ISSN: 18770-4069
www.cic.ipn.mx/rsc
www.ipn.mx

Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"

Special issue:
Advances in Intelligent and Information Technologies

E. Cuatecontzi Cuahutle
A. Cortés Fernández
J.F. Ramírez Cruz
J.H. Sossa Azuela
(Eds.)

Vol. 50

RESEARCH IN COMPUTING SCIENCE

ISSN: 18770-4069

Special issue:
Advances in Intelligent and Information Technologies

E. Cuatecontzi Cuahutle
A. Cortés Fernández
J.F. Ramírez Cruz
J.H. Sossa Azuela
(Eds.)

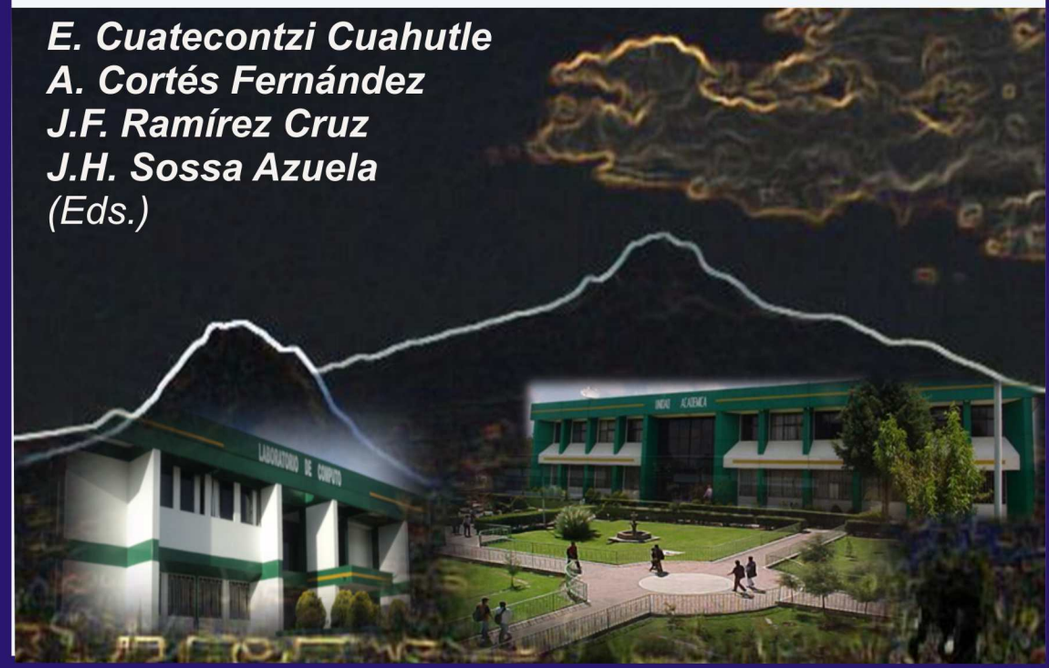


Table of Contents

Intelligent Control

Software Engineering

Intelligent Systems

Distributed Systems

Automatic Speech Recognition

Computer Vision

Artificial Intelligence

Mobile Computing

Data Mining

Control PI-Fuzzy de la Posición del rotor de un motor de Corriente Directa de Capacidad Industrial

Roberto Morales, Ivette Hernández, Edmundo Bonilla y J. Federico Ramírez

División de Estudios de Posgrado e Investigación.
Instituto Tecnológico de Apizaco.
Av. Instituto tecnológico s/n, Apizaco, Tlax. C.P. 90300
moralescaporal@hotmail.com, ivet_05_03@hotmail.com.
Paper received on 16/08/10, Accepted on 28/09/10.

Resumen. Los motores de corriente directa (CD) son usados a menudo en aplicaciones industriales tales como en robots, manipuladores y en otros sistemas accionados electrónicamente donde una amplia gama de control de velocidad y /o posición son necesarias. En este artículo se presenta el desarrollo y simulación de un controlador PI-Fuzzy de posición de un motor de CD. Usando un Control clásico PI y un Control Fuzzy en cascada es posible controlar la posición y la velocidad de una manera relativamente fácil, además de que facilita su implementación digital. Resultados simulados comprueban un buen desempeño del sistema y justifica el análisis teórico.

P**alabras clave:** Control fuzzy, control de posición, control de velocidad, lógica difusa, motor de CD.

1 Introducción

Las técnicas de inteligencia artificial se han convertido en una herramienta fundamental para tratar y modelar sistemas complejos incluyendo el área de Control Automático [1], [2]. El control automático ha desempeñado un papel vital en el avance de la ingeniería y la ciencia. Además de su gran importancia en los sistemas de vehículos espaciales, de guiado de misiles, robóticos y análogos, el control automático se ha convertido en una parte importante e integral de los procesos modernos industriales y de fabricación. Sin embargo el control convencional sigue siendo utilizado en la mayoría de las aplicaciones industriales.

La inteligencia artificial tiene varias ramas, entre ellas se encuentra la lógica difusa. En la lógica difusa el adjetivo "difuso" se debe a que los valores de verdad utilizados tienen, por lo general, una connotación de incertidumbre. La lógica difusa permite definir valores intermedios en un intento por aplicar un modo de pensamiento similar al del ser humano [3].El Control Fuzzy o Difuso se introdujo a comienzos de los años 70 como un intento para diseñar controladores para sistemas que son estructuralmente difíciles de modelar, debido a su naturaleza no lineal y a otras complejidades en la obtención del modelo. Durante los últimos años el Control Difuso ha emergido como una de las áreas de mayor investigación [4].

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp.3 -11.

2 Fundamentos de la Lógica Difusa

La lógica difusa es una técnica de la inteligencia computacional que permite trabajar información con alto grado de imprecisión, en esto se diferencia de la lógica convencional que trabaja con información bien definida y precisa.

En las teorías tradicionales se obliga a que las representaciones del mundo real que se realizan encajen dentro de modelos muy precisos, tomando la imprecisión como un factor de distorsión.

2.1 Conjuntos Clásicos

Los conjuntos en la teoría clásica son altamente restrictivos en el sentido de que un elemento o pertenece o no pertenece a un conjunto dado.

Un conjunto clásico es una colección de elementos que clasifican objetos mediante alguna propiedad. Este tipo de conjuntos se definen mediante una función característica.

Dado un subconjunto A del universo X:

$$\mu_A : X \rightarrow 0, 1 \quad . \qquad (1)$$

se define: $$\mu_A(x) = \begin{cases} 1, \text{ si } x \in A \\ 0, \text{ si } x \notin A \end{cases} . \qquad (2)$$

es decir, si $\mu_A(x) = 1$, si la afirmación "$x \in A$" es verdadera y si $\mu_A(x) = 0$, si la afirmación "$x \in A$" es falsa.

2.2 Conjuntos Difusos

De manera intuitiva se tiene el concepto de conjunto como una colección bien definida de elementos, en la que es posible determinar para un objeto cualquiera, en un universo dado, si acaso éste pertenece o no al conjunto. La decisión, naturalmente, es *"*sí pertenece*"* o bien *"*no pertenece". Un conjunto difuso se puede definir matemáticamente al asignar a cada posible individuo que existe en el universo de discurso, un valor que representa su grado de pertenencia o membresía en el conjunto difuso. Este grado de membresía indica cuando el elemento es similar o compatible con el concepto representado por el conjunto difuso.

Los conjuntos difusos son un instrumento adecuado para modelar los predicados inexactos de los lenguajes naturales [5].

2.3 Funciones de Pertenencia

La función de pertenencia de un conjunto nos indica el grado en que cada elemento de un universo dado, pertenece a dicho conjunto. Es decir, la función de per-

tenencia de un conjunto A sobre un universo X será de la forma: μ_A: X → [0,1], donde μ_A (x) = r si r es el grado en que x pertenece a A [6].

Las funciones de pertenencia son una forma de representar gráficamente un conjunto borroso sobre un universo.

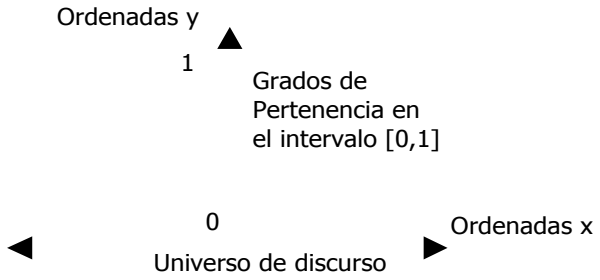


Fig. 1 Representación gráfica de un conjunto borroso sobre un universo.

A la hora de determinar una función de pertenencia, normalmente se eligen funciones sencillas, para que los cálculos no sean complicados. En particular, en aplicaciones en distintos entornos, en la figura 2 se muestran algunas funciones más utilizadas:

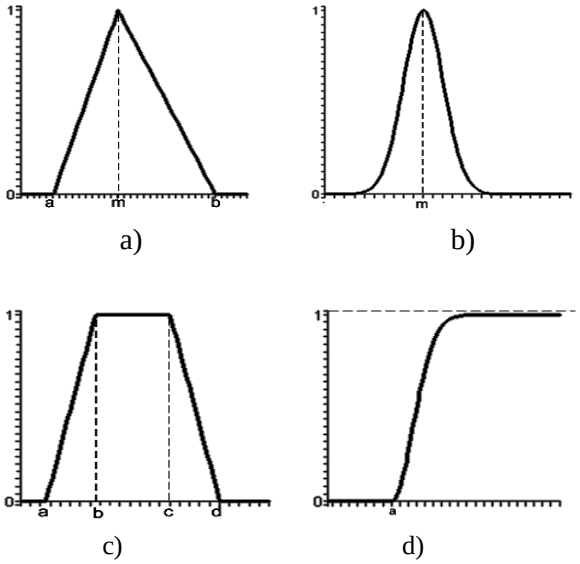


Fig. 2. Funciones de membresía más utilizadas. (a) Triangular, (b) Forma de Campana, (c)Trapezoidal, (d) Gama.

2.4 Operaciones de Conjuntos Difusos

Como ya se dijo, en la lógica difusa se emplean conjuntos, así que contamos con operaciones de complemento unión e intersección. Cuando a dos variables difusas se les aplica una operación de "unión" (que en la lógica binaria es equivalente a una operación OR), el resultado se obtiene tomando el valor más grande de entre las variables de entrada, max $x_1, x_2, \ldots, x_n$.

Para el caso de la "intersección" (que equivale a la operación AND) el valor resultante de la operación corresponde al mínimo valor de alguna de las entradas: $x_1, x_2, \ldots, x_n$.

En la operación "complemento" (equivale a una operación NOT), se toma el valor que complemente a 1, de esta forma:

$$x' = 1 - x \,. \qquad (3)$$

El principal beneficio de la lógica difusa es que con ella se puede describir el comportamiento de un sistema, mediante simples relaciones < si-entonces > o < if-then > éstas permiten describir el conjunto de reglas que utilizaría un ser humano para controlar el proceso con toda la imprecisión que poseen los lenguajes naturales y, solo a partir de estas reglas, generan las acciones que realizan el control. Por esta razón, también se les denominan controladores lingüísticos [7].

2.5 Estructura de un sistema difuso

Un sistema difuso consta de:

a) Etapa de fuzzificación: ésta etapa se encarga de la transformación de las variables controladas entregadas por el proceso, en variables de tipo lingüísticas. Como resultado de la fuzzificación se obtiene valores lingüísticos medidos.
b) Reglas: contiene las reglas difusas que encierran el conocimiento necesario por la solución del problema de control. Las reglas de control constan de reglas si <condiciones> entonces <acciones>.
c) Máquina de inferencia: realiza la tarea de calcular las variables de salida a partir de las variables de entrada, mediante las reglas y la inferencia difusa, entregando conjuntos difusos de salida. En el presente trabajo se implementó el método de inferencia de Mamdani.
d) Defuzzificación: el resultado de la inferencia difusa es retraducido de un concepto lingüístico a una salida física gracias al proceso de defuzzificación.

Por lo tanto, la figura 3 demuestra el diagrama general de un sistema difuso.

Cada una de las variables de entrada y de salida tiene una representación dentro del Sistema de Lógica Difusa en forma de Variables Lingüísticas. Una variable lingüística tiene, entre otras cosas, una colección de atributos que puede adquirir la variable, y cada atributo está representado por un conjunto difuso.

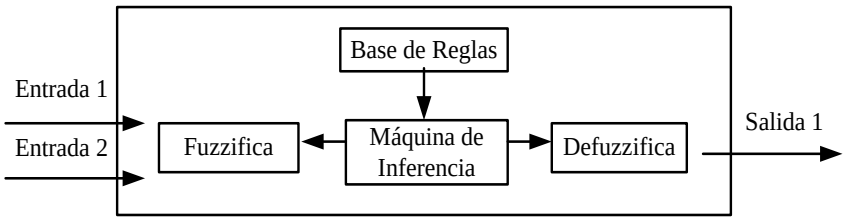


Fig. 3. Estructura de un sistema de Lógica Difusa.

3 Fundamentos de Sistemas de Control

Un Sistema es una combinación de componentes que actúan juntos y realizan un objetivo determinado. Dentro de los sistemas se encuentra el concepto de sistema de control. Un sistema de control es un tipo de sistema que se caracteriza por la presencia de una serie de elementos que permiten influir en el funcionamiento del sistema. La finalidad de un sistema de control es conseguir, mediante la manipulación de las variables de control, un dominio sobre las variables de salida, de modo que estas alcancen unos valores prefijados (consigna) [8].

Los sistemas de control se clasifican en dos grandes categorías:

- Sistemas de lazo abierto.- Es aquel en el cual la acción de control es independiente de la salida.
- Sistemas de lazo cerrado.- Es aquel en el cual la acción de control es en cierto modo dependiente de la salida, estos sistemas se llaman comúnmente, sistemas de control por retroalimentación.

En este trabajo se emplea el sistema de control de lazo cerrado como se puede ver en la figura 4 [8].

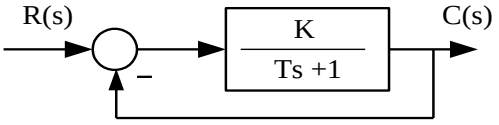


Fig. 4. Sistema en lazo cerrado.

4 Resultados de la simulación

La acción de control Proporcional Integral se define mediante:

$$u(t) = Kp\, e(t) + \frac{Kp}{Ti}\int_0^t e(t)\,. \qquad (4)$$

en donde Kp es la ganancia proporcional y Ti se denomina tiempo integral.

Tanto Kp como Ti son ajustables. El tiempo integral ajusta la acción de control integral, mientras que un cambio en el valor de Kp afecta las partes integral y proporcional de la acción de control [9]. El inverso del tiempo integral Ti se denomina

velocidad de reajuste. La figura 4 muestra un diagrama de bloques de un controlador proporcional más integral [8].

Para obtener los parámetros adecuados para el control PI se emplearon las conocidas reglas de sintonización de Ziegler-Nichols, y se obtuvieron los siguientes resultados:

Kp = 11.58, Ti = 0.45

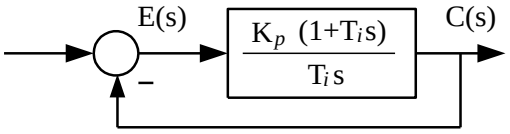


Fig. 5. Diagrama de bloques de un control proporcional-integral.

Las ecuaciones que rigen al motor CD con excitación son:

$$\frac{dI_f}{dt} = -\frac{R_f}{L_{ff}} I_f + \frac{1}{L_{ff}} V_f \cdot \qquad (5)$$

$$\frac{dI_A}{dt} = -\frac{R_A}{L_{AA}} I_A + \frac{L_{Af}}{L_{AA}} I_f \omega + \frac{1}{L_{AA}} V_A \cdot \qquad (6)$$

$$\frac{d\omega}{dt} = -\frac{B}{J} \omega + \frac{L_{Af}}{J} I_f I_A - \frac{1}{J} T_L \cdot \qquad (7)$$

En la tabla siguiente se muestran ciertas especificaciones y parámetros que se tomaron en cuenta para el desarrollo del Control PI-Fuzzy [10].

Tabla 1. Especificaciones y parámetros de un motor de CD.

Especificaciones	
Potencia P = 5 Hp.	Voltaje de armadura V_A = 240 V.
Voltaje de campo V_f =240 V.	Corriente de armadura I_A = 39.58 A.
Corriente de campo I_f = 1.0 A.	Velocidad ω = 120.14 rad/seg.
Parámetros	
Induc. de armadura L_{AA} = 0.012 H.	Resistencia de armadura R_A = 0.6 Ω.
Inductancia de campo L_{ff} = 120 H.	Resistencia de campo R_f = 240 Ω.
Momento de inercia J = 1.2 kg-m^2	Inductancia mutua L_{Af} =1.8 H.
Par externo de carga T_L = 29.2 n-m.	Fricción viscosa B = 0.35 kg-m^2/seg.

Las ecuaciones (**5**), (**6**) y (**7**) son ecuaciones diferenciales de primer orden que contienen los productos no lineales $I_f \omega$ e $I_f I_A$ de estas variables de estado.

En el controlador difuso se tienen dos variables de entrada: error e (t) y cambio en el error $\dot{e}$ (t). La variable de error tiene 7 funciones de membresía. La variable de cambio en el error presenta también 7 funciones de membresía. La salida del controlador es la señal de control u(t), el cual se ha dividido en 7 funciones de membresía. Una forma muy común de representar las reglas difusas, es mediante una representación en forma de matriz, conteniendo dos antecedentes que en este caso son e(t) y $\dot{e}$ (t) y un consecuente u(t).

Tabla 2. Reglas base.

$\dot{e}/e$	NB	NM	NS	Z	PS	PM	PB
NB	NB	NB	NB	NB	NS	PS	PB
NM	NB	NB	NM	NM	Z	PS	PB
NS	NB	NB	NS	NS	Z	PM	PB
Z	NB	NB	NS	Z	PS	PB	PB
PS	NB	NM	Z	PS	PS	PB	PB
PM	NB	NS	Z	PM	PM	PB	PB
PB	NB	NS	PS	PB	PB	PB	PB

Donde:

- NB = Negativo Grande
- NM = Negativo Mediano
- NS = Negativo Pequeño
- Z = Cero
- PS = Positivo Pequeño
- PM = Positivo Mediano
- PB = Positivo Grande

4.1 Simulación del Controlador Difuso

En la figura 6 se muestra el Controlador PI-Fuzzy de posición, el cual se implementó utilizando la herramienta MATLAB/SIMULINK [11].

En la figura 7 al realizar la simulación con el Control PI-Fuzzy, con una referencia de 0.2 rad se observa que cuando entra una carga de 1.2 N-m. en t=5s. hay una excelente reacción ya que se regula la posición llegando a su valor de referencia.

En la figura 8 al realizar la simulación con el Control PI-Fuzzy, con una referencia de 0.1 rad se observa que cuando entra una carga de 2.2 N-m. en t=5s. hay una perturbación pero después se regula la posición.

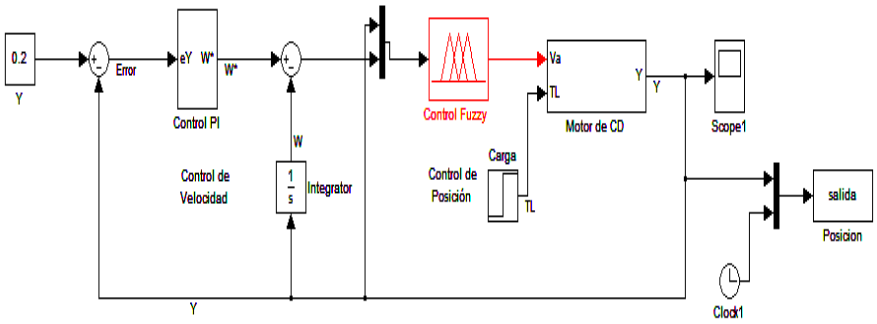


Fig. 6. Esquema de Control PI-Fuzzy propuesto.

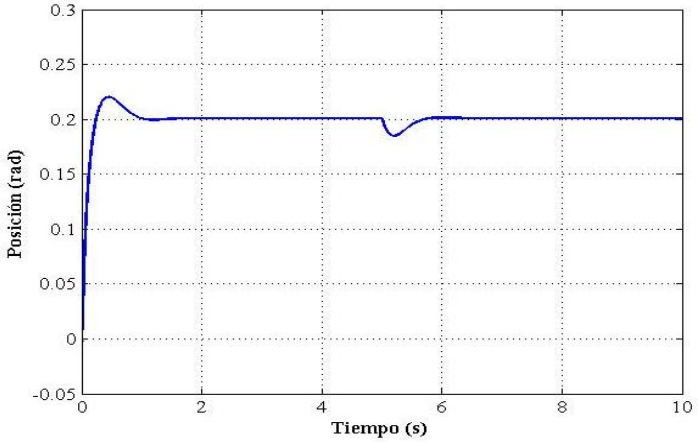


Fig. 7. Simulación del Control Difuso con un valor de referencia de 0.2 rad.

5 Conclusión

El Controlador PI-Fuzzy en cascada propuesto mejora el desempeño o la respuesta del sistema de control. En éste caso el control PI funciona como controlador de velocidad y su salida es usada como referencia del control difuso de posición. Esto se demuestra porque al inicio se llega a la referencia en un tiempo rápido y la recuperación del valor de referencia al arranque y a ciertos disturbios de carga, el control sigue la referencia.

Actualmente se está trabajando en la implementación digital del Controlador PI-Fuzzy usando en un DSP de Texas Instruments para controlar un manipulador industrial cuyos resultados serán reportados en un artículo futuro.

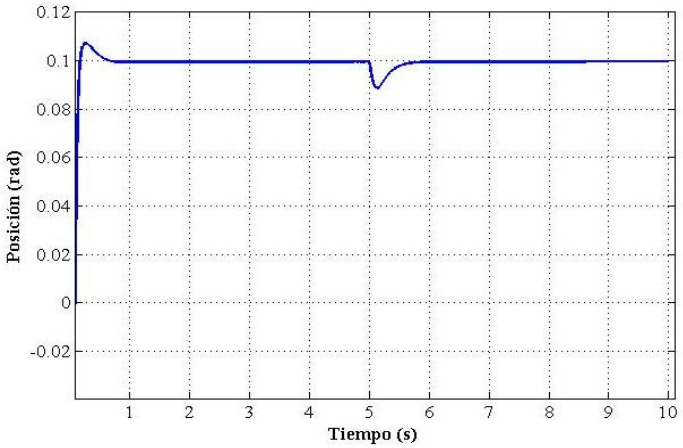


Fig. 8. Simulación del Control Difuso con un valor de referencia de 0.1 rad.

Referencias

1. C. V. Altrock, Fuzzy Logic and Neuro Fuzzy applications explained, Prentice Hall, 1995.
2. C. V. Altrock, "Fuzzy Logic and Neuro Fuzzy Technologies in Appliances" Embedded System Conference, 1999.
3. Phongsak Phakamach,Control of a DC Servomotor Using Fuzzy Logic Sliding mode Model Following Controller, World Academy of Science, Engineering and Technology. 2009.
4. WANG, Li-Xin, "Adaptive Fuzzy Systems and Control", PTR Prentice Hall, Englewood Cliffs, NJ, 1994.
5. Faraón Llorens, Lógica Multivaluadas o Polivalentes, Universidad Politécnica Superior de Alicante España.Departamento de Ciencia de la Computación e Inteligencia Artificial. 2000.
6. Ahmed Rubaai, Senior Member, IEEE, Marcel J. Castro-Sitiriche, Student Member, IEEE, and Abdul R. Ofoli, Member, IEEE, Design and Implementation of Parallel Fuzzy PID Controller for High-Performance Brushless Motor Drives:An Integrated Enviroment for Rapid Control Prototyping, IEEE Transactions on industry applications, vol.44, No., July/August 2008.
7. Hopgood Adrian A., Intelligent System for Engineers and Scientists, CRC press, 2001.
8. Katsuhito Ogata, Ingeniería de Control Moderna, 3a. Edición, Prentice Hall, 1998.
9. Karl Johan Astrom and Tore Hagglund, Automatic Tuning of PID Controllers, ISA, 1988.
10. Leander W. Matsch and J. Derald Morgan, Electromagnetic and Electromechanical Machines, Third Edition, John Miley and Sons, 1986.
11. MATLAB /SIMULINK AND POWER SYSTEM BLOCKSET, SIMULINK® Reference © COPYRIGHT 2002–2010 by The MathWorks, Inc.

ITOHealth: un sistema multimodal basado en localización para el descubrimiento de entidades médicas

José Luis Sánchez-Cervantes, Giner Alor-Hernández, Ulises Juárez-Martínez, Celia Romero-Torres, Ignacio López-Martínez

División de estudios de Posgrado e Investigación, Instituto Tecnológico de Orizaba
Av Oriente 9 No. 852, 94320, Orizaba, Ver.,
México. Télefono/Fax: +52(272) 725 7056
{isc.jolu}@gmail.com, {galor, ujuarez, cromero, imartinez}@itorizaba.edu.mx
Paper received on 29/03/10, Accepted on 05/10/10.

