsilon<10$. El principio de funcionamiento de este sensor se basa en la alteración de la frecuencia de resonancia y depende fuertemente de los cambios en la permitividad efectiva, que se altera cuando la antena diseñada se sumerge en un medio con permitividad dieléctrica diferente al aire.

El método propuesto en este trabajo, además de ser altamente sensible, es fácil de implementar y económico, lo cual brinda soluciones a los problemas a la hora de determinar cambios en la permitividad dieléctrica como consecuencia de agentes externos e internos, tales como temperatura, humedad y concentración.

Como trabajo futuro se plantea la fabricación y caracterización del sensor diseñado, para verificar experimentalmente lo predicho en el modelo elaborado, a través del método vectorial de elementos finitos, y que puede ser implementado en la caracterización eléctrica de diferentes materiales, tales como alcoholes, glucosa, combustibles y algunos gases nobles.

Agradecimientos

Este trabajo fue financiado por el Instituto Tecnológico Metropolitano a través del proyecto P1325.

Referencias

[1] Büyüköztürk, O. Yu, T & Ortega, J. (2006). A methodology for determining complex permittivity of construction materials based on transmission-only coherent, wide-bandwidth free-space measurements. Cement and Concrete Composites 28(4), pp. 349–359.

[2] Kapilevich, B. & Litvak, B. (2013). Application of resonant monopole antenna for determination of complex permittivity of liquids, powdered and granular materials. Measurement 46(6), pp. 1964–1969.

[3] Zhou, R.; Xin, H & Zhang, H. (2014). Liquid-based dielectric resonator antenna and its application for measuring liquid real permittivities. Microwaves, Antennas Propagation 8(4), pp. 255–262.

[4] Bogosanovich, M. (2000). Microstrip patch sensor for measurement of the permittivity of homogeneous dielectric materials. IEEE Transaction Instrument Measurement 49(5), pp. 1144–1148.

[5] Lobato, H.; Corona, A. & Olvera, J. (2013). Planar sensors for RFID wireless complex-dielectric-permittivity sensing of liquids. Proceedings International Microwave Symposium, pp. 2–4. Seattle, USA.

[6] Stuchly, M. et al. (1982). Measurement of radio frequency permittivity of biological tissues with an open-ended coaxial line: Part II - experimental results. IEEE Transaction Microwave Theory and Techniques 30(1), pp. 87–92.

[7] Lesnikowski, J. (2012). Dielectric Permittivity measurement methods of textile substrate of textile transmission lines. Przeglad Elektrotechniczny 3, pp. 148–151.

[8] Lee, K. et al. (2008). Glucose aqueous solution sensing by a near-field microwave microprobe. Sensors and Actuators A: Physical 148(1), pp. 28–32.

[9] Bogosanovich, M. (2000). Microstrip patch sensor for measurement of the permittivity of homogeneous dielectric materials. IEEE Transaction Instrument Measurement 49 (5), pp. 1144–1148.

[10] Lobato, H. et al. (2014). Wireless sensing of complex dielectric permittivity of liquids based on the RFID. IEEE Transaction Instrument Measurement 62(9), pp. 1–8.

[11] Xu, S. et al. (2011). Experimental measurement method to determine the permittivity of extra thin materials using resonant metamaterials. Progress in Electromagnetics Research 120(1), pp. 327–337.

[12] Torun, H. et al. (2014). An antenna-coupled split-ring resonator for biosensing. Journal of Applied Physics 124701(1), pp. 14–20.
[13] Kent, M. (1972). The use of strip-line configurations in microwave moisture measurement. Journal Microwave Power and Electromagnetic Energy 7(3), pp. 185–194.
[14] Bahl, J. & Stuchly, S. (1980). Analysis of a microstrip covered with a lossy dielectric. IEEE Transaction Microwave Theory Techniques 28(2), pp. 104–109.
[15] Sarabandi, K. & Li, E. (1997). Microstrip ring resonator for soil moisture measurements. IEEE Transaction Geoscience and Remote Sensing 35(5), pp. 1223–1231.
[16] Saeed, K.; Pollard, R. & Hunter, I. (2008). Substrate integrated waveguide cavity resonators for complex permittivity characterization of materials. IEEE Transaction Microwave Theory Techniques 56 (10), pp. 2340–2347.
[17] Kobayashi, T. & Nikawa, Y. (2010). Dynamic measurement of temperature dependent permittivity of liquid material using microwaves. Proceedings Asia-Pacific Microwave Conference, pp. 1368–1371. Yokohama, Japan.
[18] Singh, G. (2010). Double negative left-handed metamaterials for miniaturization of rectangular microstrip antenna. Journal of Electromagnetic Analysis and Applications 2(6), pp. 347–351.

© 2015 Editorial Instituto Antioqueño de Investigación

Avances en Ingeniería

Un libro resultado de investigación y producto del trabajo de divulgación de un grupo internacional de investigadores-autores en los campos ingenieriles de las Ciencias Computacionales, la Electrónica y las Telecomunicaciones.

Avances en Ingeniería

ISBN: 978-958-59127-0-0