Resumen. La diversidad de hospitales y médicos especialistas genera la dispersión o el aislamiento de la información médica de esas entidades causando su difícil localización. La ubicación de las entidades médicas es muy útil para el cuidado de la salud de las personas, sin embargo cuando la información de dichas entidades está aislada se requiere de un sistema de información que la integre para proporcionar un acceso más rápido en beneficio de los especialistas en el cuidado de la salud y de la sociedad en general. En este trabajo se propone a ITOHealth que es un sistema multimodal que proporciona los mecanismos necesarios para registrar y gestionar la información de las entidades médicas en un único sitio Web y facilita su localización a través de dos opciones: un sitio Web PHP y dos aplicaciones J2ME.

Palabras clave: ITOHealth, localización de entidades médicas, MPML, multimodalidad.

1 Introducción.

El cuidado de la salud es un campo en el que el mantenimiento de registros exactos y la comunicación son fundamentales, en el que el uso de la informática y la tecnología de red van a la zaga con otros campos. Comúnmente, en el cuidado de la salud, la información se transmite de un profesional de la salud a otro a través de notas o comunicación personal. Hoy en día, en que el registro y la comunicación electrónica son fundamentales, la industria del cuidado de la salud aún está ligada a los documentos en papel que suelen perderse fácilmente y que a menudo son ilegibles y fáciles de falsificar. Cuando profesionales de la salud y múltiples instalaciones se involucran en la prestación de asistencia médica de un paciente, los servicios de atención médica proporcionados comúnmente no están coordinados. Algunas veces esta información y el conocimiento están aislados y por lo tanto se convierte en inaccesible para los usuarios que buscan la adquisición de conocimientos para tomar

©E.Cuatecontzi Cuahutle, A.Cortés Fernández,
J. F. Ramírez Cruz, J. H. Sossa Azuela. (Eds.).
Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp.15 -25.

una decisión con respecto al cuidado de la salud. Pero el aislamiento de esta información no es el único inconveniente, también es necesario considerar que existen usuarios con limitaciones visuales o auditivas, y que dichas limitaciones constituyen un problema si cualquiera de estos usuarios requiere tener los conocimientos de la información y la ubicación de cualquier entidad médica para el cuidado de su salud. Considerando lo anterior, en este trabajo se propone un sistema multimodal basado en localización para el descubrimiento de las entidades de médicas llamado "ITO-Health". ITOHealth muestra la información de entidades médicas en dispositivos móviles y browsers Web. La multimodalidad de ITOHealth se enriquece con el uso de personajes de Microsoft Agent, la integración de voz en lenguaje natural a los personajes, el multilenguaje y el soporte para múltiples personajes. La implementación de personajes animados con voz integrada permite a los asistentes "hablar" en español o en inglés de acuerdo a las necesidades del usuario, lo cual es una ventaja para los usuarios que tienen alguna discapacidad visual.

2 Arquitectura de ITOHealth.

La Arquitectura Orientada a Servicios (SOA) es un paradigma arquitectónico para la creación y gestión de servicios de negocios que puedan acceder a los activos y elementos de información con una interfaz común independiente de la ubicación, de la plataforma operativa y el paradigma de implementación. SOA ha existido durante muchos años, pero el rasgo distintivo de la SOA es que se basa en Internet y estándares Web, en particular, los servicios Web. Los servicios Web están diseñados para proporcionar interoperabilidad entre diversas aplicaciones. Hoy en día, un número creciente de empresas, comercios y sistemas de atención médica están redefiniendo sus procesos basándose en esta tecnología.

La independencia de la plataforma y del lenguaje de programación para el desarrollo de las interfaces de los servicios Web permite la integración heterogénea de los sistemas basados en Web. ITOHealth tiene una arquitectura de integración basada en SOA. En la Figura 1 se muestran los elementos internos de la arquitectura de ITOHealth. Cada uno de los elementos tiene una función definida que a continuación se describen brevemente.

Servidor de aplicaciones Xampp.-Es el acceso a los recursos de la aplicación incluidos los servicios Web, los Midlets para la búsqueda de entidades médicas y los motores y personajes para la ejecución de la multimodalidad en el sistema.
Repositorio de información.-Este componente es responsable de mantener la persistencia de los datos como el proceso de registro de hospitales y de médicos especialistas.
Aplicación Web PHP.-La aplicación Web PHP proporciona al usuario las opciones necesarias para la gestión y la búsqueda de las entidades médicas que se registran en el repositorio de información.
Servicios Web.-Los servicios Web en conjunto con los Midlets proporcionan la funcionalidad para realizar búsquedas de las entidades médicas en base a los requerimientos del usuario.

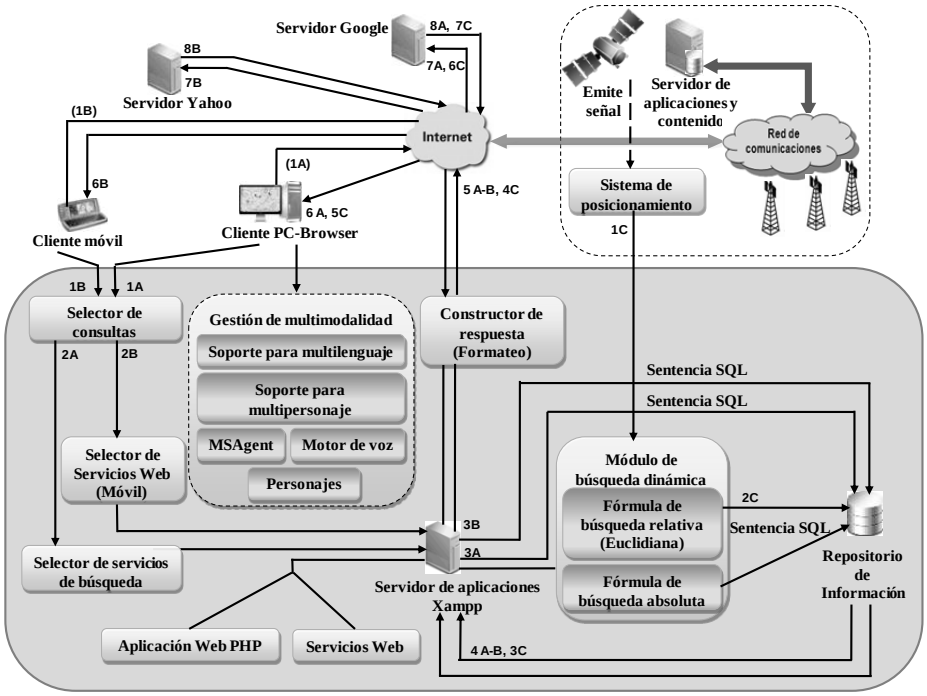


Figura. 1. Elementos internos de la arquitectura de ITOHealth.

Selector de consultas.-Proporciona la capacidad de desplegar, tanto en la aplicación Web como en los Midlets, los tipos de consulta que el usuario realizará. En el caso de la aplicación Web proporciona la ejecución de la sentencia SQL correspondiente para la obtención de resultados.

Selector de servicios de búsqueda.-El selector de servicios de búsqueda tiene la función de seleccionar el servicio de búsqueda apropiado y ejecuta la sentencia SQL correspondiente para la obtención de resultados.

Selector de servicios Web.-La función del selector de servicios Web es seleccionar el servicio Web apropiado para la búsqueda de entidades médicas a través de un dispositivo móvil.

Constructor de respuesta.-Proporciona la funcionalidad para formatear y desplegar correctamente el resultado obtenido al usuario tras la búsqueda de alguna entidad médica.

Módulo de búsqueda dinámica.-Su función es la de realizar búsquedas de entidades médicas localizadas en un determinado radio en kilómetros a partir de la ubicación geográfica del usuario. El cálculo se realiza en base a dos métodos para calcular la distancia entre dos puntos: el cálculo a través de la fórmula de la distancia euclidiana y el cálculo por medio de la fórmula de la distancia absoluta.

Gestión de multimodalidad.-La gestión de multimodalidad tiene componentes para gestionar el soporte para multilenguaje y el soporte para multipersonaje. La ejecución de la multimodalidad se complementa con dos motores: un motor permite su

ejecución en el browser del cliente y el otro motor proporciona el mecanismo de voz asociada con los personajes utilizados en el sistema. Los personajes son aplicaciones que utilizan los motores mencionados para la ejecución de animaciones y expresiones.
Soporte para multilenguaje.-Es un componente de la gestión de multimodalidad y proporciona el mecanismo para cambiar entre los lenguajes inglés y español disponibles en el sistema.
Soporte para multipersonaje.-Es otro componente de la gestión de multimodalidad que permite el intercambio de los cuatro personajes que se utilizan en el sistema.

Los componentes de la arquitectura de ITOHealth están relacionados entre sí. En la siguiente sección se proporciona una breve descripción del flujo de trabajo de cada uno de estos elementos.

3 Flujo de trabajo de ITOHealth.

La interrelación entre los componentes de ITOHealth define el flujo de trabajo para el proceso de búsqueda de entidades médicas. Obsérvese que la Figura 1 se muestra con algunas líneas y letras con números que facilitan la comprensión del flujo de los datos que se procesan durante la ejecución del sistema, el cual se describe a continuación:

1. El usuario con el browser de un equipo de cómputo (1A) o por medio de un dispositivo móvil (1B) y utilizando el componente de selección de consultas, selecciona el tipo de búsqueda de entidades médicas deseado.
2. En el caso de que se seleccione un tipo de búsqueda por medio de un browser de un equipo de cómputo, se envía al selector de servicios de búsqueda el cual realiza la solicitud correspondiente en el servidor de aplicaciones Xampp (2A). Si la búsqueda se realiza con un dispositivo móvil, se envía al selector de servicios Web que selecciona el servicio Web apropiado en el servidor de aplicaciones Xampp (2b).
3. En el servidor de aplicaciones Xampp se procesan las solicitudes y las respuestas del sistema y se conecta al repositorio de información que es responsable de ejecutar las sentencias SQL (3A), (3B).
4. El repositorio de información devuelve los resultados al servidor de aplicaciones Xampp (4A), (4B).
5. En el servidor de aplicaciones Xampp el constructor de respuesta da formato a los resultados obtenidos (5A), (5B). Posteriormente se envían los resultados con el formato apropiado para su despliegue en el browser (6A), (6B).
6. Cuando el cliente Web requiere la ubicación geográfica de alguna entidad médica, se envía una solicitud a través de Internet al servidor Google. Dicha solicitud incluye las coordenadas de latitud y de longitud de la entidad médica. El servidor Google presenta un mapa con un marca-

dor indicando la ubicación geográfica de la entidad médica solicitada (7A), (8A).

7. Cuando el cliente móvil requiere la ubicación geográfica de alguna entidad médica, se envía una solicitud a través de Internet al servidor Yahoo, dicha solicitud incluye las coordenadas de latitud y de longitud de la entidad médica. El servidor Yahoo presenta un mapa con un marcador indicando la ubicación geográfica de la entidad médica solicitada. (7B, (8B).

Si la búsqueda de las entidades médicas se realiza a través de uno de los mecanismos de localización dinámica, el funcionamiento de la arquitectura de ITOHealth se realiza de la siguiente manera:

1. Se obtiene la ubicación geográfica actual del usuario a través del sistema de posicionamiento (1C).
2. Una vez que se obtienen las coordenadas de la ubicación del usuario, se envían al módulo de búsqueda dinámica del sistema donde el usuario realiza búsquedas de entidades médicas a partir de su ubicación especificando un radio entre 5 y 200 kilómetros. Para realizar éste tipo de búsquedas, el usuario tiene dos alternativas: a) a través del cálculo de la distancia euclidiana y 2) a través del cálculo de la distancia absoluta.
3. Posterior al tipo de búsqueda seleccionada por el usuario y a la selección del radio de búsqueda, se realiza la solicitud correspondiente en el servidor de aplicaciones Xampp en el que se procesa la solicitud mediante la conexión con el repositorio de información que ejecuta la fórmula como una sentencia SQL (2C).
4. El repositorio de información devuelve al servidor de aplicaciones Xampp los resultados obtenidos (3C).
5. En el servidor de aplicaciones Xampp, el constructor de respuesta da el formato apropiado a los resultados obtenidos (4C) y posteriormente los envía con el formato apropiado para su despliegue en el browser del cliente (5C).
6. Cuando el cliente Web realiza la localización de una entidad médica con un mapa de Google, se envía una solicitud a través de Internet al servidor Google. Dicha solicitud incluye las coordenadas de latitud y longitud de la entidad médica. El servidor Google presenta un mapa con un marcador indicando la ubicación geográfica de la entidad médica solicitada. (6C), (7C).

Una innovación de esta propuesta es el uso de la multimodalidad. La multimodalidad es un proceso en el que diferentes dispositivos y personas son capaces de interactuar (en forma auditiva, visual, táctil y gestual) de manera conjunta en cualquier lugar y en cualquier momento, utilizando cualquier dispositivo móvil o un equipo de cómputo, incrementando así la interacción entre ITOHealth y los usuarios. Los componentes de la multimodalidad de ITOHealth se describen a continuación:

1. El MSAgent es una tecnología que proporciona la infraestructura básica para gestionar los componentes del sistema multimodal, como motor de la voz y los personajes.
2. El motor de voz proporciona el modo de comunicación de voz a través de los personajes y gestiona el intercambio entre los idiomas inglés y español.
3. Los personajes son los componentes que proporcionan capacidades de interacción específicos como la transmisión de expresiones y movimientos.

Este trabajo ofrece dos modalidades para facilitar el uso de ITOHealth a los usuarios: la primera modalidad es una aplicación Web desarrollada en PHP para la gestión y la búsqueda de las entidades médicas con la integración de características de multimodalidad a través del Lenguaje de Marcas de Presentación Multimodal (MPML). MPML es un lenguaje basado en XML desarrollado para permitir la descripción de la presentación multimodal utilizando agentes de carácter [1]. MPML requiere del MSAgent para su ejecución. MSAgent es una tecnología desarrollada por Microsoft que utiliza la animación, el reconocimiento de texto y de voz.

Funcionalmente MPML integra características de la interacción como el streaming de audio y vídeo con imágenes, texto o cualquier otro medio. En este contexto, el streaming se refiere a la ejecución conjunta de texto, audio, vídeo e imágenes. La otra modalidad de esta propuesta es obtener la información y la ubicación geográfica de las entidades medicas mediante el uso de dispositivos móviles. Esto se logra con el desarrollo de servicios Web que proporcionan los resultados de las búsquedas y que se invocan desde una aplicación J2ME. Un servicio Web es un sistema de software diseñado para apoyar la interacción máquina a máquina sobre una red. Tiene una interfaz descrita en un formato procesable por una computadora (específicamente WSDL) [2]. WSDL (*Web Services Description Language)* es un formato XML para describir los servicios de red como un conjunto de variables que operan con los mensajes contenidos ya sea en información orientada a documentos u orientada a procedimientos [3]. La aplicación J2ME se desarrolló usando LWUIT (*Lightweight User Interface Toolkit*). LWUIT es una biblioteca de interfaz de usuario para la plataforma Java ME que permite a los desarrolladores crear interfaces de usuario visualmente atractivas, funcionales y sofisticadas que parecen y se comportan de la misma manera en todos los dispositivos habilitados para Java ME compatibles con MIDP 2.0 (Perfil de Información Para Dispositivos Móviles) y CLDC 1.1 (Configuración para Dispositivos con Conexión Limitada). Esta biblioteca proporciona mejoras a los componentes gráficos y de diálogo, también ofrece nuevas características como la animación, la transición y el desarrollo de temas [4].
La figura 2 ilustra la multimodalidad de ITOHealth.

Figura. 2. Multimodalidad de ITOHealth.

4 Caso de estudio: búsqueda de médicos en un determinado sitio.

Para este caso de estudio sólo se considera la ciudad de Orizaba Ver. Sin embargo, ITOHealth permite localizar entidades médicas de otras ciudades en otros estados de la República Mexicana.

1. Supóngase que un turista de otro país se encuentra caminando en la ciudad de Orizaba y requiere obtener información acerca de médicos cardiólogos existentes en el estado de Veracruz, para una posible revisión médica.
2. Hay diversos médicos especialistas en múltiples sitios de la ciudad y en el estado de Veracruz. Se requiere obtener información para localizar a algún cardiólogo que le proporcione la revisión médica a este usuario.
3. El usuario se encuentra en el centro de la ciudad y posee un celular con los Midlets de búsqueda de entidades médicas instalados.
4. Hay entidades médicas registradas en el sistema ITOHealth y opcionalmente tienen su propia página Web.
5. Una alternativa de solución para este usuario es buscar entidades médicas con cualquiera de los motores de búsqueda disponibles en la Web como Google, Yahoo, entre otros, sin embargo, posiblemente la información que se encuentre sea redundante e innecesaria.

Durante su búsqueda, posiblemente al usuario le surjan algunos inconvenientes y posibles preguntas tales como:

1. ¿Cómo optimizar el tiempo de búsqueda de las entidades de médicas que requiere el usuario para una revisión médica?
2. ¿Cómo localizar las entidades médicas que ofrece una especialidad de cardiología?
3. ¿Cómo obtener información detallada y la ubicación geográfica de las entidades médicas?

Una alternativa de solución al problema anterior es utilizar el sistema ITO-Health. Para este caso de estudio se discutirá el modo de localización de entidades médicas a través de la aplicación móvil de búsqueda de entidades médicas en inglés. La descripción para la búsqueda de las entidades médicas requeridas por el usuario se describe de la siguiente manera: el usuario elige la opción de búsqueda por especialidad y selecciona la opción buscar médicos en el menú. Posteriormente, selecciona el estado de Veracruz y selecciona la especialidad de cardiología. La búsqueda despliega los resultados en una lista que muestra el nombre y el número de cédula profesional de los cardiólogos. El usuario tiene la opción de ver los detalles del médico especialista a través de su número de cédula profesional tal como se presenta en la figura 3A. En la aplicación móvil el usuario tiene la opción de obtener información acerca de la(s) especialidad(es) del médico y su ubicación geográfica mediante un mapa Yahoo.

En las figuras 3B y 3C se presenta dicha información. En las figuras 3D, 3E y 3F se muestran los resultados del caso de estudio mencionado previamente pero utilizando la aplicación Web PHP. Nótese que son los mismos resultados que se muestran en la aplicación para dispositivo móvil.

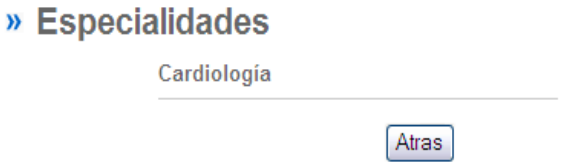
3A Detalles del médico en la aplicación móvil.

3B Especialidades del médico en la aplicación móvil.

3C Ubicación geográfica del médico en la aplicación móvil.

3D Detalles del médico en la aplicación Web PHP

3E Ubicación geográfica del médico en la aplicación Web PHP



3F Especialidades del médico en la aplicación Web PHP

Este caso de estudio muestra la utilidad de ITOHealth como una alternativa para el cuidado de la salud de las personas mediante la búsqueda de entidades médicas, utilizando un dispositivo móvil y que ITOHealth es muy útil para personas que no residen en una ciudad en particular como los turistas o visitantes. ITOHealth está disponible en: http://www.itohealth.net

5 Trabajos relacionados.

Con el transcurrir de los años, los servicios para el cuidado de la salud han sido importantes porque mejoran la vida de las personas. En este sentido, a continuación se presentan algunas propuestas importantes relacionadas con ITOHealth. En [5], [6] y [7], se utilizó una arquitectura LBS para obtener la ubicación geográfica del usuario a través de GPS o dispositivos móviles y el uso de sensores como: el WLD (Dispositivo de iluminación portátil) para monitorizar los signos vitales de los pacientes. ITOHealth tiene características de multimodalidad, obtiene la ubicación geográfica de las entidades médicas utilizando mapas Google, proporciona información detallada sobre las entidades médicas y tiene dos modalidades de uso: mediante una página Web o utilizando dispositivos móviles. En [8] se desarrolló un framework basado en la clasificación de servicios de marketing, utilizan el sistema LBS para localización de negocios, vehículos y personas que utilizan dispositivos móviles con GPS. ITOHealth proporciona a los usuarios los servicios necesarios para obtener la ubicación geográfica de las entidades médicas a través de una página Web o un dispositivo móvil. En [9] se presentó un Middleware con la adición de sensores para optimizar los sistemas de LBS dirigidos a prestar un servicio determinado. Del mismo M en

[10] se presenta un modelo para la ubicación del personal y equipo médico en un hospital a través de sensores RFDI y la tecnología Wi-Fi. ITOHealth propone una arquitectura de LBS con la integración de servicios Web para optimizar la ubicación de las entidades médicas.

6 Trabajo a futuro.

Como trabajo a futuro, consideramos integrar voz en lenguaje natural a través de VoiceXML o utilizar el motor de reconocimiento de voz de MSAgent complementado con el lenguaje MPML. Además, pretendemos incorporar la localización dinámica de las entidades médicas a través de un dispositivo GPS o por medio de geolocalización por IP también llamada Geo-IP Targeting.

7 Conclusiones

La localización de las entidades médicas es un proceso complejo que requiere tiempo y puede ser crítico cuando una persona requiere atención médica urgente. Las emergencias médicas son generalmente tratadas por una estructura basada en la comunicación entre personas, lo que comienza con la persona que da aviso de la emergencia a un operador de emergencias que se comunica con las entidades médicas correspondientes. Esta forma de comunicación no es eficiente sobre todo cuando se trata de un sistema que no es local. ITOHealth ofrece una alternativa para solucionar este problema en el que la sociedad necesita localizar y adquirir información sobre las entidades médicas para el cuidado de la salud.

Este trabajo expone la importancia de los problemas relacionados con la localización de entidades médicas para el cuidado y atención de la salud, también explica la arquitectura del sistema desarrollado para proporcionar la solución a estos problemas, así como sus componentes y flujo de trabajo.

Agradecimientos

Este trabajo fue apoyado por la Dirección General de Educación Superior Tecnológica de México (DGEST). Además, este trabajo fue patrocinado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) y la Secretaría de Educación Pública a través de PROMEP.

Referecias

1. T. Tsutsui, S. Saeyor and M. Ishizuka. MPML: A Multimodal Presentation Markup Language with Character Agent Control Functions. Proc. (CD-ROM) WebNet 2000 World Conf. on the WWW and Internet, San Antonio, Texas, USA.
2. Papazoglou, M.P. Service-Oriented Computing: Concepts, Characteristics and Directions. Proceedings of the Fourth International Conference on Web Information Systems Engi-neering. IEEE Press. 2003.

3. Curbera F., Duftler M., Khalaf R., Nagy W., Mukhi N., y Weerawarana S., "Unraveling the Web Services Web An introduction to SOAP, WSDL, and UDDI", IEEE Internet Computing, 2002.
4. Biswajit Sarkar, LWUIT 1.1 for Java Me Developers, PACKT Publishing, pp 7--10 ISBN 978-1-847197-40-5, 2009.
5. Maged N Kamel Boulos, Artur Rocha, Angelo Martins, Manuel Escriche Vicente, Armin Bolz, Robert Feld, Igor Tchoudovski, Martin Braecklein, John Nelson, Gearóid Ó Laighin, Claudio Sdogati, Francesca Cesaroni, Marco Antomarini, Angela Jobes and Mark Kinirons."CAALYX: a new generation of location-based services in healthcare" International Journal of Health Geographics, pp. 1-6, 2007.
6. Antonio Coronato, Massimo Esposito, "Towards an implementation of Smart Hospital: a localization system for mobile users and devices", Sixth Annual IEEE International Conference on Pervasive Computing and Communications, pp. 1-5, 2008.
7. Jinsoo Ahn, Jungil Heo, Suyoung Lim, Wooshik Kim, "A Study on Ubiquitous Healthcare System based on LBS", Proceedings of the World Congress on Engineering, pp. 1-4, 2008.
8. Ana M. Bernardos, José R. Casar and Paula Tarrío, "Building a framework to characterize location-based services", International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST), pp. 1-6, 2007.
9. Lorcan Coyle, Steve Neely, Paddy Nixon, Aaron Quigley, "Sensor Aggregation and Integration in Healthcare Location Based Services", pp. 1-4, 2008.
10. Vladimir Stantchev, Tino Schulz, Trung Dang Hoang, and Ilja Ratchinski, "Optimizing Clinical Processes with Position-Sensing", HealthCare IT, pp. 1-6, 2008.

Visualización de objetos de aprendizaje a través de una aplicación orientada a servicios basada en J2ME y LWUIT

Maritza Bustos-López[1], Ana María Chávez-Trejo[1], Giner Alor-Hernández[1], María Antonieta Abud-Figueroa [1], Ignacio López-Martinez[1]

[1] División de estudios de posgrado, Instituto Tecnológico de Orizaba.
Email: maritbustos@hotmail.com, achavez@itorizaba.edu.mx, galor@itorizaba.edu.mx, mabudfig@gmail.com, nachouno@hotmail.com
Paper received on 01/07/10, Accepted on 21/09/10.

Resumen. En este trabajo se propone el desarrollo de una herramienta para la visualización de objetos de aprendizaje en dispositivos móviles a través de una aplicación orientada a servicios basada en J2ME y LWUIT. El objetivo de la implementación de esta herramienta es proporcionar a médicos residentes un medio permanente de aprendizaje, portable y que le permita construir conocimiento en cualquier lugar y momento conociendo su ritmo y entorno de trabajo que requiere gran movilidad en el marco de sus rutinas diarias. Las ventajas de propiciar un contexto de aprendizaje móvil permiten emigrar a sistemas de educación moderna a la vanguardia con la tecnología actual y un sistema de conectividad e interactividad entre los médicos residentes y catedráticos del área.

Palabras clave: LWUIT, J2ME, objetos de aprendizaje, Servicios Web**.**

1 Introducción

El proceso de enseñanza aprendizaje tiene requerimientos que deben atenderse, la necesidad de emigrar a sistemas modernos con materiales educativos que perduren, fáciles de localizar, acceder e interpretar independientemente del espacio y tiempo del profesor y alumno representa una evolución obligada en esta área. Actualmente los materiales educativos existentes no se utilizan por la comunidad médica hispana en general, lo que provoca que el conocimiento se encuentre aislado lo que origina que no se comparta ni enriquezca. La reutilización de conocimiento motiva la evolución hacia nuevos esquemas de enseñanza aprendizaje. Este trabajo propone el uso de una herramienta de administración del aprendizaje para la construcción de una plataforma educativa médica con objetos de aprendizaje (OA) visualizados en dispositivos móviles. El OA encapsula un conjunto de materiales digitales que en unidad o agrupación permiten alcanzar objetivos educacionales apoyados en el uso de tecnología. Además los OA presentan atributos importantes: (1) reutilizables, (2) accesibles, (3) interoperables, (4) portables y (5) durables.

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp.27-36.

2 Trabajos relacionados

El proceso de enseñanza aprendizaje emigra a sistemas modernos, destacando la integración de plataformas de aprendizaje con tecnología móvil que facilita el proceso de enseñanza aprendizaje para realizarse en cualquier momento y lugar. A continuación se describe una serie de trabajos relacionados con el aprendizaje móvil.

En [1] se presenta un desarrollo de una arquitectura práctica de objetos de aprendizaje interactivo móvil (MILOs por sus siglas en Inglés) que se usa en un Motor de aprendizaje móvil (MLE por sus siglas en Inglés). El MLE tiene como núcleo principal presentar objetos de aprendizaje: en formatos de texto, integrados con imagen y texto, con hipervínculos y elementos con acciones específicas y un menú de audio y video para su reproducción. Su contenido incluye también la evaluación del aprendizaje que se diseña por medio de preguntas interactivas con ayudas inteligentes: preguntas (una sola elección, múltiples elecciones), gráfica para la formulación de preguntas. Un marco de trabajo auto-adaptativo de objetos de aprendizaje se describe en [2] como un contexto para el aprendizaje ubicuo, construido en WML (Wireless Markup Language) bajo el estándar LOM (Learning Object Metadata) y Metadatos Knowledge-node. En [3] se define una arquitectura que apoya el proceso de aprendizaje móvil consta de: (1) cliente móvil, (2) sistema de gestión de aprendizaje móvil permite crear, editar y eliminar objetos de aprendizaje, (3) repositorio de objetos de aprendizaje. Compatible con SCORM (Sharable Content Object Reference Model) y especificaciones Question and Test Interoperability. En [4] se presenta una comparación de cuatro LMS (Learning Management Systems): Dokeos, ATutor, ILIAS y Moodle, sistemas gestores de aprendizaje de código abierto, soportan el estándar SCORM 1.2 para objetos de aprendizaje, empaquetados con el editor RELOAD. En [5] se presenta un marco de trabajo conceptual para aplicaciones educativas móviles basadas en objetos de aprendizaje. Su diseño y arquitectura contiene una orientación tecnológica y pedagógica, planteando como objetivo principal el proveer un conjunto de herramientas, componentes y especificaciones basadas en las mejores prácticas, experiencias y estándares, en donde la estructura, contenido y la aplicación misma se basa en objetos de aprendizaje, centrándose en los aspectos educativos de la aplicación móvil. En [6] se propone un modelo conceptual para objetos de aprendizaje en ambientes móviles. En este modelo se plantean estrategias de aprendizaje como: (1) aprendizaje situado, (2) aprendizaje colaborativo y (3) aprendizaje personalizado. Implica también tres modelos para diseño y uso de objetos de aprendizaje en móviles: (1) personalización, (2) interacción y colaboración. Un modelo para objetos de aprendizaje colaborativos en dispositivos móviles (MbCLO) se propone en [7] los aspectos más importantes del enfoque son ((LO, por sus siglas en inglés) → objetos de aprendizaje, CSCL → aprendizaje colaborativo soportada por computadora y m-learning → aprendizaje móvil). En [8] se presentan objetos de aprendizaje médico en el que propone el uso de XSLT-transformaciones de los objetos cuando se ejecutan. Finalmente propone un servidor de objetos de aprendizaje médicos que proporciona el desarrollo de una interfaz para: (a) el proveedor de contenidos: la creación y la actualización de los objetos de aprendizaje médico, (b) el usuario: la taxonomía y la ontología basada en la interfaz de búsqueda y (c) las aplicaciones médicas: una API para la búsqueda de objetos de aprendizaje médico. En

[9] se centra en la utilización del sistema gestor de aprendizaje Moodle y el Ariadne Knowledge Pool (Ariadne repository). Además de LOV, arquitectura original que implementa dos LMS (INES y Moodle), y cuatro LOR, repositorio de objetos de aprendizaje (MERLOT, ARIADNE, Edna y NIME).

3 Objetos de Aprendizaje

Los objetos de aprendizaje se identifican por la reutilización de recursos educativos en múltiples contextos de aprendizaje, la cual se apoya en la Web, junto con estándares que permite, principalmente, describir los recursos, facilitar su localización y tecnologías como los entornos de aprendizaje móvil que personalizan la manera de aprender al crear, enriquecer, distribuir y mostrar material en dispositivos móviles.

La tecnología que se propone en este trabajo para el desarrollo y visualización de objetos de aprendizaje a través de dispositivos móviles es J2ME, LWUIT, Servicios Web y NuSOAP. J2ME contiene una mínima parte de las APIs (Application Programming Interface) de Java, porque la edición estándar de APIs de Java ocupa 20 Mb, y los dispositivos pequeños disponen de una cantidad de memoria mucho más reducida. J2ME usa 37 clases de la plataforma J2SE (Java 2 Platform, Standard Edition) provenientes de los paquetes java.lang, java.io, java.util. Otras diferencias con la plataforma J2SE es el uso de una máquina virtual distinta de la clásica JVM denominada KVM. KVM tiene restricciones que hacen que no tenga todas las capacidades incluidas en la JVM [10]. LWUIT es una biblioteca gráfica la cual desarrolla Sun Microsystems, Inc. con el fin de proveer a los programadores de una herramienta con la cual mejorar las interfaces gráficas en cuanto a usabilidad y calidad para aplicaciones móviles con tecnología J2ME [11]. Los servicios Web se usan para desarrollar nuevos componentes de software o para construir un puente de comunicación con los sistemas internos de los socios comerciales en la cadena de suministro con el fin de exponer sus procesos de negocios de manera externa. También los servicios Web se utilizan internamente para proveer interfaces programables a sistemas ya existentes y para integrar aplicaciones Web con estos sistemas [12]. NuSOAP es un kit de herramientas para desarrollar servicios Web bajo el lenguaje PHP. Está compuesto por una serie de clases que permiten a los desarrolladores crear y consumir servicios Web SOAP. Provee soporte para el desarrollo de clientes (aquellos que consumen los servicios Web) y de servidores (aquellos que los proveen). NuSOAP está basado en SOAP 1.1, WSDL 1.1 y HTTP 1.0/1.1. Soporta gran parte de la especificación SOAP 1.1. Genera WSDL 1.1 y también lo consumen para su uso en la serialización. Tanto los servicios de RPC / encoded y document/literal son compatibles [13].

4 Arquitectura de la plataforma

La arquitectura de la plataforma educativa permite la comprensión general del proceso de visualización de los objetos de aprendizaje. En la figura 1, se presenta un panorama general de los elementos que constituyen la arquitectura de la herramienta

desarrollada, donde cada uno de ellos tiene una función específica que a continuación se describe.

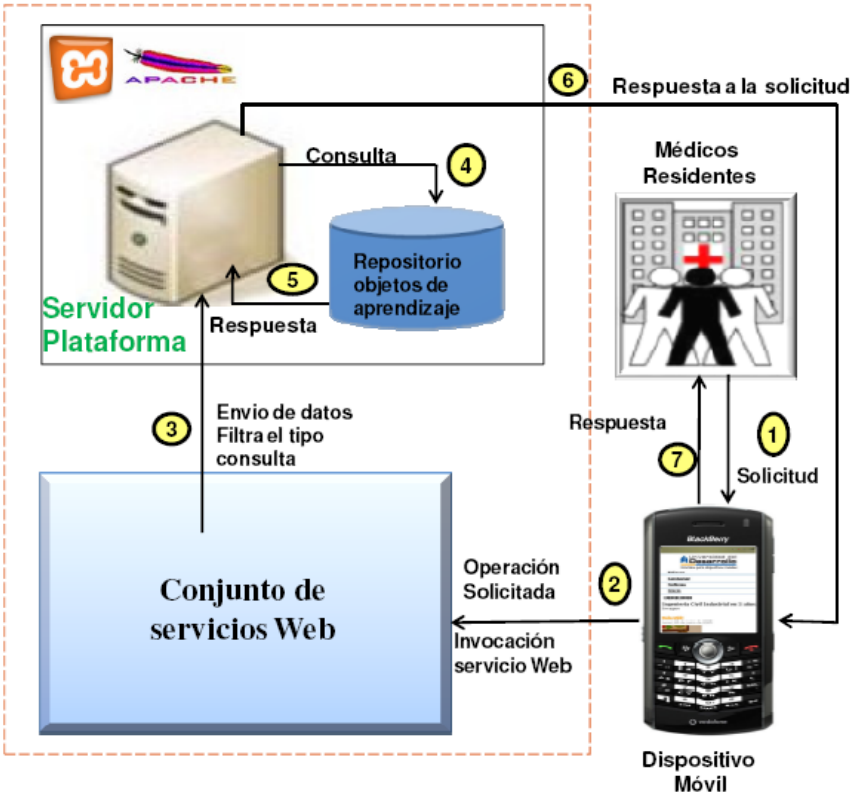


Fig. 1 Arquitectura del sistema

4.1 Componentes de la arquitectura

La arquitectura para la visualización de objetos de aprendizaje en la plataforma educativa consta de los siguientes componentes de Hardware y Software:

- Servidor de la plataforma: El servidor tiene instalado el servidor Web Apache en XAMP, este componente atiende y responde a las diferentes peticiones, proporcionando los recursos que le solicitan, se mantiene en comunicación directa con el repositorio de OA para dar respuesta a las solicitudes.
- Conjunto de servicios Web: Proporciona el mecanismo para filtrar lo tipos de consulta que realiza el usuario en la localización de los objetos de aprendizaje.
- Repositorio de Objetos de aprendizaje: Este módulo registra la información de los objetos de aprendizaje.

En la arquitectura, cada elemento tiene una interrelación que se detalla a continuación:

4.2 Flujo de información

La relación entre los componentes de la arquitectura describe el flujo de información del sistema gestor de aprendizaje a través del cual se detalla los datos necesarios para realizar los procedimientos, operaciones y solicitudes de parte de los especialistas y médicos residentes actores principales del sistema. A continuación la descripción del flujo de información:

1. El médico residente efectúa una solicitud mediante un dispositivo móvil para visualizar los objetos de aprendizaje en formato de texto, video, audio o imagen y así continuar con su proceso de aprendizaje.
2. El dispositivo móvil envía esta solicitud por medio de una conexión WIFI al conjunto de servicios Web para que detecte la operación que requiere el usuario.
3. En los servicios Web se invoca el tipo de función necesaria para realizar la consulta que requiere el usuario.
4. Con base en la operación detectada, el servidor procesa la solicitud conectándose con el repositorio de objetos de aprendizaje para consultar información de los objetos de aprendizaje, la cual se realiza mediante una sentencia SQL.
5. El repositorio devuelve los resultados de la sentencia SQL solicitada al servidor de la plataforma.
6. El servidor envía a través de Internet la respuesta a la solicitud que realiza el especialista.
7. Finalmente, una vez obtenido la respuesta del servidor en el dispositivo móvil se muestra un listado de OA que corresponden al criterio de búsqueda, al seleccionar un objeto específico se visualiza en pantalla información del metadato, reproducción total del objeto.

Esta arquitectura se diseñó y desarrolló con el fin de que alguna entidad médica que cuente con un departamento de enseñanza pueda utilizarla para mejorar el proceso educativo entre los especialistas y médicos residentes.

El visualizador de objetos de aprendizaje instalado en el dispositivo móvil ofrece al usuario tres tipos de localización de objetos, cada búsqueda plantea un criterio de consulta, el cual es atendido por una función particular del servicio Web para dar respuesta a la solicitud del cliente móvil, a continuación se presenta las descripciones de las funcionalidades de los servicios Web desarrollados para la búsqueda de OA:

Consulta por área de conocimiento: En este tipo de búsqueda el usuario selecciona de una lista desplegable el área al igual que introduce en una caja de texto una palabra clave para dar inicio a la localización del objeto que desea.

Consulta por tipo de formato: En esta búsqueda se selecciona de una lista desplegable el formato del OA al igual que introduce en la caja de texto una palabra clave. Esta función realiza la invocación al servicio Web para realizar la consulta de un OA en formato de video o audio o imagen o texto.

Consulta por palabra clave: En este tipo de búsqueda el usuario introduce en una caja de texto una palabra clave para dar inicio a la localización del objeto que desea. Esta función realiza la invocación al servicio Web para realizar la consulta de un OA.

Cada servicio Web provee mecanismos de manejo de excepciones ya que primero se verifica que exista una conexión hacia el repositorio de OA, que exista una palabra clave para extraer por medio de una consulta los registros existentes en una lista bajo el criterio que realiza el usuario, obteniendo datos del OA como: el identificador, título, descripción, autor, área, sub-área, tipo formato y URL, para visualizar, reproducir e imprimir en pantalla información del OA.

A continuación se presenta un caso de estudio que permite observar la funcionalidad de la herramienta educativa desarrollada. Este caso de estudio se basa en la enseñanza de procedimientos médicos.

5 Resultados

En esta sección se presenta un caso de estudio mediante el cual un médico residente visualiza objetos de aprendizaje a través de la aplicación desarrollada.

Caso de estudio: Procedimiento de sutura.

a) Suponga que sólo existe un médico por especialidad en una entidad médica.
b) Este médico especialista tiene bajo su supervisión médicos residentes los cuales tienen diferentes jornadas de trabajo e intensidad horaria.
c) En cada jornada de trabajo, cada residente obtiene una información diferente del especialista.

En este escenario, ¿cómo el médico especialista pudiera ofrecer sus conocimientos a sus residentes de una manera óptima y que además, el conocimiento transmitido se pudiera reutilizar y sea homogéneo para futuros residentes, para el personal del hospital?

Con nuestra propuesta, se ofrece una solución a esta pregunta. Para llevar a cabo lo anterior, el especialista crea los objetos de aprendizaje necesarios para transmitir el conocimiento a los médicos residentes o en su caso al personal del hospital regional de Río Blanco, suponga que se trate del procedimiento de sutura de una herida. Estos OA se editan a través de la plataforma desarrollada y se almacenan en el repositorio de la arquitectura planteada. Una vez que el especialista o personal del área educativa del hospital hayan creado, validado los objetos de aprendizaje y se encuentren almacenados en el repositorio, un médico residente o usuario accede desde su dispositivo móvil y ejecuta la aplicación visualizadora de OA que previamente se instaló. Al ejecutar la aplicación ingresa a la interfaz de inicio como se muestra en la figura 2a y 2b. En la figura 2b se muestra la interfaz del menú principal el cual presenta tres botones uno para conocer acerca de la aplicación, el segundo conecta a la interfaz para localizar los OA y el tercero para salir de la aplicación. Una vez que el médico residente se encuentra en la interfaz del menú principal selecciona el botón búsqueda Objeto de aprendizaje que lo direcciona hacia la interfaz de localización de OA. En esta interfaz el usuario establece el criterio de búsqueda y ofrece tres tipos de consulta: 1) Consulta por área, 2) consulta por tipo y 3) consulta por palabra clave. Al seleccionar buscar OA por área, el usuario indica en una lista desplegable

el área e introduce en una caja de texto una palabra clave para delimitar la consulta y observar un listado de todo aquel OA que corresponda a este criterio y así dar inicio al proceso de solicitud del OA y realizar la conexión al repositorio e invocar el servicio Web correspondiente al tipo de búsqueda. En la figura 2c y 3a se muestra un ejemplo de la localización de un OA que pertenece al área de enfermería, su palabra clave es procedimiento cuyo contenido es un video en formato MPEG-4 (MP4), el cual ofrece al médico residente conocimiento acerca de la técnica e instrumentos utilizados en el procedimiento de suturar una herida. Luego de establecer los criterios de búsqueda la herramienta visualizadora realiza la conexión al repositorio de OA e invoca al servicio Web correspondiente para dar respuesta a la solicitud. Una vez recuperados los nombres de los OA que cumplen con los criterios de búsqueda anteriores se presenta una interfaz con una lista de OA de la cual el usuario tiene la opción de seleccionar los OA a visualizar como se observa en la figura 3b.

Figura 2a. Visualizador de OA del HRRB

Figura 2b. Menú principal

Figura 2c. Localización de un OA por área

Figura 3a. Criterios de búsqueda de un OA

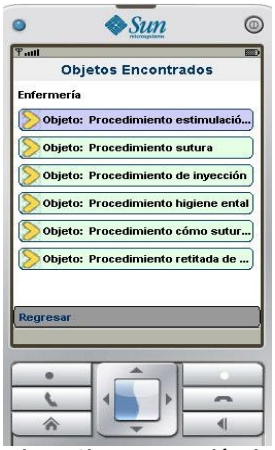


Figura 3b. Recuperación de los OA según criterio de búsqueda

Figura 3c. Interfaz de reproducción del OA en formato video

El usuario al seleccionar el OA a visualizar en su dispositivo móvil se descubre la interfaz de menú de reproducción del video como se observa en la figura 3c y 4a. El menú correspondiente al formato de video consta de cuatro items play, stop, regresar y detalles que presenta información del OA como título, descripción, área, sub-área, formato y URL se muestra en la figura 4b. El usuario desde su celular tiene la opción de reproducir varias veces el OA así como también descubrir más elementos que proporcionen conocimiento del tema procedimiento de sutura en los otros tipos de búsqueda tal como se presenta en la figura 4c. En la figura 5a se presenta una localización de un OA en formato JPG que proporciona al médico residente un recurso para identificar la técnica de suturar e instrumentos utilizados en el procedimiento, de igual forma esta interfaz realiza una invocación al servicio Web y establece la conexión al repositorio. Posteriormente, el usuario consulta por palabra clave otro elemento del OA, proporcionando una palabra en la caja texto de su interfaz da inicio a buscar un nuevo OA en donde se presenta la interfaz de consulta para OA de tipo de texto tal como se presenta en le figura 5b. En la figura 5c se presenta la visualización del procedimiento de retirada de sutura en formato de texto. Finalmente, el usuario desde su dispositivo móvil cuenta con una herramienta que hace parte de un recurso en el proceso de aprendizaje y construcción de conocimiento ya que ofrece consultas a OA propios del hospital en casos clínicos, procedimientos y conocimiento en general del contexto de la medicina en las diferentes áreas de forma atractiva en formatos de audio, video, texto e imagen.

Figura 4a. Interfaz reproducción del video Cómo suturar una herida

Figura 4b. Interfaz de información del OA Etiquetas del metadato LOM

Figura 4c. Consulta OA por tipo

Figura 5a. Visualización OA procedimiento sutura

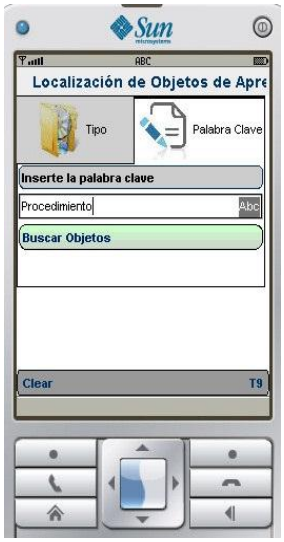


Figura 5b. Interfaz de consulta de OA por palabra clave

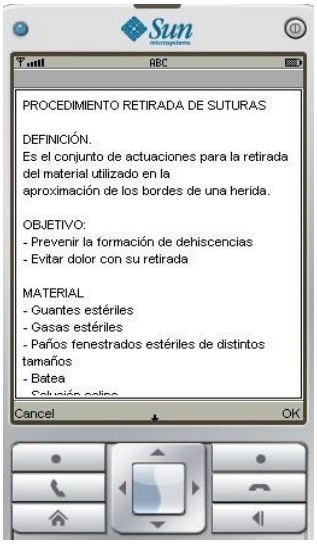


Figura 5c. Visualización de OA en formato de texto

Como se observa también para este caso de estudio, la propuesta planteada es de gran ayuda en la mejora del proceso de enseñanza aprendizaje en el contexto médico. Además de permitir gestionar el conocimiento bajo el paradigma de objetos de aprendizaje se genera una galería de recursos educativos que tiene como ventajas que el médico residente construya conocimiento, a partir de un instrumento de aprendizaje permanente y portable bajo un ambiente de aprendizaje móvil.

6 Conclusiones

En esta propuesta, se describe la necesidad que presentan las entidades médicas en el área educativa. Como solución al problema planteado, se desarrolló una herramienta que permita la visualización de objetos de aprendizaje mediante dispositivos móviles, mediante el cual se alcanza el objetivo educativo de obtener un sistema moderno de aprendizaje apoyado en tecnología el cual permite la construcción de un proceso de enseñanza aprendizaje permanente para los médicos residentes y para los especialistas en las diferentes áreas, creando una red social de aprendizaje en la que se permite gestionar contenidos educativos.

Referencias

1. Andreas Holzinger, Alexander Nischelwitzer, Matthias Meisenberger. “Mobile Phones as a Challenge for m-Learning: Examples for Mobile Interactive Learning Objects (MILOs)”. Proceedings of the 3rd Int’l Conf. on Pervasive Computing and Communications Workshops (PerCom 2005 Workshops).

2. Cao Jiuxin, Mao Bo, Luo Junzhou. “The Self-adaptive Framework of Learning Object Based on Context”. 2008 International Conference on Computer Science and Software Engineering IEEE.
3. Luis de Marcos, José Ramón Hilera, José Antonio Gutiérrez, Carmen Pagés, José Javier Martínez. “Implementing Learning Objects Repositories for Mobile Devices”. Department of Computer Science, University of Alcalá. Alcalá de Henares (Madrid). Spain.
4. Luís A. ÁLVAREZ G, Daniela P. ESPINOZA P., Sandra G. BUCAREY A. “Empaquetamiento y Visualización de Objetos de Aprendizaje SCORM en LMSs de Código Abierto”. Universidad Austral de Chile Valdivia, Chile.
5. René Cruz Flores, Gabriel López Morteo. “Framework para aplicaciones educativas móviles (m-learning): un enfoque tecnológico-educativo para escenarios de aprendizaje basados en dispositivos móviles”. Virtual Educativa 2007. Brasil.
6. Sergio Castillo, Gerardo Ayala. “A Personalization Model for Learning Objects in Mobile Learning Environments”. 16th International Conference on Computers in Education, ICCE 2008, October 27-31, Taipei, Taiwan.
7. René Cruz Flores, Gabriel López Morteo. “A model for collaborative learning objects based on mobile devices”. IEEE DOI 10.1109/ENC.2008.32 Mexican International Conference on Computer Science.
8. Juha Puustjärvi, Leena Puustjärvi. “Managing Personalized and Adapted Medical Learning Objects”. Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007) 0-7695-2916-X/07.
9. Olivier Catteau, Philippe Vidal, Julien Broisin. “Learning Object Virtualization Allowing for Learning Object Assessments and Suggestions for Use”. Eighth IEEE International Conference on Advanced Learning Technologies. DOI 10.1109/ICALT.2008.192.
10. S. Gálvez Rojas, L. Ortega Díaz “Java a tope: Java 2 Micro Edition”. Dpto de Lenguajes y Ciencias de la Computación E.T.S. de Ingeniería Informática, Universidad de Málaga.
11. LWUIT home pagehttps://lwuit.dev.java.net/ [accesed] Octubre 6, 2009.
12. Samtani, G and D. Sadhwani, “Enterprise Application Integration and Web Services,” in Web Services Business Strategies and Architectures, P. Fletcher and M. Waterhouse, Eds. Birmingham, UK:Expert Press, LTD, pp. 39-54,2002.
13. NuSOAP. http://www.scottnichol.com/nusoapintro.htm, [Accessed: April 7, 2009].

Análisis de comportamiento de datos generados por códigos con complejidad O(n)

C. Bustillo-Hernández, L.C. Cota-Gómez, J. Figueroa-Nazuno
Centro de Investigación en Computación
Instituto Politécnico Nacional
Unidad Profesional Adolfo López Mateos
{chbustillo004,lcota,jfn}@cic.ipn.mx
Paper received on 23/07/10, Accepted on 05/10/10.

Resumen. Uno de los aspectos importantes de las técnicas de prueba conocidas para la evaluación de código es su funcionamiento. Sin embargo no se puede probar todos los escenarios posibles en los que el código será utilizado; esto es, existen casos no frontera que provocan comportamiento no esperado. Este trabajo presenta resultados de códigos extraordinariamente simples con comportamiento complicado y la aplicación de diferentes técnicas de análisis, no lineal sobre éstos, utilizando una máquina de Tag. Los sistemas Tag representan el modelo de máquina de Turing universal, fundamento de la Computación Moderna; los cuales a pesar de que su codificación es de complejidad O(n), pueden generar datos con comportamiento complicado.

Palabras Clave: código, validación, técnicas de evaluación, sistemas Tag, máquina de Tag.

1 Introducción

El desarrollo de código implica una serie de actividades específicas previas, tal como el análisis y diseño del algoritmo, y aunque la funcionalidad esperada se vea reflejada en la codificación, algunas veces el resultado de la salida del programa no es el esperado; esto en algunos casos no depende precisamente de errores en el proceso de desarrollo sino en la naturaleza del problema abordado.

Las técnicas de prueba conocidas para la evaluación del código, se enfocan en que el programa cumpla con el propósito planteado, por medio de pruebas sistemáticas que comprueben la lógica interna y verifiquen los dominios de entrada y salida. En base a esto último, los métodos conocidos de ingeniería de software, no son suficientes para evaluar la calidad de los resultados arrojados por programas sumamente sencillos, cuyos datos generados tienen diferentes comportamientos con parámetros de entrada similares.

De programas muy sencillos se pueden obtener datos de salida con varias características distintas y muy complejas.

La contribución de este trabajo, es la aplicación de diferentes técnicas de análisis no lineal sobre datos producidos por la máquina clásica de Tag, que permitirán obte-

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 37-45.

ner una caracterización más completa por medio de valores cuantitativos y estudiar el comportamiento no esperado de códigos simples y cortos.

2 Problema Abordado

En la literatura se ha encontrado que expresiones formales sencillas pueden producir comportamiento complicado o caótico en el sentido estricto[3], tal es el caso de las ecuaciones de Lorenz. Por otra parte en el área de electrónica lo anterior también se ha manifestádo en los circuitos muy simples como el de Chua [8,9].Así mismo Wolfram demuestra que estos comportamientos se presentan en Autómatas Celulares clasificándolos como estácionarios, cíclicos o "muy complicados"[7]. Por otro lado, De Mol ha realizado experimentaciones exhaustivas basadas sólo en ajustes de parámetros [2]. Sin embargo ninguno de los trabajos anteriores efectúa un análisis métrico cuantitativo.

Los sistemas Tag, han sido formalmente demostrados como equivalentes a una máquina de Turing universal (fundamento teórico de la Computación Moderna)[11] y son útiles para el estudio de problemas sobre Decibilidad, Halting, etc.

Un sistema Tag se define parte conjunto finito de símbolos $\{0,1,\ldots..,\mu\}$ y de reglas de la forma $\{0 \rightarrow \sigma_0,\ldots.,\ \mu \rightarrow \sigma_\mu\}$, donde $\sigma_0,\ldots..,\ \sigma_\mu$ es una secuencia de cadenas finitas compuestá de símbolos, incluyendo la cadena vacía, junto con un entero.[1]

El modelo aquí abordado, mismo que Post encontró como intratable y que ha sido ampliamente estudiado, es el sistema con los símbolos {0,1}, las reglas $\{0 \rightarrow 00,\ 1 \rightarrow 11001\}$ y ν =3. El algoritmo de máquina de Tag implementado, se describe a continuación:

Paso 1. Se ingresa una cadena binaria de cualquier longitud (ver Fig. 1)

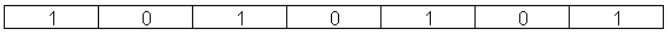


Fig. 1. Cadena binaria ingresada para ejemplificar el algoritmo.

Paso 2. Se verifica el primer elemento de la cadena y se aplica la regla correspondiente de acuerdo a: $0 \rightarrow 00$, $1 \rightarrow$ 11001.

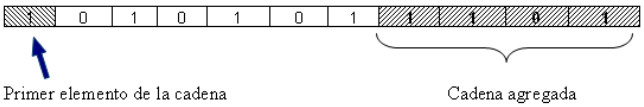


Fig. 2. Selección del primer elemento y cadena agregada.

Paso 3. Cortar ν =3 elementos al principio de la cadena como se ilustra en la Fig.3

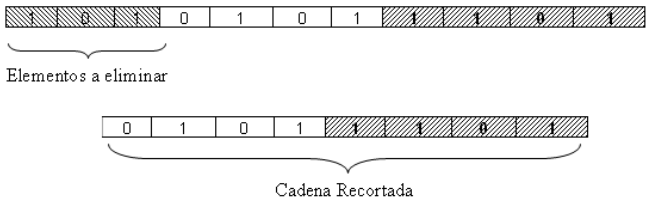


Fig. 3. Corte de la cadena.

Paso 4. Una vez cortada la cadena, se convierte a su representación decimal.

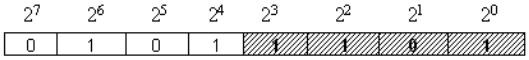


Fig. 4. Cadena recortada con su representación en decimal.

Paso 5. Se repite el paso 1, con la cadena recortada obtenida en el paso 3.

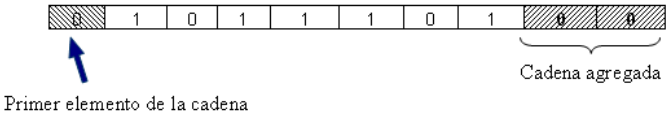


Fig. 5. Cadena obtenida, aplicando regla y corte.

En este trabajo para generar las diferentes cadenas de entrada se definió un ordenamiento sencillo para presuponer cierta estructura previa que pueda tener relación con la salida, tomando en cuenta que las reglas del sistema son fijas. En este caso se utilizó una secuencia númerica muy simple.

Algunas aspectos importantes abordados son: la aplicación de técnicas de análisis no lineal para obtener valores cuantitativos ortogonales que proporcionen una caracterización más completa sobre el comportamiento de las series generadas; la representación decimal de las cadenas generadas, en las que se pueda utilizar técnicas de análisis en el espacio continuo y no en el discreto; se hace un estudio de la representación decimal de las cadenas generadas y no sobre la longitud de éstas. [2,7]

3 Metodología

3.1. Generación de datos

Se realizaron varios experimentos de los cuales se escogió como ejemplo las cadenas de entrada 10010, 11100,100110, 110000 y 111010, correspondientes a los números 18, 38, 48 y 58 respectivamente. Se buscó un ordenamiento sencillo que procurara una relación aditiva de cantidades en 10 entre los números, en otros casos de entrada los resultados son los clásicos que se abordan en [2,7], lo cuales aquí no

se presentan. Las series de salida producidas consisten en 1000 números, convertidos a base decimal desde la cadena binaria.

3.2. Implementación del Algoritmo

A continuación se presenta el pseudocódigo del programa que implementa el algoritmo de Tag anteriormente descrito.

```
for (i=1;i<n;i++)
        if(cadena[0]==0)
            concatena(cadena,00)
        else
            concatena(cadena,1101)
        end-if

    borra_elementos(cadena,v)
    convierte_decimal(cadena)

  end-for
```

3.3. Técnicas de Análisis No Lineal aplicados a los datos

Una vez generadas las series y su representación en decimal de cada número, se obtienen los parámetros de Tiempo de Retraso y Dimensión Embebida para la construcción del Espacio de Fase, utilizando las técnicas de AMI (Average Mutual Information) y FNN (False Nearest Neighbor) respectivamente.
Posteriormente se construye un Mapa Recurrente a partir de Espacio de Fase obtenido y se calculan las propiedades de Porcentaje de Recurrencia, Porcentaje de Determinismo, Entropía, Porcentaje de Laminaridad y Tendencia.

Por último, se aplicó directamente a los datos de las series, la técnica de Análisis Gramatical.

A continuación se describen brevemente cada una de las técnicas utilizadas.

3.3.1 Tiempo de Retraso

El Tiempo de Retraso es usado para calcular la Dimensión Embebida. Sin embargo escoger el Tiempo de Retraso óptimo puede ser problemático. Si el Tiempo de Retraso es muy pequeño, las coordenadas usadas para cada vector reconstruido no será lo suficientemente independiente para llevar cualquier nueva información acerca de las trayectorias del sistema en el Espacio de Fase; si el Tiempo de Retraso es muy grande, las coordenadas podrían convertirse en aleatorias con respecto una de la otra. El método más comúnmente utlilizando para estimar el Tiempo de Retraso es calcular la función de Información Mutua [3].

3.3.2 Dimension Embebida.

Los valores de la Dimensión Embebida (*m*) y el Tiempo de Retraso (*d*) son usados para la reconstrucción en del Espacio de Fase.
La regla es obtener la Dimensión Embebida adecuada, es tal que $m \leq 2*N+1$, donde *N* es el número de variables operativas. En la mayoría de los casos *N* es desconocido y solamente es estimado [4].

3.3.4 Espacio de Fase

Es una técnica que produce gráficas que representan la dinámica de una serie de datos en un espacio *n*-dimensional, una vez determinado el valor óptimo del Tiempo de Retraso y Dimensión Embebida [4].

3.3.5 Mapa Recurrente (MR)

Es una técnica que se basa en una matriz, donde cada [i],[j]-ésimo elemento es calculado como la distancia entre vectores $\vec{V}_i$ y $\vec{V}_j$ de las series de datos reconstruidas en el Espacio de Fase [4].

A través de la estructura en el Mapa Recurrente se pueden obtener diferentes medidas como son: el Porcentaje de Recurrencia, Porcentaje de Determinismo, la Entropía, Tendencia, Porcentaje de Laminaridad, entre otros.

3.3.6 Porcentaje de Determinismo (%DET)

Está métrica mide el porcentaje de [i],[j]-ésimos elementos recurrentes que llegan a formar estructuras [4]. Se puede interpretar como una medida de determinismo en la estructura de los datos y está dada por:

$$DET = \frac{\sum_{l=l_{\min}}^{N} lP(l)}{\sum_{i,j}^{N} R_{i,j}^{m,\varepsilon}} \qquad (11)$$

donde:
$P(l)$ es la frecuencia de distribución de las longitudes l de las estructuras diagonales en el MR y *N* es el número de líneas diagonales.

3.3.7 Porcentaje de Recurrencia (%REC)

Es definida como una medida global de [i],[j]-ésimos elementos recurrentes contenidos en el MR [4], está dada por:

$$\%REC = \frac{1}{N^2}\sum_{i,j=1}^{N} R_{i,j}^{m,\varepsilon} \quad (2)$$

donde:
N es el número total de datos.
i, j son los índices de los elementos del MR.
m es la dimensión embebida
ε es el radio (umbral)

3.3.8 Entropía (ENT)

Está medida se refiere a la entropía de Shannon de la distribución de frecuencias en la longitud de las líneas diagonales [4] y está dada por:

$$ENT = -\sum_{l=l_{\min}}^{N} P(l) In P(l) \quad (3)$$

$$P(l) = \frac{P(l)}{\sum_{l=l_{\min}}^{N} P(l)} \quad (4)$$

donde:

$P(l) = \{l_i; i = 1,...., N\}$ es la frecuencia de distribución de las longitudes l de las estructuras diagonales del MR.

N es el número de líneas diagonales.

3.3.9 Porcentaje Laminar (%LAM)

Está medida se define como el porcentaje de [i],[j]-ésimos elementos recurrentes que forman las líneas verticales [4], y está dada por:

$$LAM = \frac{\sum_{v=v_{\min}}^{N} vP(v)}{\sum_{v=1}^{N} P(v)} \quad (5)$$

Donde $P(v)$ representa la frecuencia de distribución de las longitudes de línea verticales. Cada línea vertical indica que un estádo no cambia o cambia muy lentamente, es decir, el estádo permanece constante durante algún tiempo.

3.3.10 Tendencia (TEN)

Es una medida que indica que tan rápido desaparecen y ocurren cambios desde la diagonal principal. La tendencia, como su nombre sugiere, ayuda a detectar la no estácionalidad en los datos:

$$TREND = \frac{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)(RR_i - RR_i)}{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)^2} \quad (6)$$

Donde $\tilde{N} < N$ y N es el número de áreas diagonales. Si los [i],[j]-ésimos elementos recurrentes están homogéneamente distribuidos a través del MR, el valor de tendencia deberá aproximarse a cero. En cambio, si los [i],[j]-ésimos elementos recurrentes están distribuidos heterogéneamente el valor de tendencia será muy lejano a cero . La tendencia es calculada como la pendiente de la regresión de los mínimos cuadrados del porcentaje local de recurrencia como una función de desplazamientos ortogonales desde la diagonal central [4].

3.3.11 Análisis Gramatical

Está técnica encuentra patrones dentro de una serie de datos, mismos que quedarán representados como reglas de producción dentro de una gramática. Es muy eficaz para la caracterización de una serie de datos, ya que mientras más compleja sea, más reglas de producción se necesitan para construir su gramática. Por lo tanto, el número de producciones es un valor cuantitativo de su complejidad [5].

4. Resultados.

En la Tabla 1 se muestran los valores obtenidos con las técnicas de análisis no lineal.

Se obtiene el valor de Tiempo de Retraso, indispensable para la obtención de la Dimensión Embebida (*m*). En los resultados se caracteriza de forma estricta el comportamiento de los datos, donde si *m*<=3 el fenómeno se clasifica como caótico, en caso contrario se trata de un sistema complejo. De las métricas obtenidas de Mapa Recurrente se puede observar que el Porcentaje de Recurrencia en el conjunto de datos estudiado presentan la misma estructura de repetibilidad, pero sin ser estacionarias.

Los valores altos de porcentaje de Determinismo indican que los datos generados tienen definida una estructura, pero esto no implica que sean predecibles. Los resultados de Entropía presentan valores bajos y de esto se interpreta que hay poco "ruido" en los datos.

Tabla 1. Resultados de diferentes medidas con las Técnicas de Análisis No Lineal, donde porcentaje de Recurrencia (%REC), procentaje de Determinismo (%DET), Entropía (ENT), porcentaje de Laminaridad (%LAM) y Tendencia(TEN)

Cadena Binaria	Número	Comportamiento	Mapa Recurrente (MR)							Análisis Gramatical
			Tiempo de Retraso	Dimensión Embebida	%REC	%DET	ENT	%LAM	TEN	Número de Reglas de Producción
10010	18	Complejo	4	9	16,103	99,789	7,304	0	-1,012	11
11100	28	Ciclico	0	0	0	0	0	0	0	9
100110	38	Caótico	4	2	16,152	99,795	7,349	0	-0,958	12
110000	48	Complejo	4	4	16,081	99,793	7,331	0	-1,077	11
111010	58	Complejo	4	4	16,081	99,793	7,331	0	-1,077	11

Así mismo, los porcentajes de Laminaridad muestran que hay homogeneidad dentro de las series de datos, esto también se observa en los resultados de Tendencia, donde se presenta un grado bajo de desorden.

Por otro lado, en el Análisis Gramatical se observa que el caso con menor número de reglas de producción corresponde a la serie de datos que se clasificó como ciclica ya que su dimensión embebida es cero.

En general, se puede observar que aunque las medidas son ortogonales entre sí, proporcionan cierta congurencia en la interpretación de los resultados.

Todos los resultados nos indican que aunque entre las series de datos se da una variabilidad de comportamientos, cada una de ellas presenta cierta estructura instrínseca.

5. Conclusiones.

La máquina de Tag es un procedimiento muy simple que puede ejemplificar el comportamiento de ciertos programas que se retroalimentan, que bajo ciertos parámetros pueden provocar salidas no esperadas. Esto se pudo caracterizar con diferentes técnicas de análisis no lineal mediante valores cuantitativos.

Se demostró que la estructura de cada una de las series generadas no tiene ninguna relación con la secuencia numérica presupuestá entre las cadenas iniciales, a pesar de que el procedimiento consiste en reglas fijas. Por lo tanto la máquina de Tag, así como otros códigos simples, pueden no ser deterministas y sin embargo no producir aleatoriedad.

Los resultados son congruentes con la Teoría del Caos moderna que ha demostrado que en sistemas matemáticos, electrónicos o de autómatas celulares, que son modelos totalmente deterministas se pueden obtener comportamiento caótico e impredecible [3].

El análisis de máquina de Tag en una de sus formas más simples, demuestra que este problema muy pequeño y determinista produce comportamiento que tiene estructura y diferentes características de comportamiento caótico.

Referencias

1. Post, E. L.: Formal Reductions of the General Combinatorial Decision Problem. American Journal of Mathematics 65(2) 197-215
2. De Mol, L.: Tracing Unsolvability: A Mathematical, Historical and Philosophical analysis with a special focus on Tag Systems. PhD thesis, University Gent (2007)
3. Kantz, H. y Schreiber, T.: Nonlinear Time Series Analysis. University Press Cambridge (2005)
4. Takens, F.: Detecting Strange Attractors in Turbulence. Lecture Notes in Mathematics 898 366-381
5. Delfín-Santiesteban, O. R.: Análisis de la Predictibilidad de Series de Tiempo usando Algoritmos de Extracción de Reglas Gramaticales. Tesis Maestría, Centro de Investigación en Computación, Instituto Politécnico Nacional (2006)
6. De Mol, L.: On the boundaries of Solvability and Unsolvability in Tag Systems. Preprint submited to Elsevier (2008)
7. Wolfram, S.: A New Kind of Science. Wolfran Media Inc. (2002)
8. E. Bilutta, P. Pantano.: A gallery of Chua Attractors, serie A. (61). World Scientific on Nonlinear Science (2008)
9. L. Fortuna, M. Frasca, M. Gabriella Xibilia,: Chua´s Circuit Implementation. Yesterday, Today and Tomorrow. serie A. (65), World Scientific on Nonlinear Science (2009)
10. De Mol L.: Tag systems and collatz-like functions. Theoretical Computer Science 390(1) 92-101 (2008)
11. Minsky M.: Recursive Unsolvability of Post's problem of 'Tag' and other Topics in Theory of Turing Machines. Annals of Mathematics 74 (3) (1961)

Developing a SOA-based m-commerce Android application

Jorge Fernando Ambros-Antemate, Giner Alor-Hernandez, Ulises Juarez-Martinez, Ana Maria Chavez-Trejo

Division of Research and Postgraduate Studies,
Instituto Tecnológico de Orizaba, México
jfambros@gmail.com,{galor, ujuarez, achavez}@itorizaba.edu.mx
Paper received on 01/07/10, Accepted on14/09/10.

Abstract. The breakthrough in wireless technology increases the number of users with cellular phones, which presents a scenario for the development of m-commerce applications. Nowadays these devices are able to manipulate more quantity of information allowing adapt technologies of communication such as Web services. Web services were available only in personal computer applications, but thanks to technological advances presented by mobile communication devices, Web services are available on mobile devices. This paper proposes a service-oriented architecture for an m-commerce application developed under the Android platform.

Keywords: Android, SOA, SOAP, Web services.

1. Introduction

The mobile commerce (m-commerce) is any electronic transaction of goods, products or services through mobile telecommunications networks. The key to distinguish between m-commerce and e-commerce is the use of a mobile device instead of using a Web browser from a personal computer.

Currently there is a growing number of cell phone users, which opening a potential market for m-commerce applications, such operations have the features that the user accesses the Internet from any location at any time from a mobile communication device to make the purchase of goods and / or services. The essence of m-commerce revolves around the idea of reaching consumers, suppliers and employees regardless of their location. M-commerce applications offer advantages not available in traditional e-commerce applications [1]: (1) **Ubiquity** is the main advantage of m-commerce. Users get any kind of information, on which they are interested at

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp.47-57.

any time, (2) **Accessibility**, through mobile devices, business entities are able to reach consumers anywhere. (3) **Location**, applications developed in localization offer significant value to m-commerce, knowing the user's location offering new services, such as displaying user preference products just by being close to the company.

To improve this kind of applications the application development platforms play a very important role, especially adapting to **new** technologies that improve application performance, such as web services that offer full interoperability between applications through the use of open standards, this allows the data contained in e-commerce server could be displayed directly on a mobile commerce application. Based on this understanding, this paper proposes a Service Oriented Architecture (SOA) using Web services for an m-commerce application developed under the Android platform.

This paper is structured as follows: section 2 presents an introduction to the Android development platform. Section 3 describes the m-commerce Android application and the architecture to use. Section 4 presents a case study that describes the functionality of a developed m-commerce application. Section 5 discusses related work. Finally we present the conclusions of this paper.

2. Android platform

Android is a software stack for mobile devices including an operating system, middleware, user interface and applications. The Android Software Development Kit provides the tools and APIs necessary to develop applications on the platform using the Java programming language [2]. Each Android application runs in its own process, with its own instance of the Dalvik virtual machine. The platform enables the device to run multiple virtual machines efficiently and optimize files with Dalvik Executable format (. Dex) for minimum memory consumption.

Currently there are several platforms for developing of mobile applications; Table 1 presents a comparison among Android and the most popular platforms.

Table 1. Comparison of development platform for mobile phones.

	Android	J2ME	iPhone	Symbian
Developer	Google	Sun controls spec	Apple	Nokia
Programming language	Java	Java Micro Edition	Objective C	C++
Open Source Licence	Not	Not	Yes	Not
API for the use of hardware components	Yes	Yes	Yes	Not

As we can see in Table1, each platform provides different mechanisms for developing applications. For example, Apple constraints to developers releasing the contents of the SDK, J2ME provides a good abstraction for programming, but it

does not use a shortcut to the capabilities of the device. Symbian is Open Source, however it constraints the access to the hardware components of the device and does not use a standard C++ version, each company will make changes to the SDK for adapt it to the cell phone. Finally, Google and Open Handset Alliance offer Android source code under an Apache license, giving the facility to cell phones manufacturers to integrate their own changes to the device, this license allows the developer to modify the source code of the platform when the APIs included do not provide some functionality.

However, Android also has disadvantages:

1. As a newly developed platform has had little penetration and this implies that cell phones are expensive.
2. As an open platform allows examining the source code to find vulnerabilities.

3. SOA with Android

A SOA is a combination of consumers and services that work together, it is guided by principles and supports various standards [3], unlike traditional information systems which have their embedded business process. SOA separates these processes from automated functions and organized them into individual modules, listed in a dictionary of services that allow its use by the entire organization.

The key of SOA architecture is "processes abstraction", this allows the execution, management, monitoring and process changes are handled directly at business level in a versatile way, rather than being embedded [4]. But despite of the benefits mentioned previously, cell phones do not integrate technology for the use of Web services (e.g. SOAP), because many of them have limited processing capabilities and are unable to process data obtained from Web services.

The mobile devices with the Android operating system does not support SOAP processing directly to enable it, the kSOAP library is required [5] and must be included in a project. Figure 1 shows the relationship between the application and kSOAP, where Activities (main classes in Android applications) communicate directly with kSOAP to handle HTTP requests and responses to the Web services invocations.

Our m-commerce system uses a service-oriented architectural style providing the following benefits:

1. Using industry standards like XML, WSDL (Web Services Description Language) and SOAP: This ensures interoperability between different services and applications that use them.
2. The communication protocols used by Web services are independent of the operating system, platform and programming language.
3. The services to be loosely coupled enable that the applications that use them will be easily scalable because there are little dependence between these elements.

4. There is a high reuse to enable the services are used by the application of mobile commerce and other applications.

The proposed architecture has a layered-design to allow the exchange of data, in this way each layer provides its services to its adjacent layers. In Figure 2 show these components.

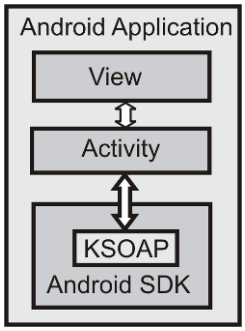


Fig. 1. kSOAP library in Android application.

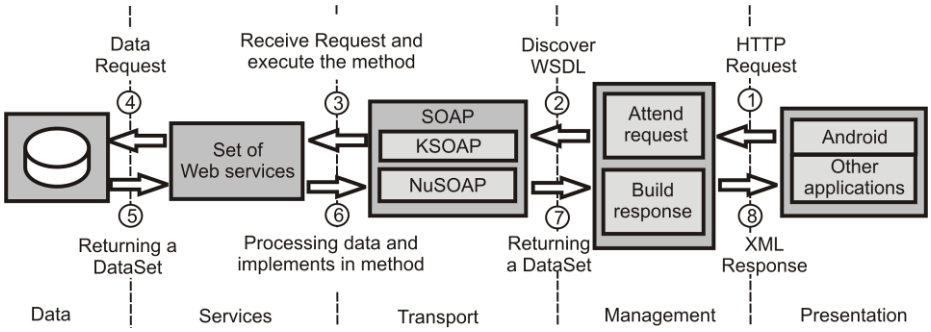


Fig. 2. Layered-design of the m-commerce application of the SOA architecture.

The layers are described below:

Presentation: This layer is responsible for managing the data flow from the user, the application consists of a light client responsible for interacting with the service provider, the application is designed with the pattern Model - View - Controller (MVC), the Model (Activities in the application) handles all requests made by the Controller and is responsible for the handling of data obtained from the functions library kSOAP, to extract each one of the ele-

ments contained in the data a SoapPrimitive class is used and then are presented through the views on the device.

Management: This layer has as fundamental purpose managing SOAP requests and providing access to WSDL documents, also classify, catalog and manage Web services so they are discovered and consumed by applications.

Transport: This layer is responsible for managing Web services using SOAP, the Web server uses a group of classes called NuSOAP to develop Web services with PHP language, the client uses kSOAP which is a function library of SOAP-based Web services for Java environments, for Android applications kSOAP use a version designed for this operating system [5].

Services: This layer is responsible to contain Web services. All business transactions are done through Web services (responsible for exposing all the functionality of the system to the pone and other applications). In order to develop each Web service, Internet standards are used: XML as the standard format for data, SOAP as a protocol for data exchanging and WSDL documents to describe the public interface of services.

Data: This layer is responsible of managing the data storage in e-commerce application and the mobile commerce.

The application functionality is explained below:

Step 1. The Android application makes an HTTP request through kSOAP, it specifies the URL of the WSDL document location.

Step 2. KSOAP discover on the Web server the WSDL document that contains the method requested by the application.

Step 3. The Web server receives the request by NuSOAP, the business logic processes the request and executes the method requested by the customer.

Step 4. The business logic communicates with the database server.

Step 5. The database server receives the request, processes it and returns the requested records.

Step 6. The result of the database server is processed and is implemented in the method by the business logic.

Step 7. NuSOAP builds the response and returns the data via web service using SOAP as an XML document to the library of functions kSOAP.

Step 8. KSOAP gets the XML document and transforms it to an object with the structure element = anyType (key0; key1 = element1; ...; Keynes-1 = elementN-1), to extract each one of the elements a special treatment is required using SoapPrimitive class, in this way the keys are obtained separately along with their elements and are used in activities.

In the next section presents a case study, which shows the functionality of the application.

4. Case study: looking for products in an application of m-commerce

The m-commerce application is based in the procedures for electronic purchasing proposed by Blommestein[6]. Figure 4 shows the whole procedure, the area dotted shows the processes that implements the m-commerce application, Delivery and invoicing and payment are relate to internal processes of the company

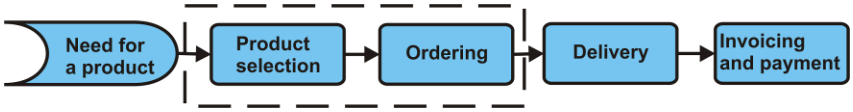


Fig. 4. Procedures for electronic purchasing proposed by Blommestein.

Figure 5 show in the area dotted the parts of the process of product selection implemented in our m-commerce application.

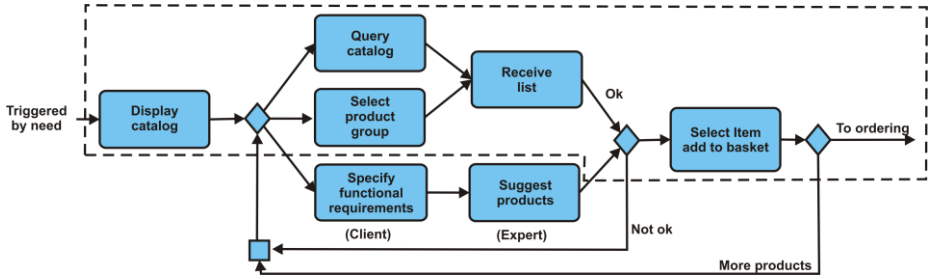


Fig. 5. The product selection process.

The processes are explained below:

1. Display Catalog: This process shows a product list offered by the enterprise. This product list can be displayed in a hierarchical way.
2. Query catalog: The client can search in the catalog using keywords, when a search query is entered the result is a product list.
3. Select product group: The products groups can be classified according to functional and technical characteristics.
4. Receive list: The product list required by the client.
5. Specify functional requirements (client) and Suggest products (Expert): The client can call on help from a product expert in the buyer or supplier organization. To do this, the client formulates a query in which he indicates the required product function. The expert concerned sends him by return a product list or groups that meet his requirements.

6. Select item, add to basket: The client selects the required product and adds the product to his shopping basket.

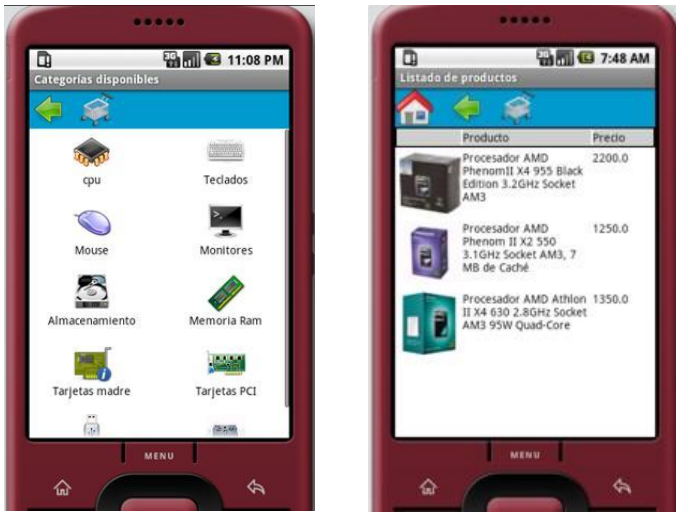
The following case study describes how the m-commerce application carries out the product selection process. Figure 6 shows the e-commerce portal which offers a catalog of products. In this case study, we covered the following premise: How does a user using a phone with Android operating system and m-commerce application retrieves products with the same features of e-commerce?

The process begins when the Android application invokes a Web service through an HTTP request, the function library kSOAP takes the URL and sends the request to the server where there are implemented the functions library NuSOAP, it communicates with the business logic and runs the appropriate method for the search of categories, to obtain records and builds the response is sent through the Web service in XML format to the library of kSOAP functions of the Android application, it extracted each one of the elements and are presented in the user view. Figure 7 (a) shows the result of this invocation.

Fig. 6. The electronic commerce portal in a Web browser

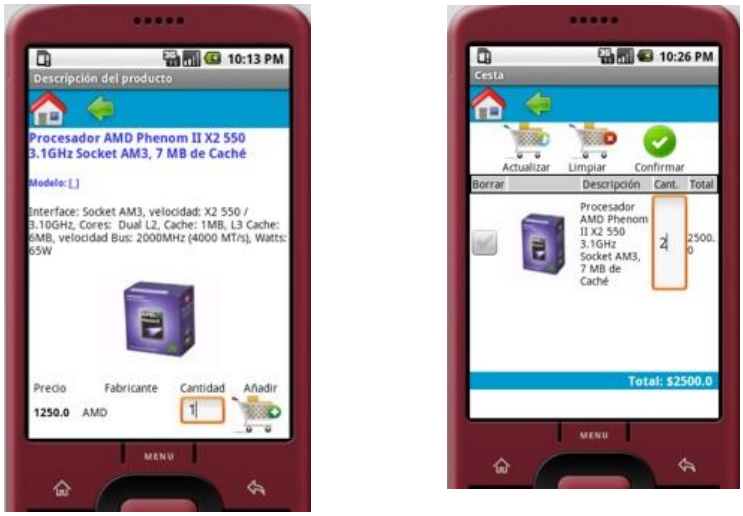
Figure 7 (b) shows the list of products available related to the category; when the client selects a product of the list, the complete description of the product is displayed, figure 8 (a) show the result of this selection. Figure 8 (b) shows the shopping cart when the user can add or delete products and clean the basket.

Once confirmed the shopping cart the client proceed to carry out the order, figure 9(a) shows the e-mail and password validation, the figure 9(b) shows the order summary.



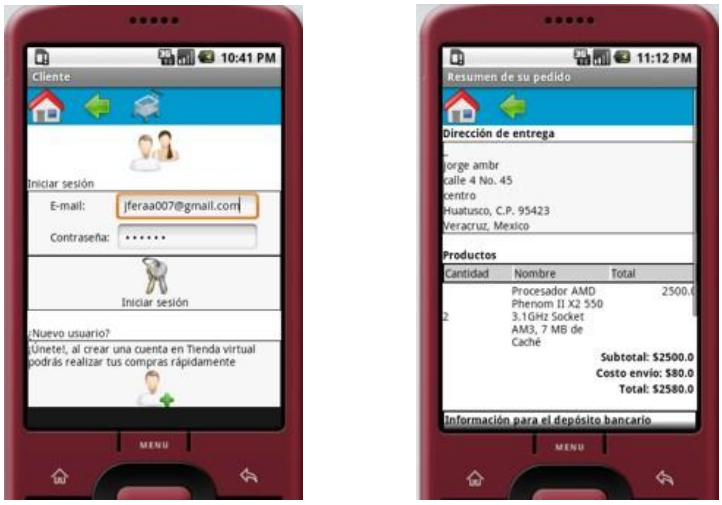
(a) (b)

Fig. 7. (a) Products categories, and (b) Products List by category of m-commerce application.



(a) (b)

Fig. 8. (a) Product description, and (b) basket.



(a) (b)

Fig. 9. (a) E-mail and password validation , and (b) order summary.

The m-commerce application is helpful to the user since provides a similar functionality of an e-commerce application. With an m-commerce application the user can performs the purchasing process of products available in stock by using an Android phone without requiring a desktop computer

Next section shows the related work using a SOA in mobile devices.

5. Related works

Currently, the amount of mobile phones are increasing daily, this allows for a new platform for business applications so that these devices need to adapt to technologies and communication protocols, below it presented works that are related with mobile applications that are developed under a SOA.

In [7] a review of kSOAP and J2ME Web services APIs (WSA or JSR 172) is presented for the use of Web services on mobile devices, also described a case study with a BlackBerry, with BlackBerry MDS Studio it creates enabled applications to communicate using a simple and compact protocol with BlackBerry Enterprise Server (BES) using Web services. In [8] a review of an application called WIPdroid is presented, the application uses Web services and real-time communication that allows to mobile computing platform support for distributed computing services, the tool gSOAP is integrated developed in C/C++, and it supports multiple platforms including embedded systems and small operating systems. In [9] a review of a framework is presented for a service-oriented using the REST technology (Representa-

tional State Transfer), in this framework, mobile applications communicate with business applications, plus inserts the execution of local and remote services, provides companies the ease to control services and data that users access through their mobile devices. In [10] a review of a SOA is presented for mobile devices through a COM implementation, for the tests using a Symbian S60 phone to demonstrate the effectiveness of the SOA for mobile applications. In [11] a review of a SOA framework is presented for mobile services, the objective is to investigate how the construction of mobile services is benefited from Service-Oriented paradigm. The framework designed by components provides services, giving the facility to be replaced by other instances or implementations. A review of an architecture for Web services consumption on mobile devices is presented in [12]. J2ME Web Services APIs are implemented on the device to provide access to Web services using SOAP. In [13] a review of a mobile Lightweight architecture SOA is presented for business applications running on devices with J2ME. It is a framework in which objects sent to the mobile device are serialized, compressed and transferred to the client as a message RDF. Communication is established using SOAP messages and all work is done through a Java midlet that exchanges data asynchronously with the server, minimizing the data transferred and stored on the device. Finally in [14] a review of a location-based mobile application is presented to enrich the experience of tourists where the application uses the principles of web service oriented architecture and data providers to support interactions among tourists.

6. Conclusions

This paper presents a service-oriented architecture for an Android application. The architecture provides a new way of communication for mobile commerce applications by invoking Web services using SOAP, the whole implementation is done only with open source functions libraries such as NuSOAP and kSOAP. This is expected that companies interested in trade Android mobile phones implement this architecture to offer customers an alternative to buying new products or services.

Acknowledgements

This work is supported by the General Council of Superior Technological Education of Mexico (DGEST). Additionally, this work is sponsored by the National Council of Science and Technology (CONACYT) and the Public Education Secretary (SEP) through PROMEP.

References

1. Lim E. Siau K. Advances Mobile Commerce Technologies. United States of America: Idea Group Publishing. 2003.
2. What is Android? | Android Developers, Octubre 2009. [on Web]. Available: http://developer.android.com/guide/basics/what-is-android.html. Date of consultation: October 2009.

3. Bean J. SOA and Web Services Interface Design. United States of America: Elsevier Inc. 2010.
4. Fernández J., Surroca A. Cómo reformular la Arquitectura Corporativa para alcanzar el alto Rendimiento. Centro de Alto rendimiento: 2008.
5. ksoap2-android, agosto 2009. [on Web]. Available: http://code.google.com/p/ksoap2-android/. Date of consultation: October 2009.
6. Blommestein F., "Procedures for electronic purchasing". Leidschendam, 2000.
7. Kozel T., Slaby A., "Mobile devices and Web services". 7th WSEAS International Conference on APPLIED COMPUTER SCIENCE. 2007.
8. Chou W., Li L.. "WIPdroid–A Two-way Web Services and Real-time Communication Enabled Mobile Computing Platform for Distributed Services Computing". 2008 IEEE International Conference on Services Computing. 2008.
9. Ennai A., Bose S. "MobileSOA: A Service Oriented Web 2.0 Framework for Context-Aware, Lightweight and Flexible Mobile Applications".
10. LIU, Z., GAO, X., "SOA Based Mobile Device Test". 2009 Second International Conference on Intelligent Computation Technology and Automation. 2009.
11. Thanh D., Jorstad I., "A Service-Oriented Architecture Framework for Mobile Services". Advanced Industrial Conference on Telecommunications/Service Assurance with Partial and Intermittent Resources Conference/ELearning on Telecommunications Workshop. 2005.
12. Tergujeff T., Haajanen J., Leppänen J., Toivonen., "Mobile SOA: Service Orientation on Lightweight Mobile Devices". 2007 IEEE International Conference on Web Services. 2007.
13. Natchetoi Y., Kaufman V., Shapiro A., "Service-Oriented Architecture for Mobile Applications". SAM'08 ACM. 2008.
14. Paganelli F., Parlanti D., Francini N., Giuli N., "A SOA-Based Mobile Guide to Augment Tourists' Experiences with User-Generated Content and Third-Party Services," iciw, pp.435-442, 2009 Fourth International Conference on Internet and Web Applications and Services, 2009.

Uso de inteligencia colectiva en el desarrollo de una aplicación Web 2.0 empleando técnicas de búsqueda basada en contenidos para la recuperación de mamografías digitalizadas

Yuliana Pérez-Gallardo, Giner Alor-Hernández, Ruben Posada-Gomez, Guillermo Cortes-Robles

División de estudios de Posgrado e Investigación,
Instituto Tecnológico de Orizaba
Av Oriente 9 No. 852, 94320, Orizaba, Ver.,
México. Télefono/Fax: +52(272) 725 7056
{gallardo.yuli}@gmail.com, {galor, rposada, gcortes}@itorizaba.edu.mx
Paper received on 02/07/10, Accepted on 21/09/10.

Resumen. En la actualidad los motores de búsqueda tradicionales tales como Google han facilitado enormemente la recuperación de imágenes sobre la Web. Sin embargo, los resultados obtenidos por éstos no son a menudo útiles para los usuarios, por lo general debido a dos problemas: (1) no es posible definir una consulta acerca del contenido de una imagen, y (2) La semántica de las palabras clave no se considera, existiendo problemas de polisemia y sinonimia. Este trabajo presenta una arquitectura basada en capas con el fin de desarrollar una aplicación Web 2.0 que use técnicas de búsqueda basada en contenidos para la recuperación de imágenes a través de Inteligencia Colectiva. Como caso de estudio, los usuarios realizan búsquedas de imágenes médicas (mamografías digitalizadas) intentando eliminar los problemas semánticos y de contenido.

Palabras Clave: CBIR, Web 2.0, Inteligencia Colectiva, Mamografías.

1. Introducción

Actualmente los motores de búsqueda tradicionales facilitan de gran manera la recuperación de imágenes sobre la Web, ejemplos de ello son Google y Bing. Sin embargo, los resultados recuperados por éstos motores no son a menudo útiles para los usuarios, comúnmente debido a dos problemas: en primer lugar 1) la semántica de las palabras clave no se toma en consideración, cuando el usuario proporciona a

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp.59-69.

un motor de búsqueda tradicional, una palabra clave ambigua, la respuesta consiste en todos los elementos relacionados con los diversos significados de la palabra, mientras que el usuario está probablemente interesado en sólo uno de ellos. Esta ambigüedad se genera por la polisemia, que es la propiedad que poseen las palabras que, en distintos contextos tienen varios significados diferentes 69], y la sinonimia, que es la relación semántica de identidad o semejanza de significados entre determinadas palabras (llamadas sinónimos) u oraciones [2]. En segundo lugar, 2) no es posible definir una consulta acerca del contenido de una imagen, se necesita que los resultados de recuperación provean las imágenes que una persona seleccionaría, al buscarlas por sí mismo en un repositorio donde se almacenan. Esta situación presenta un problema, ya que es difícil de codificar en un algoritmo de manera automática, la idea que tiene una persona sobre el significado semántico de una imagen. Las personas tienden a describir las imágenes basándose en los objetos que están representados en ellas, este tipo de descripción produce resultados de correspondencia entre la percepción humana y el contenido de la imagen.

Basado en la problemática anterior, este trabajo propone una aplicación Web 2.0 que use técnicas de Inteligencia Colectiva para la búsqueda de imágenes basadas en contenido. Para comprender mejor la importancia de la propuesta, se describe en la sección 2 los trabajos relacionados, los principales conceptos que aborda este trabajo en la sección 3, la arquitectura del sistema y el caso de estudio en las secciones 4 y 5 respectivamente y finalmente se presenta en las secciones 6 y 7 el trabajo a futuro y conclusiones.

2. Trabajos relacionados

Dada la importancia que tiene la Inteligencia Colectiva en la actualidad, existen múltiples investigaciones dentro de los diferentes campos en los que se utiliza la misma. La recopilación y análisis de artículos relacionados con la propuesta permite obtener información relevante y adquirir conocimiento acerca de técnicas de desarrollo y enfoques en los que en la actualidad se trabaja. En [3] se propone el diseño de nuevos modelos de negocio dentro del comercio electrónico, para los integradores de servicios utilizando los principios de Inteligencia de Enjambre (*Swarm Intelligence* en inglés) para proveer medios que aprovechen la Inteligencia Colectiva de los clientes, de esta forma se desarrollan nuevos productos y servicios basándose en las preferencias de sus clientes maximizando las posibilidades de éxito, tal como actualmente lo hacen empresas como Adidas, Lego Factory y Amazon. En [4] presenta el uso de IC para la colaboración intercultural. Esta investigación se enfoca en el obstáculo que todavía representa el lenguaje en la comprensión mutua entre sociedades. En los sistemas de traducción automática la traducción no es perfecta, debido al desajuste que existe entre los significados de las oraciones y/o palabras dentro de las comunidades. Proponen la producción colaborativa de materiales de traducción personalizadas, tomando en cuenta a la propia comunidad nativa para que colabore en el desarrollo del material de traducción de su lengua a otras. Dentro del enfoque de CBIR, en [5] proponen que mediante la agrupación de documentos etiquetados en una folksonomía, es posible extraer los conjuntos de etiquetas relacionadas con los diferentes contextos en donde existen etiquetas ambiguas. Con su propuesta intentan

solucionar dicha ambigüedad construyendo clasificadores que usan datos recolectados desde sistemas de etiquetado colaborativo. Como pruebas, realizaron experimentos en búsquedas de Google, en donde los resultados muestran que el método es capaz de clasificar los documentos devueltos con alta precisión. Utilizan el algoritmo de k-vecino más cercano para clasificar los documentos devueltos por el motor de búsqueda. En [6] se enfocan en la extracción de elementos de contenido de bajo nivel. Se extraen las características que representan el contenido de las imágenes en vectores y para medir su similitud utilizan la distancia euclidiana. Dentro de este mismo enfoque proponen una prueba de correlación en la distancia convencional que corresponde al mecanismo de recuperación de imágenes basado en contenido. En [7] se propone un sistema de recuperación de imágenes basada en una búsqueda combinada de textos y contenidos. La idea es utilizar el texto presente en el título, descripción y etiquetas de las imágenes para la mejora de los resultados obtenidos de la búsqueda. La búsqueda basada en texto proporciona resultados con similitud semántica, mientras que la búsqueda basada en contenido proporciona resultados con similitud gráfica. Debido a la independencia entre estos enfoques, su combinación mejora el rendimiento de un sistema de búsqueda por el beneficio de ambos enfoques. Finalmente, en [8] se presenta un sistema de CBIR basado en componentes configurables. La principal novedad reside en su contenido basado en búsqueda de imágenes de componentes (CBISC) que apoya las consultas sobre las colecciones de imágenes. CBISC se basa en la Iniciativa de Archivos Abiertos (OAI), permitiendo así consultas planteadas a través de peticiones HTTP y las respuestas codificadas en XML.

3. Web 2.0, Inteligencia Colectiva y CBIR

La Web ha experimentado una evolución a través del tiempo [9] para cubrir las necesidades de los usuarios, en sus orígenes con la Web 1.0 con páginas HTML estáticas, pasando por la Web 1.5 hasta ahora con la Web 2.0, en donde las aplicaciones Web se centran en los usuarios invitándolos a participar activamente interactuando, mostrando su punto de vista y preferencias, de tal forma que convierten a los usuarios en colaboradores, generando en contenido que ellos mismos consumen. Así, la Web 2.0 son todas aquellas utilidades y servicios de Internet que se sustentan en una base de datos, la cual es modificada por los usuarios del servicio, ya sea en su contenido (añadiendo, asociando, cambiando o borrando información), en la forma de presentarlos, o en ambos [9]. Pero, surge una pregunta ¿Qué hacer con todos los datos recolectados por la participación e interacción de los usuarios? Esta información se pierde si no se convierte en inteligencia y se canaliza hacia la mejora de la aplicación.

Lévy [10] sugiere que la Inteligencia Colectiva (IC) es una especie de sociedad anónima en donde cada accionario aporta como capital su conocimiento, sus conversaciones, su capacidad de aprender y enseñar. La suma de inteligencias no se somete ni se limita a las inteligencias individuales, sino por el contrario, las exalta, las hace

fructificar y les abre nuevas potencias, creando una especie de cerebro compartido. Debido a los beneficios que nos brinda la suma de las inteligencias individuales, éstas son aplicadas a la Recuperación de Imágenes Basada en Contenido (CBIR, en inglés Content Based Image Retrieval) para tener un mejor resultado en las búsquedas. CBIR es un sistema que recupera imágenes basadas en sus propiedades visuales. El objetivo de las técnicas de consulta basadas en contenido de las imágenes es encontrar, de forma eficiente, las imágenes en una base de datos que son similares a un criterio de búsqueda. A diferencia de las típicas consultas en bases de datos, en este caso se utiliza un criterio de similitud que no tiene por qué ofrecer una coincidencia exacta. Uno de los elementos más importantes en CBIR es la extracción y la evaluación de la medida de similitud de las características de la imagen. Las características de la imagen son aquellos elementos que describen el contenido de una imagen. El significado de las características depende del contexto en el que se manejen, existen características visuales que se derivan directamente de la imagen, el ejemplo más simple es el valor de intensidad de los pixeles de la imagen, aunque este tipo de características visuales son muy sensibles al ruido, brillo, tono y saturación de los cambios. También hay otras características que son específicas de los dominios de aplicación y requieren cierto conocimiento especial. Una buena característica contiene suficiente poder discriminante para distinguir entre las imágenes similares y diferentes y es invariante a la transformación espacial, como la traslación, rotación y cambios menores relacionados con el medio ambiente de iluminación, donde se captura la imagen.

4. Arquitectura del sistema

Para esta propuesta, se propone el desarrollo de una aplicación Web 2.0 que use técnicas de IC para la recuperación de mamografías digitalizadas basadas en su contenido. En la Figura 1, se presenta la arquitectura basada en capas de la propuesta y se describe cada uno de los componentes que la integran. En esta arquitectura, cada capa tiene una función específica que a continuación se describe:

1. **Capa del Cliente:** Está integrada por los módulos a) Cargar imagen, b) Retroalimentación, c) Solicitud de búsqueda, d) Criterio de búsqueda, e) Sugerencia Semántica y f) Resultado de la consulta. Esta capa brinda la interacción con el usuario con el fin de permitirle acceder a los servicios que ofrece la aplicación.
2. **Capa de Control de Imágenes:** Contiene los módulos a) Análisis de la imagen y b) Controlador de la imagen. Esta capa se refiere a las diferentes operaciones que el sistema lleva a cabo con respecto al análisis de las imágenes, por ejemplo la binarización de las imágenes, obtención de regiones de interés, almacenamiento de imágenes y características, entre otras.
3. **Capa de Etiquetado:** Consta de los módulos a) Etiquetado, b) Asignación del vector de características y c) Actualización de características de la imagen. En esta capa se lleva el control del etiquetado colectivo de las imágenes.
4. **Capa de Recuperación de Imágenes:** Integrada por los módulos a) Análisis de características, b) Algoritmo de similitud, c) Búsqueda de imágenes

con características similares, d) Análisis semántico y e) Imágenes. Esta capa contiene todos los módulos necesarios para la recuperación de imágenes basados en contenido y su análisis semántico.

5. **Capa de Datos:** En esta capa se maneja la persistencia tanto de los datos como de las imágenes.

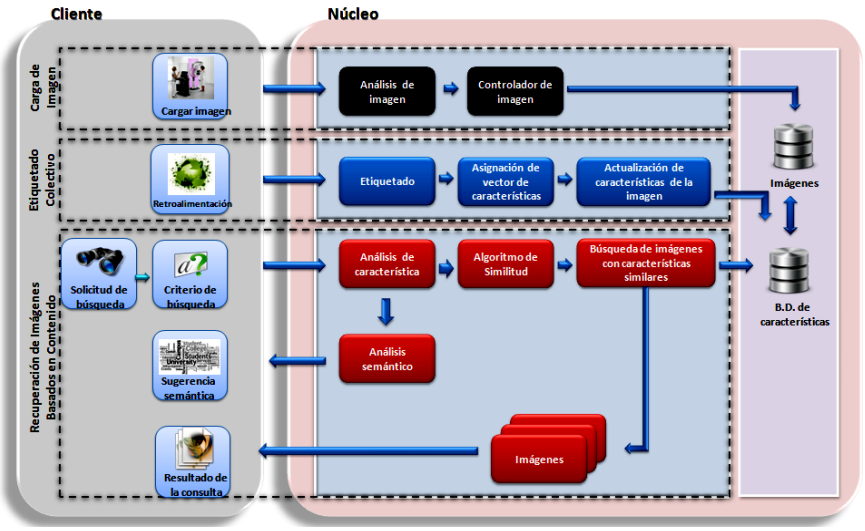


Figura 1. Elementos internos de la arquitectura.

Las capas interactúan entre sí para brindar el funcionamiento total del trabajo propuesto, por medio de diferentes flujos de trabajo. Los flujos de trabajo que contiene el sistema son: carga de imagen, etiquetado colectivo y recuperación de imágenes basados en contenido.

4.1. Flujos de trabajo

A continuación se describe la funcionalidad de los flujos de trabajo que conforman la propuesta.

Carga de imagen. Este flujo de trabajo tiene por objetivo la gestión de las genes. Está integrado por los módulos análisis de la imagen y controlador de la gen. Esta propuesta será empleada para la recuperación de mamografías sadas en contenido. Las imágenes que se emplean son de tipo ICS, que es el Estándar para Imágenes de Citometría (Image Cytometry Standard, en inglés).

formato ICS se propuso por [11] en 1990 y es capaz de almacenar 1) Imágenes tidimensionales y multicanal, 2) Imágenes en 8, 16 o 32 bits enteros, 32 o 64 flotante y datos punto flotante complejos y 3) Parámetros tomados dimamógrafo acerca de la formación de la imagen. El formato original ICS archivos independientes: un archivo de encabezado de texto con extensión .ics otro, con los datos de la imagen real, con la extensión .ids. El flujo de trabajo za cuando el usuario dentro del módulo Cargar imagen, carga una imagen con mato ICS y la envía al servidor al módulo de Análisis de la imagen en donde, lo muestra la

Figura **2.** Detección de características, se obtienen las características, por ejemplo micro calcificaciones o tumores; esto se lleva a cabo aplicando el método de Otsu al separar dentro de la imagen el objeto de interés y el fondo, obteniendo así una imagen binarizada, después se aplica un filtro de erosión y otro de dilatación para tener una mejor definición de los contornos de los objetos, por último se analiza cada pixel y sus vecindades para definirlos o no como áreas de interés.

Una vez obtenidas se envían al módulo de Controlador de imagen, que se encarga de almacenarlas en la base de datos de imágenes y caracteríticas respectivamente. Para manipular las imágenes con formato ICS, se utiliza la librería ImageJ, que es una biblioteca de funciones para el procesamiento de imágenes basado en Java y desarrollado en el Instituto Nacional de Salud de EE.UU. (NIH, National Institutes of Health, en inglés). ImageJ es una plataforma ideal para desarrollar, probar nuevos algoritmos y técnicas de procesamiento de imágenes, también se utiliza en la investigación y desarrollo de aplicaciones en muchos laboratorios del mundo, especialmente en imágenes biológicas y médicas.

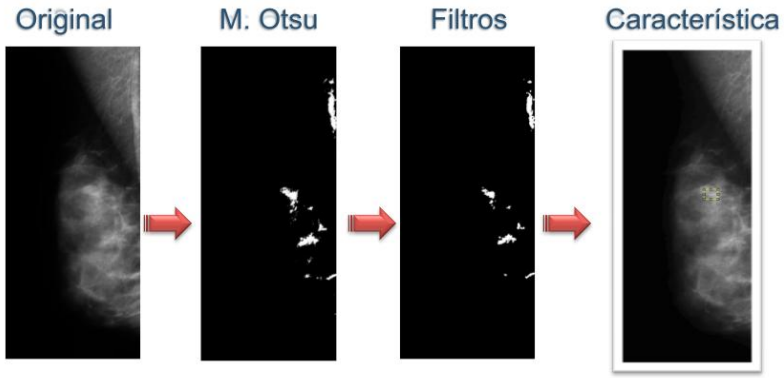


Figura 2. Detección de características

Etiquetado colectivo. Este flujo empieza en el módulo de Retroalimentación en la capa del cliente, la retroalimentación también denominada feedback es el proceso

de compartir observaciones y sugerencias, con la intención de recabar información, a nivel individual o colectivo, para mejorar el funcionamiento de un proceso. Dentro del contexto de la propuesta, se utiliza la retroalimentación del usuario al agregar o modificar etiquetas a las imágenes con el fin de describir su contenido.

El objetivo de este flujo de trabajo es realizar un sistema de etiquetado colaborativo, entro del módulo de Etiquetado y por medio de una intefáz se muetran las mamografías digitalizadas, en donde se permite a los usuarios seleccionar características dentro de la imagen y asociarlas a etiquetas con significado semántico con el fin de describirlas; resultando en una clasificación generada por los usuarios que se conoce comúnmente como folksonomía. El etiquetado colaborativo permite que distintos usuarios describan las imágenes sobre las cuales tienen conocimiento, generando con cada contribución un grado mayor de inteligencia para toda la comunidad. Una vez que el usuario selecciona las características de la imagen, éstas siguen al módulo llamado Asignación de vector de características, en donde se asocian las características obtenidas a la imagen correspondiente para después dentro del módulo de Actualización de características de la imagen, almacenarlas o actualizarlas, según sea el caso, en la base de datos correspondiente.

Recuperación de imágenes basado en contenido. El objetivo de este flujo de trabajo es recuperar un conjunto de imágenes basándose en el contenido de éstas, es decir, en sus características. El inicio de este flujo se define cuando el usuario dentro del módulo de Solicitud de búsqueda en la capa del cliente, solicita una nueva búsqueda, después dentro del módulo de Criterio de búsqueda, el usuario proporciona un criterio, que para una nueva búsqueda, es una palabra clave (mientras la búsqueda se refine éste criterio crecerá en más ítems), la cual envía al módulo de Análisis de características donde se reparten a dos módulos más el módulo de Análisis semántico y el de Algoritmo de similitud. En el Análisis semántico se calcula la semántica de la palabra clave, es decir si la palabra tiene algún problema de polisemia y/o sinonimia, éstos se resuelven y se sugieren los posibles contextos en que el usuario la utilice para delimitar el sentido semántico. Dicha sugerencia se hace al usuario dentro de la aplicación por medio de una nube de etiquetas, para guiar el filtrado de la búsqueda dependiendo el contexto. Mientras que, en el módulo de Algoritmo de similitud, se lleva a cabo la comparación por distancia entre la palabra clave y las etiquetas de características almacenadas en la base de datos con el fin de obtener y mostrar en el módulo de Resultado de consulta aquellas imágenes con contenido similar al criterio de búsqueda propuesto por el usuario.

5. Caso de estudio

Dentro del campo de la medicina existe una gran cantidad de información visual como radiografías, mamografías, resonancias magnéticas, tomografías y muchas más disponibles para análisis médicos. La producción de imágenes crece más rápido

que los métodos para administrar y procesar esa dicha información, imponiéndose un nuevo reto para su eficiente representación, almacenamiento y recuperación.

Para exponer el funcionamiento de la propuesta, se lleva a cabo un caso de estudio dentro del contexto de la medicina, para la recuperación de mamografías digitalizadas con formato ICS, como lo muestra la Figura 3. **Mamografía original digitalizada**

Supóngase que un usuario ha percibido una calcificación en dicha mamografía pero para asegurar su diagnóstico, quiere compararlo con otras mamografías con el mismo diagnóstico. En un caso normal, el usuario busca expedientes de algunas pacientes diagnosticadas de la misma forma, y tomar sus mamografías, con el inconveniente que los expedientes no estén disponibles o que no se trate de una calcificación similar.

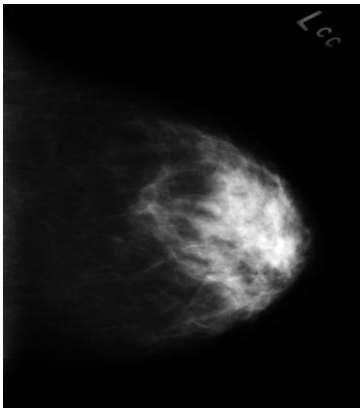


Figura 3. Mamografía original digitalizada

Una alternativa de solución al problema anterior es utilizar la aplicación propuesta en este artículo. Para este caso de estudio se discutirá el modo en que se realiza la asignación de característica y la búsqueda de imágenes similares.

Asignación de características. El usuario carga la imagen en la aplicación, la cual se envía al servidor, en donde por medio de un proceso de binarizacion de la imagen y del análisis de los pixeles, se realiza una sugerencia de los sectores en donde posiblemente se encuentran calcificaciones y/o tumores, tal y como lo muestra la Figura 4. **Proceso de binarización y obtención de áreas de interés.** Permitiendo de esta forma, que el experto corrobore la información modificando la selección y/o reemplazándola como crea conveniente; y agregando la etiqueta correspondiente que describa la característica seleccionada.

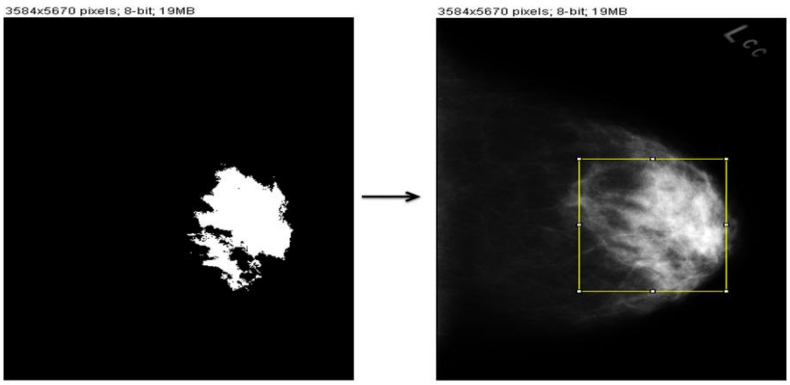


Figura 4. Proceso de binarización y obtención de áreas de interés.

Más adelante otros usuarios si así lo desean, modifican, agregan o eliminan las diferentes características que presenten las imágenes, de acuerdo con su conocimiento para beneficio de la comunidad. La Figura 5. **Etiquetado de características** muestra el proceso de etiquetado, en donde existen opciones de exponer diferentes aspectos de la imagen, en este ejemplo se muestra su contorno. Así las consultas posteriores serán cada vez más exactas y completas debido a la suma de conocimiento de los usuarios participantes en el etiquetado.

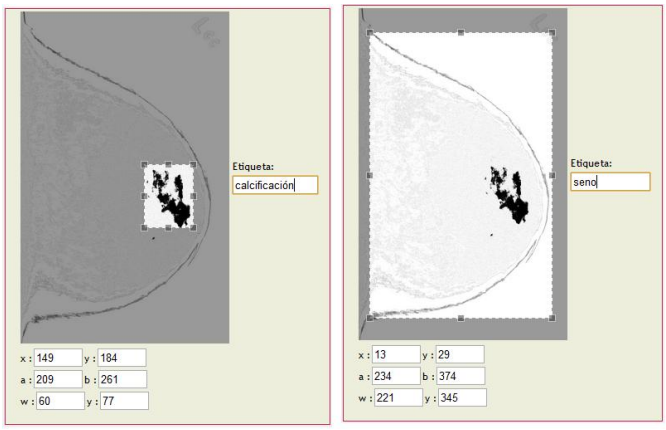


Figura 5. Etiquetado de características

Búsqueda de imágenes similares. Para este caso si el usuario necesita consultar todas aquellas imágenes que contengan *mastitis* como característica, el usuario pro-

porciona la palabra clave en la aplicación, para que, a partir del cálcuo de la distancia euclidiana se obtenga la distancia entre la palabra clave y las etiquetas almacenadas en la base de datos con el fin de obtener aquellas imágenes con contenido milar a la palabra clave propuesta por el usuario. Además la aplicación sugiere, a partir de la palabra clave, los posibles contextos en que el usuario la utilice para delimitar el sentido semántico.

Este caso de estudio muestra la utilidad de la aplicación propuesta como una alternativa para la búsqueda de imágenes médicas basadas en contenido, utilizando Inteligencia Colectiva dentro de la aplicación Web 2.0. Es útil para usuarios que deseen recuperar imágenes con significado semántico y visual similar, basándose en el conocimiento y experiencia de otros usuarios.

6. Trabajo futuro

Como trabajo futuro está el desarrollo de la interfaz para el flujo de trabajo Recuperación de imágenes basado en contenido y sus correspondientes módulos para agregarlos a la aplicación Web, de tal forma que se solucione los posibles problemas de semántica para las palabras clave y además se aplique el algoritmo de distancia entre la palabra clave y las características almacenadas en la base de datos, para obtener las imágenes con contenido similar. Así también, la implementación de un sistema de reputación para los usuarios que participan en el etiquetado de las imágenes, de tal forma que sea posible conocer que tan confiables son sus opiniones para etiquetar las imágenes. Dicha reputación se adquiere de las calificaciones de una encuesta realizada a otros usuarios, quienes realizan las búsquedas y expresan su opinión acerca de la certeza de dichas etiquetas.

7. Conclusiones

El factor determinante en el interés por las técnicas CBIR es el rápido incremento del tamaño de colecciones de imágenes digitales, así como en la importancia y disponibilidad de las imágenes en diversos campos aplicación como la medicina. Por lo que existe una necesidad de gestionar datos de imagen digital, dado que la digitalización en sí no facilita la gestión, aunque posibilita derivar automáticamente información de las imágenes en sí mismas. Esta propuesta intenta resolver problemas semánticos y de contenido en la realización de búsquedas, específicamente de mamografías digitalizadas, mediante el uso de Inteligencia Colectiva y al implementar un algoritmo CBIR, permitiendo que los usuarios sean las personas encargadas de administrar y conformar la información existente dentro de la aplicación, obteniendo así un mayor grado de conocimiento para la comunidad. Los beneficios que presenta la propuesta dentro del contexto de la medicina en donde los diagnósticos se hacen frecuentemente diferenciales, son gran ayuda debido al uso de la inteligencia colectiva de los especialistas.

Agradecimientos

Este trabajo fue apoyado por la Dirección General de Educación Superior Tecnológica de México (DGEST). Además, este trabajo fue patrocinado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) y la Secretaría de Educación Pública a través de PROMEP.

Referencias

1. Polisemia. Enciclopedia Microsoft Encarta Online 2009 http://mx.encarta.msn.com. 1997-2009. Microsoft Corporation. Reservados todos los derechos.[Consulta: sábado, 31 de noviembre de 2009]
2. Revista Española de Lingüística. "Sinonimia y teoría semántica en diccionarios de sinónimos de los siglos XVII y XIX". http://www.uned.es/sel/pdf/ene-jun-94/24-1-Gonzalez.pdf. 1994. [Consulta: miércoles, 23 de diciembre de 2009]
3. Ulrike, B., Sandro, G., Henrik, I., Reinhard, J.: Design of new business models for service integrators by creating information-driven value webs based on customers collective intelligence. In: Proceedings of the 42nd Hawaii International Conference on System Sciences. 2009.
4. Ishida T.: Service-Oriented Collective Intelligence for Intercultural Collaboration. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2008.
5. Ching-man, Yeung A., Gibbins, N., Shadbolt, N.: A k-Nearest-Neighbour Method for Classifying Web Search Results with Data in Folksonomies. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2008.
6. Jian, C., Ming, Y.: A New Similarity Measurement Based on Distance and Correlation Test for Content-Based Images Retrieval. In: Congress on Image and Signal Processing. IEEE. 2008.
7. Barrios, J.M., Díaz-Espinoza, D., and Bustos, B.: Text-Based and Content-Based Image Retrieval on Flickr: DEMO. In: Second International Workshop on Similarity Search and Applications. IEEE. 2009.
8. Da S-Torres, R., Bauzer-Medeiros, C., Goncalves M. A., Fox, E.: An OAI Compliant Content-Based Image Search Component. In: Join ACM/IEEE Conference on Digital Libraries. ACM. 2004.
9. Ribes Xavier. "La Web 2.0. El valor de los metadatos y de la inteligencia colectiva". http://www.telos.es/articuloperspectiva.asp?idarticulo=2&rev=73. 2008. [Consulta: jueves, 03 de septiembre de 2009]
10. Cobo Romaní Cristobal, Pardo Kuklinski Hugo. "Planeta Web 2.0. Inteligencia colectiva o medios fast food". Grup de Recerca d'Interaccions Digitals, Universitat de Vic. Flacso México. Barcelona / México DF. 2007.
11. Dean, P., Mascio, L., Ow, D., Sudar, D., Mullikin, J.: Propsed standard for image cytometry data files, Cytometry. In: Wiley-Liss, Inc. n.11, pp. 561-569, 1990.

Arquitectura para un Sistema de Información Web de medición de datos en línea.

Jesus Leonardo López Hernández, Ignacio Rodrigo Cervantes Mendieta, Ana María Chávez Trejo, Giner Alor Hernández

Instituto Tecnologíco de Orizaba
Division de Estudios de Posgrado e Investigación
jleonardolh@gmail.com, igrodcer@gmail.com,
achavezt@prodigy.net.mx, gineralor@hotmail.com
Paper received on 02/07/10, Accepted on 05/10/10.

Abstract. El Departamento de Operación de PEMEX Refinación Sector Mendoza, realiza actividades de transporte por ducto y requiere un proceso de análisis de las condiciones operativas, en base a reportes operativos que contienen la información de cada Estación de Bombeo y Terminal de Almacenamiento y Reparto. Este trabajo presenta una propuesta para la arquitectura de un Sistema de Información Web de medición de datos en línea que permita la comunicación en tiempo real con los dispositivos físicos (Sensores) y la computadora. Finalmente se describe el dispositivo electrónico utilizado en el proceso de adquisición y conversión de datos analógico-digital.

Palabras Clave: Arquitectura, Tarjeta de Adquisición Datos, Aplicación Web.

1. Introducción

La actualidad informática de las empresas requiere que sus sistemas informáticosdíganse aplicaciones Web o aplicaciones locales, cubran las necesidades operativas ya que dependen de los resultados obtenidos de estos sistemas. En PEMEX Refinación Sector Mendoza,el Departamento de Operación se encarga de la lógica del transporte por ducto y para llevar a cabo esta actividad realiza una monitorización constante que permite a los encargados conocer los datos de campo generados. Una problemática actual, es la generación de reportes operativos diarios, que requieren de personal para realizar esta tarea repetidas veces, es aquí donde eltrabajo se vuelve complicado, porque requiere de comunicación diaria y constante vía Trunkingo Radio a cada EB (Estación de Bombeo) y TAR (Terminal de Almacenamiento y Reparto), para conocer las condiciones operativas diarias y novedades, esto causa errores de inconsistencia y problemas de tiempo de entrega y sincronización de captura en otras aplicacionesinformáticas. La interpretación de las variaciones de presión entre las EB, es otra problemática encontrada ya que en ocasiones se necesita verificar visualmente si existe una toma clandestinao los equipos de la empresa presentan fallas mecánicas que generandatos erróneos. El requerimiento de un Sistema de In-

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp.71.-80

formación que cubra la necesidad de la presentación de reportes finales de condiciones operativas y la detección y seguimiento de variaciones de presiones de bombeo, será de gran impacto para el Departamento de Operación al facilitar la monitorización de las condiciones operativas durante el proceso de bombeo de producto (crudo o procesado) tanto para el área de Poliducto y Oleoducto. El sistema detectará presiones fuera de los rangos permitidos para la operación y generará avisosque permitan tomar medidas pertinentes, esto se traducirá en la reducción de la pérdida de producto por robo. Además reducirá el tiempo de captura e inconsistencias producidas al generar los reportes, ya que éstos actualmente se elaboran en forma manual y requieren la comunicación constante entre operadores de diferentes estaciones para llenar todos los datos necesarios. Todo esto aunado a que el sistema tendrá la capacidad de realizar cálculos específicos para la presentación visual de estos informes finales.

Un "Sistema de Información de medición en línea" se define como un Sistema Informático que interactúa con medios físicos (dispositivos electrónicos) por medio de los cuales recibe datos para procesarlos y presentar resultados [1].

Cada Sistema de Información debe diseñarse en base a una arquitectura con el objetivo a cumplir para los fines planteados. La siguiente sección presenta la Arquitectura Propuesta para el Sistema, en base al estilo arquitectónico en capas y el patrón arquitectural MVC (Modelo-Vista-Controlador)

2. Arquitectura Propuesta para el Sistema de Información Web

La arquitectura que se propone para el sistema Web, se basa en la arquitectura general de aplicaciones ASP.Net MVC, el cualse instala en el mismo servidor que contiene al Sistema Gestor de Base de Datos (Oracle), para así mediante un mapeo Objeto-Relacional (O-R)realizar la conectividad entre estas tecnologías.Cualquier usuario que tenga acceso y haga una petición de la aplicación Web no se le tendrá que instalar el software para hacer uso de la aplicación.Los componentes principales para las aplicaciones que se desarrollan en el marco de trabajo .Net MVC, tienen la característica de distinguir a éste, no permiten que se mezclen los elementos propios de cada componente o capa, es decir, si un elemento vista requiere a un elemento de la base de datos (Modelo), esta petición se tiene que hacer vía controlador, ya que es el encargado de cómo su nombre lo indica, controlar las peticiones entre la aplicación.

La Figura 1, muestra la arquitectura en capas propuesta y dentro de cada capa los elementos software o elementos electrónicos para la aplicación.

- **Capa Vista.** En esta capa se tienen a los componentes que desplegaránla presentación o interfaz de usuario y los datos solicitados en base al tipo de usuario.
- **Capa Controlador.** Esta es la capa más importante del marco de trabajo, ya que funciona como un intermediario entre las diferentes capas de la aplicación, cualquier petición de consulta que se realice por parte del usuario o de almacenamientoa la base de datos tendrá que ser gestionada por el controlador, ésterealiza las llamadas a los métodos o instancias necesarias para hacer dichas tareas.

- **Capa Modelo.** En esta capa se tienen a los componentesnecesarios para construir el mapeo Objeto-Relacional (O-R), para manejar como objetos a las entidades de la base de datos. La representación O-R es como el marco de trabajo sugiere se trabaje hacia las bases de datos.
- **Capa ADC.** La capa Convertidor Analógico-Digital, se encarga de adquirir los datos de los dispositivos físicos, procesarlos y almacenar mediante un módulo software (LAB-View) a la base de datos. El componente USB establece la comunicación entre la tarjeta de adquisición de datos y la computadora. El micro-controlador PIC18F2550 es el dispositivo electrónico que realiza las tareas de adquisición y comunicación.

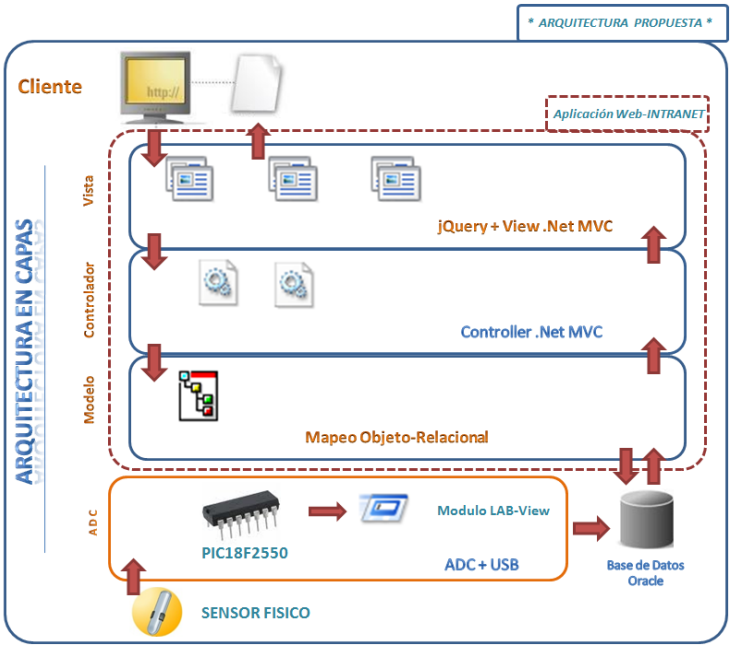


Fig. 1. Arquitectura Propuesta.

La tecnología de esta aplicación permite integrar a este proyecto a jQuery, un *framework* quepermite construir mejoresinterfaces de usuario.

A continuación la sección 3 describe la tecnología usada para la construcción de la Tarjeta de Adquisición de Datos de la capa ADC, en base al micro-controladorPIC18F2550, detallándose las características más importantes.

3. Tecnología para la construcción de la Tarjeta de Adquisición de Datos

El núcleo de la tarjeta de adquisición de datos es un micro-controlador PIC18F2550 de Microchip. La función de este dispositivo es convertir los datos analógicos a digitales, y establecer una comunicación mediante el protocolo USB con una computadora personal. El micro-controlador mencionado dispone de un módulo para realizar la comunicación serial compatible con el SIE (serial interface engine o máquina con comunicación serie) USB "full-speed" (2.0) y "low-speed" (1.0).El micro-controlador PIC18F2550 puede alimentarse con un voltaje de 2.5 hasta 5.5 Volts, por lo que la tarjeta de adquisición de datos será alimentada por medio del puerto USB o desde una fuente de alimentación externa. En la tabla siguiente se muestran las características del micro-controlador PIC18F2550.

Tabla. 1. Descripción de características del PIC 18F2550.

Tipo de memoria programable	Flash
Memoria de programa (bytes)	32.768
Memoria RAM de datos (bytes)	2.048
Frecuencia de operación	Hasta 48 MHz
Líneas de E/S	24
Canales de Conversión A/D de 10 bits	10 canales
Puertos USB	1, Full Speed, USB 2.0
Temperatura de operación (c)	-40 hasta 85
Rango de voltaje de operación	2 hasta 5.5
Numero de pines	28

A la tarjeta de adquisición de datos se agrega un cristal de cuarzo de 20 MHz encargado de llevar el ritmo del PIC 18F2550, éstetrabaja cuánticamente pasando de un estado a otro en un lapso de tiempo, determinado por el control del Cristal, que interpreta cada cuánto va a dar su siguiente salto, en nuestro caso dado que el cristal es de 20 MHZ cambiará de estado una vez cada 0,00000005 segundos (0.05µs), que es lo que se conoce como el Ciclo de Reloj.En la figura 2, se describen los elementos de la tarjeta de adquisición de datos para la aplicación. A continuación se mencionaran los componentes más importantes.

1. Componente USB es un conector tipo B, el cual tiene la funcionalidad de proveer la comunicación con la computadora, para así transmitir los datos que se están adquiriendo desde los medios físicos.
2. Elemento indicador del estado de la tarjeta de adquisición de datos es un "LED" que indica el estado de la tarjeta,activa o inactiva.
3. Botón de reinicio del PIC, cargará nuevamente las funciones de la tarjeta en el caso de que se encuentre bloqueada o los datos obtenidos sean inconsistentes.

4. Para el proceso de lectura de los datos,la tarjeta contiene un grupo de borneras, cuya función es obtener el dato analógico del medio físico.

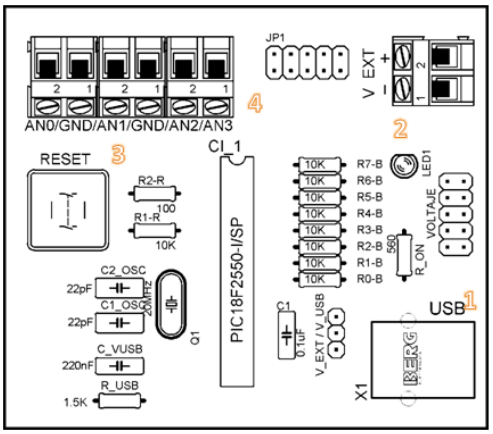


Fig.2. Diagrama de la Tarjeta de Adquisición de Datos.

4. Resultados

A continuación se presentan los objetivos alcanzados hasta este momentopara el proyecto. Cabe mencionar que todos los elementos contemplados para la arquitectura, se ajustana las tecnologías con las que la empresa ya tiene sistemas implementados, por lo que se siguióla misma línea de trabajo, el elemento hardware como la Tarjeta de Adquisición de Datos se construyó en base a las necesidades de la empresa. Con el análisis de los datos operativos y de campo necesarios para la presentación de informes finales y el proceso de adquisición de datos, serealizó el diseño y codificación de la base de datos bajo la tecnología Oracle. Uno de los objetivos primordiales de este proyecto es la eliminación de la inconsistencia y redundancia de datos, ya que en la actualidad el Departamento de Operación no hace uso de una base de datos. A continuación en la Figura 3, se presenta el diseñode la base de datos, para el Sistema de Información Web.

Para el proceso de conversión de datos analógico-digitalesel micro-controlador PIC18F2550 cubre en gran medida los requerimientos técnicos para el desarrollo de esta aplicación, la integración de este dispositivo con distintos lenguajes de programación justifican su uso, en este caso el componente para conectividad por USB a la computadora viene listo para su implementación.

Además este micro-controlador proporciona confiabilidad en la lectura de datos según el ámbito de aplicación y problema a resolver, la alimentación con la que trabaja puede ser suministrada por el componente USB que se usa para el envío de los datos, con esto no se quiere decir que sea la única opción a utilizar para este proyecto, pero en base a las necesidades de la empresa se optó por este dispositivo electrónico.

La vista física de la arquitectura del Sistema de Información Web, se presenta en la Figura 4, esta vista muestra características de distribución de los componentes

software con el hardware, así como el medio de comunicación de la Intranet en la cual va a interactuar la aplicación. Esta aplicación será utilizada desde máquinas de EB y TAR. El ambiente Intranet, en el cual el Sistema trabajara tiene el objetivo de brindar seguridad a los datos ya que sólo tendrán acceso los usuarios autorizados.

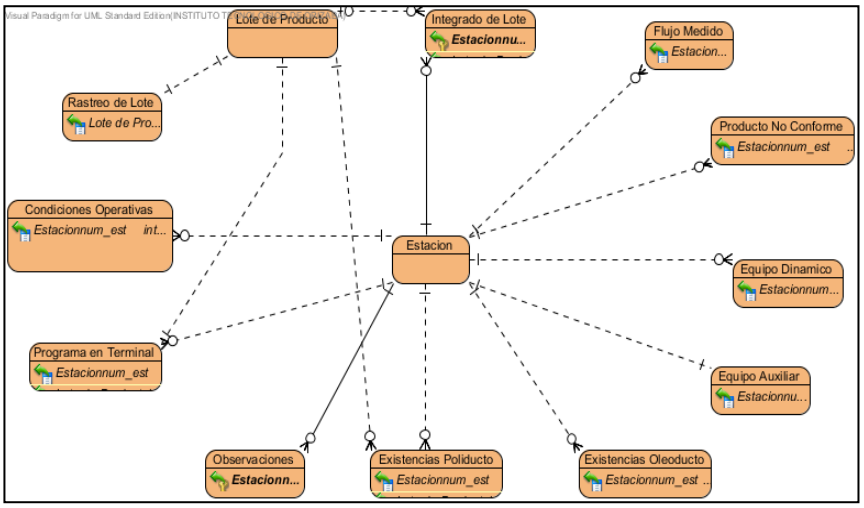


Fig. 3 Esquema lógico de la Base de Datos.

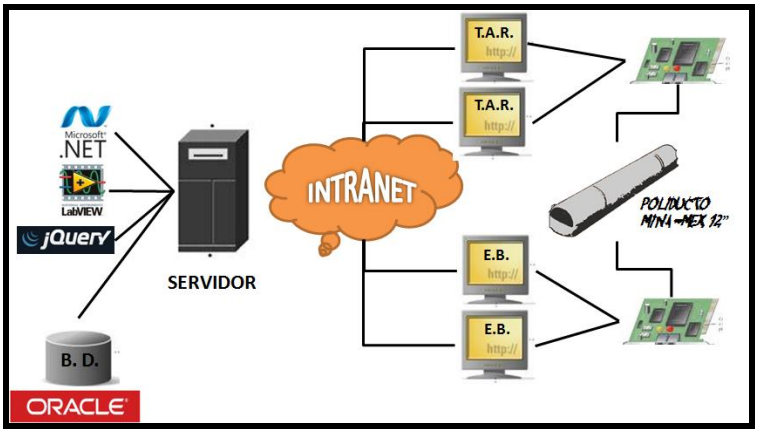


Fig. 4. Arquitectura física del Sistema

Se presenta a continuación el diseño de interfaces de usuario del sistema mediante algunos prototipos para el proceso de captura de datos y para el proceso de consulta. El proceso de captura se realiza de manera independiente por cada EB y

TAR, a excepción de los datos (Presiones) que se adquieran mediante el proceso de conversión analógico digital, los cuales no podrán ser manipulados por los usuarios.

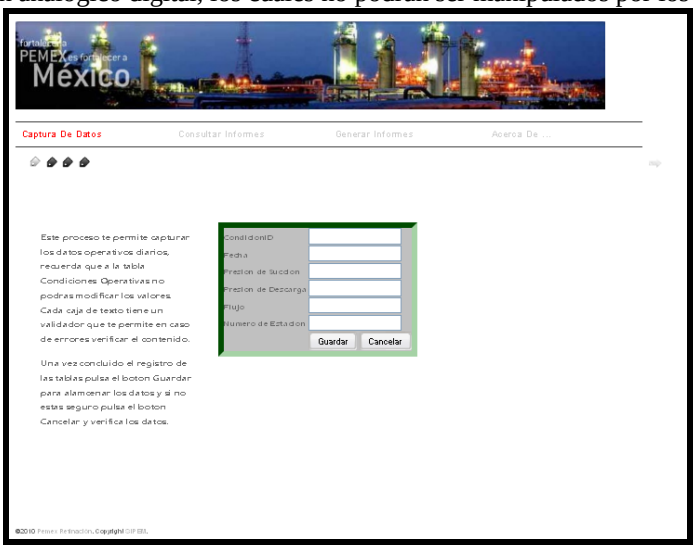


Fig. 5 Prototipo de Interfaz de usuario para Captura de Datos.

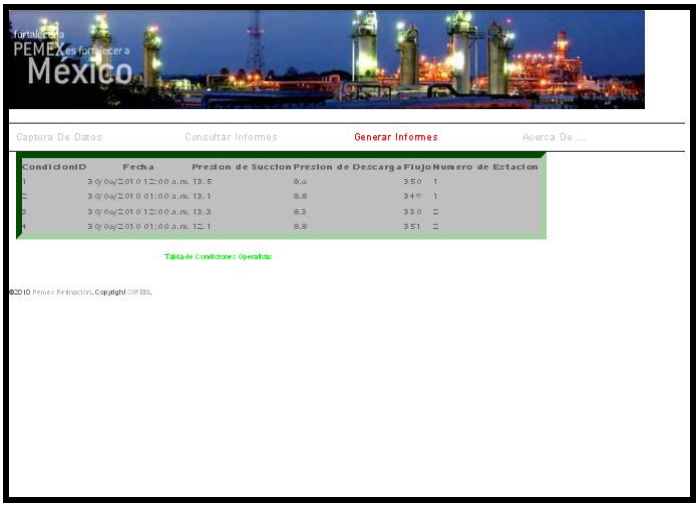


Fig. 6 Prototipo de interfaz de usuario para consultas.

5. Trabajos Relacionados

En [9] se presenta un sistema de información, para seguridad industrial y protección ambiental, que minimiza el proceso de elaboración manual de elaboración de reportes operativos e informes, para el área de Mantenimiento, el cual calcula au-

tomáticamente las condiciones de desempeño, vinculando la eficacia de las acciones a las metas establecidas.

El SISPA (Sistema de información industrial protección al medio ambiente), fue desarrollado para un ambiente Web, con una tecnología ASP, que permite su uso en las estaciones que comprende el Sector Mendoza, cabe mencionar que el gestor de base de datos que ocupa es Oracle.

En [10] se describe un sistema de información gerencial llamado SIGOR (Sistema de Información Gerencial de Operación de Refinación), la tarea de los sistemas de información gerenciales es el apoyo al procesar la información para el apoyo a la toma de decisiones importantes que influyan en el funcionamiento de la empresa.

En [11] se presenta una técnica para el desarrollo de una aplicación Web para un sistema de información en tiempo real, que permita la conexión remota y concurrente de diferentes equipos en la red a la base de datos histórica del sistema, sin necesidad de que se instale ningún componente de software en el equipo remoto del usuario que realiza la consulta.

En [12] se describe SITRAC (Sistema de Transferencia en Custodia para PEMEX Refinación) un sistema de información de aplicación general para Pemex Refinación, cuyo objetivo es obtener un balance de los hidrocarburos transportados, por cada TAR (Terminal de Almacenamiento y Reparto) requiere capturar la información con referencia los productos Px. Magna, Px. Premium y Px. Diesel.

En [13] se muestra como un sistema basado en sensores inalámbricos que tiene por objetivo detectar, localizar y cuantificar las fugas y otras anomalías comunes durante el proceso de transporte de agua por tuberías. También se utiliza para monitorizar la calidad del agua y para medir los niveles en los colectores del alcantarillado. Se menciona que el sistema recoge vibraciones hidráulicas y acústica y mencionan los algoritmos que ocuparon para el análisis de estos datos para detectar y localizar fugas.

XStream [14] es un sistema de gestión de flujos de datos que soporta eficientemente aplicaciones de procesamiento de señales de rango amplio. Las aplicaciones de motivación para este sistema provienen de una variedad de dominios como los son industriales, científicos, y de ingeniería. Estas aplicaciones usan sensores integrados de vibración, sísmicos, de presión, magnéticos, acústicos, de red y sensores médicos que obtienen muestra de datos que están en el orden de millones de muestras por segundo. Estos flujos de muestra de datos de alta frecuencia se procesan y analizan en cuestión de milisegundos.

El sistema combina la secuencia de eventos y el procesamiento de la señal. XStream tiene como objetivo mejorar tanto la productividad del programador, facilitando el desarrollo de funciones de procesamiento definidas por el usuario; así como un alto desempeño, procesando varios millones de muestras por segundo en una computadora estándar.

6. Trabajo a Futuro

De los trabajos a futuro pendientes para concluir este proyecto, se encuentra la integración de la tarjeta de adquisición de datos con elmódulo de software (LabVIEW), que permitirá almacenar en la base de datos del sistema, la información ad-

quirida de los medios físicos,ésta es una de las tareas de mayor prioridad. Una vez realizada esta tarea de almacenar los datos desde el sensor se procederá a realizar las vistas de informes finales de condiciones operativas.

7. Conclusiones

Con la tecnología .Net MVC y la arquitectura propuesta se permite el uso de la aplicación Web a múltiples usuarios en una manera concurrente, se realizarán consultas a la base de datos del sistema, sin tener que instalar componentes de software en los equipos y sin generar ningún conflicto entre ellos.

Debido a que todos los procesos son ejecutados en el Servidor utilizando la tecnología de .Net el "Common Language RunTime", permite a los usuarios invocar a funciones por medio de los objetos de la interfaz gráfica de la aplicación Web, que de igual manera se encarga de visualizar los resultados.

Agradecimientos.

Se agradece al Consejo Nacional de Ciencia y Tecnología (CONACyT) por las becas otorgadas para la realización de los estudios de posgrado.

Se agradece la colaboración del Dr. Gerardo Águila Rodríguez y el Ing. Juan Pablo Gómez Nieto, del Laboratorio de Interfaces del Instituto Tecnológico de Orizaba, por su apoyo para la creación de la tarjeta de adquisición de datos, para la aplicación Web.

Referencias

1. A. Burns y A. Wellings. Real-time Systems and their Programming Languages. Addison Wesley. 1996.
2. Definiciones de Arquitectura de Software, http://www.sei.cmu.edu/architecture/start/community.cfm
3. Carlos Billy Reynoso, Introducción a la Arquitectura de Software, Marzo de 2004.
4. Visual Studio .NET Introducción a la plataforma Microsoft Visual Studio .NET.
5. Morales-Chaparro, R.; Linaje, M.; Preciado, J. C.; Sánchez-Figueroa F MVC Web design patterns and Rich Internet Applications, http://www.sistedes.es/sistedes/pdf/2007/SCHA-07-Morales-MVC.pdf
6. Data Sheet PIC 18F2550,http://ww1.microchip.com/downloads/en/DeviceDoc/39632b.pdf.
7. PIC18F2550, http://www.microchip.com/wwwproducts/Devices.aspx?dDocName=en010280.
8. Alfredo Espinosa R., Brisa M. Silva F. y Agustín Quintero R, Desarrollo de una aplicación Web para un sistema de información en tiempo real.
9. PEMEX Refinación "Sistema Industrial de Seguridad y Protección al Medio Ambiente" http://sispanet.dco.pemex.com/sispa2/iniciodelsispa/Default.html
10. PEMEX Refinación "Sistema de Información Gerencial de Operación de Refinación" SIGOR, 2004
11. Alfredo Espinosa R., Brisa M. Silva F. y Agustín Quintero R. "Desarrollo de una aplicación web para un sistema de información en tiempo real". Boletín IIE, abril-junio del 2007 pp. 52-58.

12. CIATEQ “Sistema de transferencia de custodia” http://www.mexicocyt.org.mx/investigaciones/2974
13. I. Stoianov, L. Nachman, S. Madden, and T. Tokmouline, “PIPENET: A wireless sensor network for pipeline monitoring,” in IPSN ’07: Proceedings of the sixth international conference on Information processing in sensor networks. New York, NY, USA: ACM Press, 2007
14. L. Girod, Y. Mei, R. Newton, S. Rost, A. Thiagarajan, H. Balakrishnan, S. Madden, “XStream: a Signal-Oriented Data Stream Management System” In Proceedings of the International Conference on Data Engineering (ICDE'08), Cancun, Mexico, April 2008.

Monitoreo de procesos de negocios a través de un enfoque orientado a metas

Alicia Martínez, Nimrod González, Hugo Estrada

CENIDET, Cuernavaca, Morelos, México
{amartinez, nimrod, hestrada}@cenidet.edu.mx
Paper received on 13/08/10, Accepted on 04/10/10.

Resumen. Hoy en día, la utilización de sistemas de software para dirigir y monitorear de manera automática los procesos de negocios se ha convertido en una opción muy común entre las empresas modernas. Este tipo de sistemas (*workflow*) son diseñados para monitorear las tareas, documentos, reglas, y otros recursos que son utilizados en un proceso de negocio, el cual es modelado y ejecutado en un entorno automatizado. Es importante hacer notar que en su mayoría, los sistemas *workflow* están orientados a capturar las tareas necesarias para ejecutar un proceso de negocio desde un punto de vista basado en funcionalidades. Sin embargo, estos sistemas han descuidado la representación de las metas y objetivos que la empresa desea alcanzar con el desempeño de sus tareas. De esta manera, en los enfoques actuales de *workflow*, resulta muy complicado determinar si las tareas modeladas en realidad satisfacen los objetivos de negocio. En este artículo, el framework Tropos es utilizado para modelar las metas de los procesos de negocio y además, para establecer el mecanismo para controlar y monitorear las metas de cada uno de los actores involucrados en los procesos del negocio. Con esto, es posible medir el impacto del cumplimiento de las metas del proceso en la satisfacción de las metas del negocio. Una herramienta de software llamada *GoalFlow* fue desarrollada con el objetivo de validar nuestra metodología propuesta. Con este enfoque es posible administrar y tomar las decisiones apropiadas acorde al nivel de cumplimiento de las metas del negocio.

Palabras clave: Metas del negocio, sistemas *workflow*, tareas de monitoreo

1. Introducción

Hoy en día es habitual que las organizaciones recurran a sistemas de software como auxiliares en el desarrollo y administración de sus procesos. Recientemente se ha popularizado el uso de herramientas de tipo *workflow*, las cuales ayudan en el control y monitoreo de los procesos de una organización. Estos sistemas son muy útiles para el desempeño de las funciones de las empresas, ya que permiten guiar y controlar de forma automática los componentes de un proceso de negocio (personas, tareas, documentos, normas y computadoras) [1].

Algunos de los enfoques más conocidos para modelado de *workflow* son BPEL, XPDL, YAWL, openEDMS, OnBase, LiquiOffice, Cardiff y Bonita [2][3][4][5]. Todas estas herramientas ofrecen una base sólida para modelar procesos y son cada día más populares. Sin embargo, a pesar de las ventajas conocidas de los sistemas de *workflow*, existen algunas problemáticas, las cuales necesitan ser resueltas para permitir a estos sistemas realizar análisis más profundos de aspectos claves de la estructura del negocio y de los comportamientos

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp.81-91.

organizacionales, tales como metas, dependencias, tareas, etc. Casi la totalidad de los sistemas de *workflow* tienen limitaciones de modelado, debido principalmente a que estos están basados en representar tareas de negocio y han descuidado el uso de metas para tener una vista más completa de los procesos de negocio de una empresa. Por lo tanto, en algunos casos, estos enfoques basados en tareas no son suficientes para modelar un contexto específico.

En este artículo se propone un enfoque basado en metas para monitorear y controlar procesos definidos en un sistema *workflow*. Para hacer esto, el framework Tropos [6] es usado para determinar el impacto de la satisfacción de las metas de los actores en la satisfacción de las metas del negocio. Es importante mencionar que el framework i* puede ser utilizado para los mismos propósitos. En este trabajo, una herramienta de software, denominada *GoalFlow* ha sido desarrollada para validar la metodología propuesta.

El artículo está estructurado de la siguiente forma: los fundamentos teóricos son descritos en la siguiente sección. El modelado de procesos de negocio es presentado en la sección 3. La sección 4 muestra el desarrollo de un caso de estudio, y finalmente, las conclusiones se presentan en la sección 5.

2. Fundamentos teóricos

A continuación se presentan los dos conceptos básicos utilizados en este trabajo de investigación: el framework Tropos y los sistemas *workflow*.

2.1. El framework Tropos

El framework Tropos está formado por dos diagramas que nos permiten modelar el ambiente organizacional con un enfoque basado en metas y dependencias [7], [8]:

Diagrama de actores. Es una representación gráfica donde se muestran los actores y sus metas, y las dependencias entre los actores.

Diagrama de metas. Es una representación gráfica donde se analizan con mayor detalle las metas, planes y dependencias de cada actor.

A continuación se presentan los conceptos básicos utilizados en este framework.

Actor. Es una entidad que tiene metas estratégicas e intenciones dentro del sistema o dentro del conjunto organizacional. La representación gráfica de un actor es un círculo de líneas punteadas.

Hardgoal/Softgoal. Estas metas representan los intereses estratégicos de un actor. Las *hardgoals* se distinguen de las *softgoals* porque en las segundas no existe un criterio claro para definir si ellas han sido satisfechas. Las *softgoals* son dibujadas como una nube, mientras que las *hardgoals* se muestran como un rectángulo con las puntas redondeadas.

Plan. Representa una forma de hacer algo en un nivel abstracto. La ejecución del plan puede ser una manera de satisfacer una *hardgoal* o una *softgoal*. Los planes son dibujados como hexágonos.

Recurso. Representa una entidad física o de información. Son dibujados como un rectángulo.

La fig. 1 ilustra la notación gráfica de estos conceptos de modelado.

Fig. 1. Notación gráfica de los elementos básicos de Tropos

Dependencias. Es una relación intencional y estratégica entre dos actores. Este tipo de relación indica que un actor depende de otro actor con el objeto de alcanzar una meta, ejecutar un plan u obtener un recurso. El primer actor es llamado *depender*, mientras que el actor del cual se depende se denomina *dependee*. El objeto alrededor del cual se centra la dependencia se denomina *dependum*. Existen 4 tipos de dependencias: dependencia *hardgoal*, dependencia *softgoal*, dependencia de recurso y dependencia de plan. La representación gráfica de las dependencias se ilustra en la fig. 2.

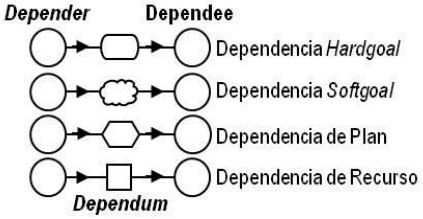


Fig. 2. Notación gráfica de las relaciones de dependencia

2.2. Herramientas de *workflow*

Las aplicaciones de *workflow* [9], permiten automatizar las actividades involucradas en procesos de negocios con el objetivo de mejorar la forma en las que estas se llevan a cabo. Dichos procesos comprenden las secuencias de acciones, actividades o tareas utilizadas para la ejecución de los mismos, incluyendo el seguimiento del estado de cada una de sus etapas y la aportación de las herramientas necesarias para gestionarlos. En este contexto, una definición de *workflow* señala que *"la automatización de proceso de negocios, en forma parcial o total, implica el conocimiento de los documentos, información o tareas son pasados de un participante a otro para realizar acciones de acuerdo a un conjunto de reglas procedimentales"* [10]. De esta manera el propósito de un *workflow* es lograr una mayor eficiencia para unir personas, procesos y computadoras con el objeto de reducir el tiempo y acelerar la realización del trabajo.

3. Modelado de procesos de negocio orientado a metas

Las principales contribuciones de este trabajo de investigación son:

a) El uso del framework Tropos para controlar, seguir y monitorear metas de negocio. Es importante mencionar que Tropos ha sido extendido para considerar actividades de monitoreo sobre acciones, tareas y metas.

b) Hacer explícito cómo las tareas de los actores tienen impacto en la satisfacción de las metas del negocio. La idea principal de este enfoque es determinar la satisfacción de las metas de los procesos basada en el cumplimiento de sus tareas y actividades, y finalmente
c) Este enfoque es una herramienta útil para el administrador de la empresa para tener una retroalimentación inmediata acerca del desempeño de los procesos, tareas y actividades, el grado de cumplimiento de las metas, el impacto de las relaciones entre actores en el desempeño de los procesos, etc. Toda esta información es de utilidad para tener una idea clara de la información que responde las preguntas: qué, quién, porqué y cuando se realizan las actividades en el negocio.

Uno de los conceptos relevantes en esta propuesta son los planes, los cuales, a fin de asociarlos con elementos de la tecnología *workflow*, son manejados bajo el nombre de tareas. En esta propuesta, se añadió el concepto de acción para refinar las tareas de Tropos.

El método de modelado de procesos de negocios está compuesto de siete fases. La entrada de este método es la información de los procesos de negocio que requieren ser modelados, y la salida es el modelo de procesos de negocios orientado a metas.

Análisis de la información: esta fase permite obtener, explorar y analizar todas las fuentes posibles que describan a las actividades del proceso y su entorno.

Definición del proceso: esta fase especifica el objetivo del proceso y sus características; por ejemplo, quien es el actor responsable del proyecto, tiempos de ejecución, etc.

Identificación de actores: esta fase permite identificar a los actores que se encargan de desempeñar las actividades del proceso.

Refinamiento de actores: esta fase permite definir las capacidades y habilidades de cada actor.

Identificación de metas: durante esta fase se extraen todas las actividades necesarias para la realización del proceso y se organizan como *hardgoals* y *softgoals* según el actor responsable de su ejecución.

Refinamiento de *hardgoals*: en esta fase se detalla cada *hardgoal* especificando el conjunto de tareas cuya ejecución es necesaria para el cumplimiento de dicha meta y posteriormente se especifican las acciones requeridas para desempeñar cada tarea.

Identificación de relaciones: consiste en extraer las dependencias entre los actores especificando la dirección de la dependencia, el tipo de dependencia y el objeto alrededor del cual gira la dependencia.

La fig. 3 muestra los pasos propuestos en nuestra propuesta de investigación.

4. Métricas para controlar y monitorear los procesos de negocios orientados a metas

En esta propuesta de investigación se utilizan métricas y axiomas para implementar los procesos de control y monitoreo de los procesos de negocio definidos en los pasos previos de la metodología.

Métricas. Las métricas que se han definido permiten cuantificar el progreso en la satisfacción de cada una de las metas del negocio, además de cuantificar el progreso de cada uno de los procesos que fueron modelados. El enfoque GQM (Goal-Question-Metric) de Basili, Caldiera y Rombach [11] fue utilizado para especificar los objetos de medición. Esta es la primera fase del proceso de definición de métricas para procesos de negocio. El enfoque GQM es de gran utilidad para derivar preguntas y caracterizar las metas en forma cuantificable. La tabla 1 muestra el resultado de la aplicación del enfoque GQM para nuestra propuesta.

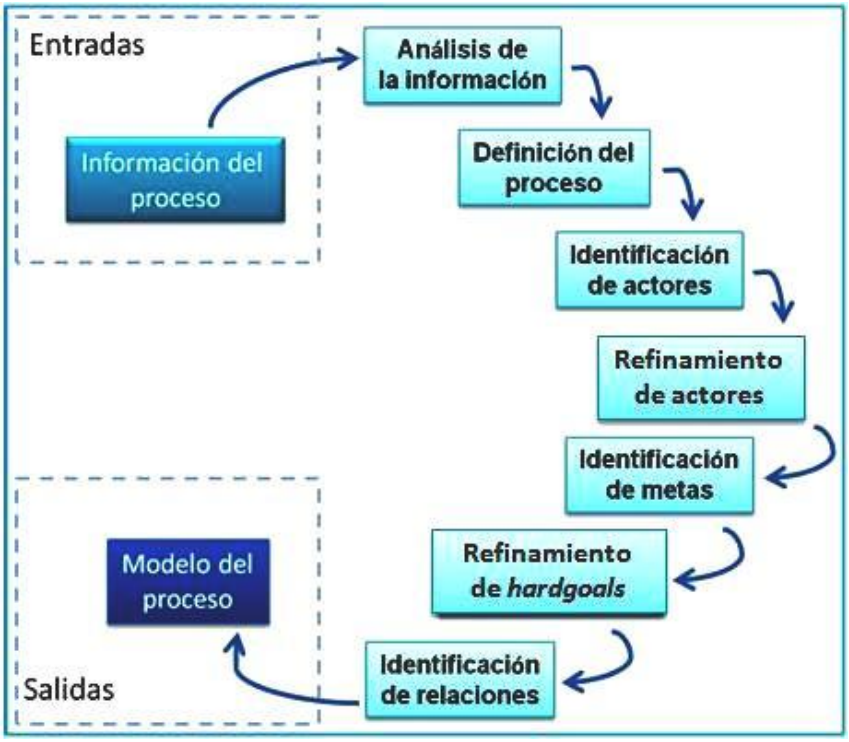


Fig. 3. Etapas del modelado de procesos orientado a metas

Una vez que las preguntas han sido formuladas, el proceso siguiente consiste en asociar las preguntas con las métricas apropiadas. En nuestro enfoque, la efectividad de un proceso es determinada por la satisfacción de todas sus *hardgoals*. Estas metas representan los objetivos que deben ser logrados para proveer un servicio o producir un producto. Por su parte, la eficiencia de un proceso es determinada indirectamente a través de la satisfacción de las *softgoals*. Estos factores pueden ser considerados como la calidad de un proceso.

Los factores utilizados en este trabajo de investigación para implementar el método GQM son la efectividad y la eficiencia. Estos factores pueden ser medidos por las siguientes métricas (tabla 2):

Las métricas han sido definidas en forma específica para este proyecto de investigación, sin embargo, estas pueden ser utilizadas en otros enfoques que usen metas en modelado de negocios.

Tabla 1. Niveles y objetos de medición para el proceso de negocios orientado a metas

Niveles	Objeto de Medición
Conceptual (Metas)	- Medición del progreso en la satisfacción de metas y tareas. - Validar el cumplimiento de las metas y tareas. - Medir el progreso y la ejecución de los procesos.
Operacional (Pregunta)	- ¿Cómo puede ser medido el progreso de una meta o tarea? - ¿Cuándo se sabe que una meta o tarea ha sido completada?
Cuantitativo (Métrica)	- Medir el tamaño del proceso (cantidad de *hardgoals*, tareas y acciones). - Medir la eficiencia alcanzada (cantidad de *softgoals*). - Medir el proceso de acciones, tareas, metas y procesos.

Tabla 2. Métricas para el proceso de negocios orientado a metas

Métrica	Descripción
Tamaño del proceso	El tamaño total del proceso medido de acuerdo al número de acciones de las que está formado.
Desempeño del proceso	El progreso en la ejecución de las metas y tareas, así como la relación entre el número de tareas que han sido ejecutadas y el total de acciones de un proceso.
Eficacia en la completitud del proceso	El progreso en el proceso de ejecución como la relación entre el número de acciones involucradas en las *hardgoals* y el tamaño del proceso.

Validación del cumplimiento de una meta o tarea

Para que una tarea T se cumpla, requiere que las *m* acciones que la componen se cumplan. If (A == terminada) then tarea T = terminada.

Para que una meta M se cumpla, requiere que las *n* tareas que la componen se cumplan. If (T == terminada) then meta M = terminada.

En una dependencia el *dependee* no puede satisfacer una meta si el *depender* no cumple con su asignación correspondiente en dicha dependencia, es decir, si no ha "*enviado*" el objeto de la dependencia (*dependum*).

If (*dependum*!= enviado) then *dependee* = pausa

En una contribución, la *hardgoal* puede cumplirse aun cuando la *softgoal* no se cumpla.

Para que una descomposición *D* del tipo *AND* se cumpla, requiere que las *d* metas que componen se cumplan.

If(meta[d1 , d2…dn] ==terminada) then descomposicion D = terminada

Para que una descomposición *D* de tipo *OR* se cumpla, requiere que al menos 1 de las metas que la conforman se cumpla.

If(meta[1…c] == terminada) then descomposicion D =terminada

$c \geq 1$

Para que una descomposición *D* de tipo *XOR* se cumpla, requiere que 1 y sólo 1 de las metas que la conforman se cumpla.

If(meta[c] == terminada) then descomposicion D =terminada

$c \geq 1$

Axiomas. Las métricas definidas en la sección anterior permiten a los analistas determinar el progreso de cada uno de los procesos. Para la determinación de este progreso es necesario tomar en cuenta los siguientes axiomas:

- El actor principal es aquel sobre el cual se centra el proceso de desarrollo.
- El actor principal puede o no ser el actor sistema de software.
- Una tarea *t* está compuesta por un conjunto de acciones *X* y estas tareas representan $1/x$ de una tarea *t*.
- Una tarea puede ser ejecutada sólo cuando todas las acciones que la componen son ejecutadas.
- Una *hardgoal* es satisfecha cuando todas las tareas que la componen son ejecutadas
- Cada actor puede desempeñar *N* metas ($0 \leq N \leq \infty$)
- Cada actor puede tener *N* relaciones con otros actores ($0 \leq N \leq \infty$)
- Las relaciones pueden ser: dependencias o contribuciones
- Una relación se define especificando el actor origen y el actor destino.
- En una relación de dependencia, el actor origen es el *depender* y el actor destino el *dependee*.
- En una relación de contribución, el actor origen proporciona la *softgoal* y el actor destino la *hardgoal*.

En este trabajo de investigación, se utilizan las *softgoals* como la base de medición de la calidad de los procesos de negocios. En este enfoque la eficiencia es tomada como base para incrementar la calidad en el desarrollo de los procesos de negocio. De esta manera, las *softgoals* contribuyen en la satisfacción de las *hardgoals* y es posible asignarles un valor porcentual a cada una de las *softgoals* con un valor del 100% de su eficiencia.

El cálculo del valor de la eficiencia dependerá del número de *softgoals* consideradas para su implementación en los procesos, además del nivel de contribuciones que cada meta tenga con otras metas.

5. Desarrollo de un caso de estudio

Este trabajo de investigación ha sido probado con varios casos de estudio reales, sin embargo en este artículo por cuestiones de espacio solo se presenta el caso de estudio llevado a cabo en la empresa Laminados Inoxidables de Morelos S.A de C.V, específicamente en el departamento de ventas. En este caso de estudio se modelan los procesos que se desarrollan en la empresa para cubrir un pedido solicitado, el cual excede la cantidad de piezas existentes en el almacén.

Para ello se obtienen del almacén la cantidad de piezas disponibles y se solicita al departamento de producción que fabrique las piezas faltantes. Como primer paso estudiamos toda la información que se tiene del proceso desde la perspectiva del framework Tropos, identificando el objetivo del proyecto y los actores que intervienen en los procesos. De esta manera la información introducida a la herramienta *GoalFlow* sería: objetivo del proyecto: satisfacer un pedido solicitado. Actores: ventas, producción, almacén.

Primeramente se deduce una descripción del ambiente organizacional a partir de la información obtenida, donde se busca identificar quienes intervienen en el proceso, qué es lo que hacen y cómo se relacionan con otros participantes.

Siguiendo los conceptos de Tropos, se puede realizar una visualización del entorno organizacional, al analizar la información de las actividades realizadas por los actores y se obtiene el diagrama de actores mostrado en la fig. 4:

Después de refinar las metas, se introduce la información correspondiente a las dependencias entre actores. Al especificar una dependencia se requiere introducir: el nombre de la dependencia, el actor *depender*, el tipo de dependencia, el objeto *dependum*, y el actor *dependee*.

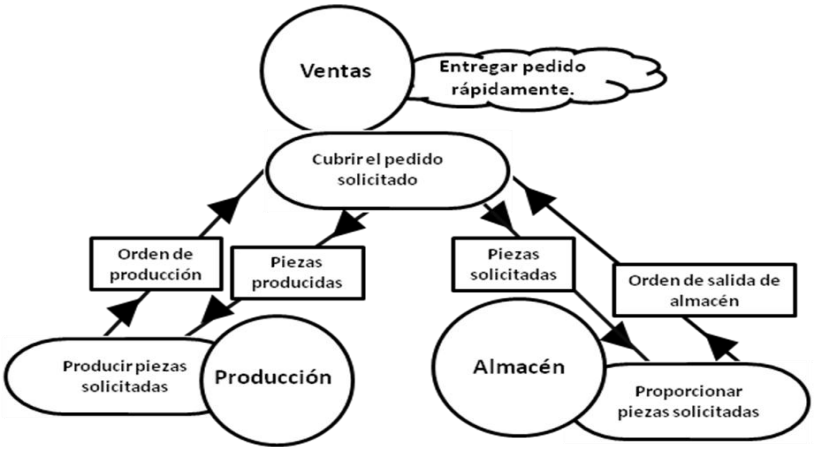


Fig. 4. Diagrama de actores del caso de estudio

Modelado de metas. El objetivo de esta etapa es desarrollar el modelo de metas del framework Tropos, enfocándose principalmente en definir cómo las metas de los actores contribuyen a las metas del negocio. En esta etapa se genera un modelo que refleja la información correspondiente al modelo de metas. Para el caso de estudio Laminados Inoxidables de Morelos S.A de C.V., Goalflow identifica al actor departamento de ventas como el actor principal y se crea el modelo mostrado en la fig. 5:

Monitoreo y control de un proyecto en la herramienta *GoalFlow*. El monitoreo de un proyecto lo lleva a cabo el administrador de un proyecto. Él podrá ver la lista de actores que participan en un proyecto, en la que se señala al actor principal y la cantidad de *hardgoals* asignadas a cada actor, su cumplimiento de las metas y el estado de las dependencias existentes en el flujo de trabajo como se aprecia en la fig. 6 (a). Para llevar a cabo el Control en la herramienta cada actor debe ir marcando el progreso de sus actividades desde su propia terminal. Al accesar a un proyecto, podemos ver detalles sobre el mismo, como la cantidad de metas, tipo y ejecutantes de las mismas, además de las dependencias y el estado en que se encuentran (fig. 6 (b)).

Por motivos de espacio en este artículo se han omitido varias de las funciones con las que cuenta la herramienta *GoalFlow*.

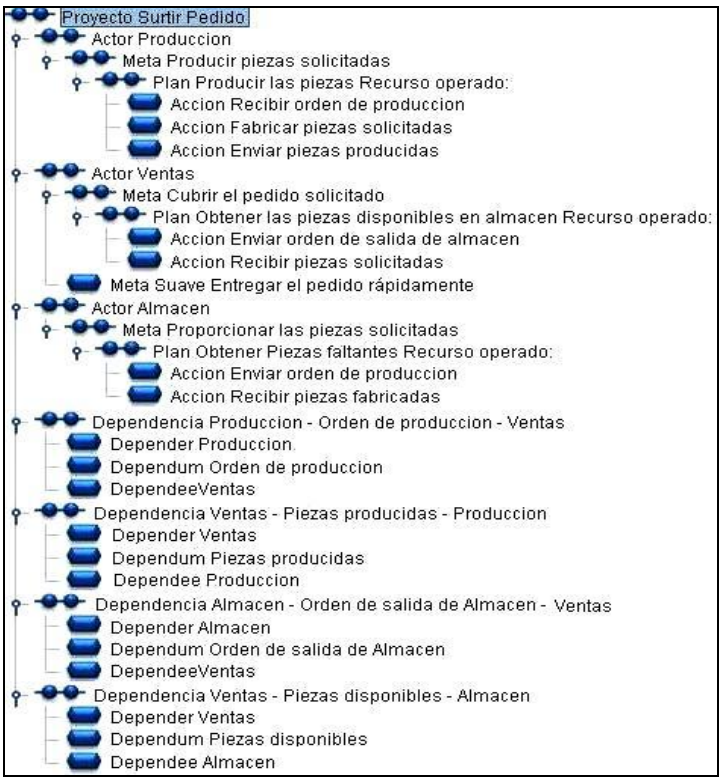
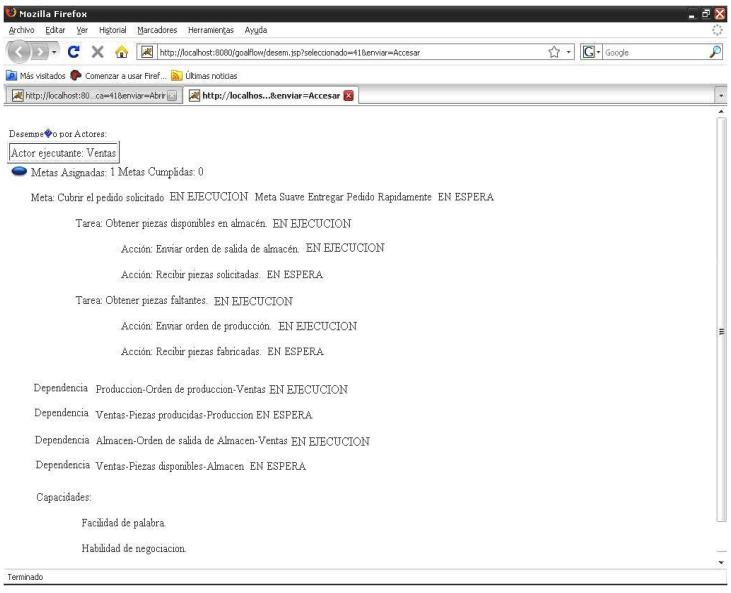
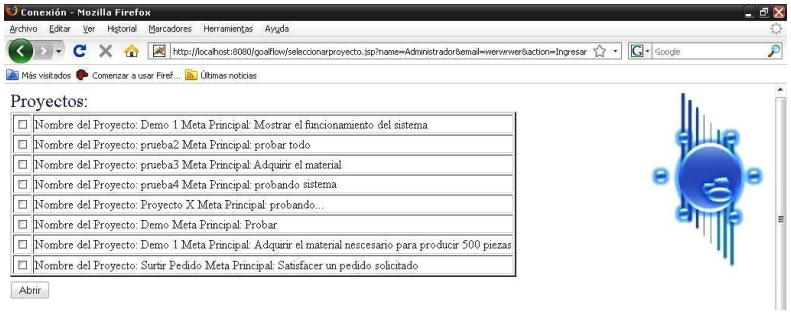


Fig. 5. Modelo de metas de la empresa Laminados Inoxidables de Morelos



a)



b)

Fig. 6. (a)Ventana de monitoreo (b) Ventana de control del caso de estudio Laminados Inoxidables de Morelos

6. Conclusiones

En este artículo se describe nuestra propuesta de combinar los principios del framework Tropos (el cual es una metodología que ayuda al modelado de una organización) con un enfoque orientado a metas y los conceptos de un enfoque *workflow*, lo cual nos permiten tener una vista global de la organización, así como identificar, separar y explicar distintas etapas en un proceso de negocio, con el objetivo de tratarlas de forma acorde con sus características particulares y dar una mayor atención a aquellas que resulten fundamentales en dicho proceso.

Además, *GoalFlow* permite tener un acercamiento detallado de lo que cada actor debe realizar para satisfacer sus dependencias con otros actores y facilita el entendimiento organizacional. Esta herramienta da al responsable de la ejecución de un flujo de trabajo la oportunidad de tomar medidas que inicien y continúen las acciones requeridas para que se ejecuten las metas, como dirigir, instruir o ayudar a los actores para el desarrollo y mejoramiento de sus capacidades, las cuales, también se toman a consideración en la metodología orientada a metas. Es posible utilizar la información generada por la herramienta para definir estrategias de calendarización de metas, técnicas para medir la calidad del desempeño en un proceso, métodos de planeación de tareas orientados a metas, sistemas de calendarización de proyectos, manejo de roles y funciones de monitoreo que amplíen el esquema de cobertura de la herramienta propuesta.

Referencias

1. BonitaSoft: Bonita Workflow, Available: http://www.bonitasoft.com, site visited 12-02-2010.
2. White S. Using BPMN to Model a BPEL Process. BPTrends, 3(3):1-18, 2005.
3. Altimate OpenADMS CMS, http://www.altimate.ca/Workflowdemo.html
4. Van der Aalst W., Ter Hofstede A. YAWL: yet another workflow language. Information Systems, 2005.
5. Jung J.: Mapping of Business Process Models to Workflow Schemata – An Example “Using MEMO-OrgML and XPDL.” Research Reports of the IS Research In-

stitute, University of Koblenz-Landau, No. 47, 2004BonitaSoft: Bonita Workflow, Available: http://www.bonitasoft.com, site visited 12-02-2010.

6. P. Bresciani, P. Giorgini, F. Giunchiglia, J. Mylopoulos, and A. Perini: Tropos: An agent-oriented software development methodology. J. Autonomous Agents and Multi-Agent Systems 8(3):203-236, May 2004.
7. F. Sannicoló, y otros. "The Tropos Modeling Language. A user guide", Technical Report #DIT-02-0061, Universidad de Trento, Italia, 2002.
8. A. Martínez, H. Estrada, y L. A. Gama, "Una guía rápida de la metodología Tropos", Gerencia Tecnológica e Informática, Vol. 7, Nº 19, ISSN 1657-8236, Diciembre 2008.
9. Frank J Romano, National Association for Printing Leadership, Erika Kendra, National Association for Printing Leadership, Publicado por NAPL, 2001.
10. Workflow Management Coalition Terminology and Glossary (WFMC-TC-1011). Technical report. Brussels, Belgium: Workflow Management Coalition, 1996.
11. Basili V., Caldiera, G. y Rombach H.: The Goal Question Metric Approach, in: Marciniak, J. J. (ed.), Encyclopedia of Software Engineering, Wiley, pp. 528–532, 200.

An Ontology proposal for Semantics Checking on Imposed Restrictions by Application Design in Component-based Software Assembling

Manuel Corona-Perez[1], Edgar Castillo-Barrera[1]

[1]Departamento de Tecnologías de Información
CUCEA – Universidad de Guadalajara
Periférico Norte 799 – Mod. L305
Los Belenes 45100
Zapopan, Jalisco, México
manuelcorona@cucea.udg.mx, ecastillo@uaslp.mx
Paper received on 20/08/10, Accepted on 04/10/10.

Abstract. Component-based Software has demonstrated to be a good solution in Software reuse, but this solution fails in guarantee Semantics imposed by the application design. We propose an Ontoloy-based Semantics model to check restrictions to support changes from one application to another. This semantic checking is be developed as a process during the Software Component assembling design which is called Componization.

Keywords: component-based software, ontology, semantics checking, componization

1. Introduction

Component-based software development (CBSD) [5] [1] has shown to be a good solution in software reuse and this help to build applications not necesary from the scratch. We build a new applications using executable pieces of software previosly developed and used in other applications builded from third parties. This software is known as Components Off The Shelf (COTS) [8]. When we reuse Software Components and we assemble them into an application framework we need to adapt them with the aim to get the software solution that we are building. This process of adaptation is called Composition. But not only the goal is to adapt these components, we need also to validate operational semantics, if we want to guarantee that these operations will be satisfied.

Some restrictions come from intercomponent operations, defined by *contracts*, and other restrictions are defined from the System Architect of the application.

To validate these operational restrictions, we propose an *Ontology-based Semantics*.

An Ontology is an explicit specification of a conceptualization [4], It set up semantics on specific domain knowledge in a formal and all-porpouse way [2]. We

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 93-98.

need an Ontology to represent components and the operational semantics between components at the user level(System Architect). To represent the meaning and consistency of things we use *Semantics*, and an Ontology is one o the best way to model semantics. Our purpose is an *Ontology-Based Semantics* [6] to model semantic validation on Component-based Software Assembling.

2. Context Semantics Model in Components Assembling

When we use Components in a Software Applications, we must guarantee that some constraints previously imposed for the Application must be fulfilled.

This way we can focus in three different ways where semantics may occur. The first one is *Component to Component Assembling*, the second one is *Component to Framework Assembling*, and the third one is *The Whole application*.

Semantics must occurs between the Components, between the Component and the Framework. In essence we call this, *Referential Semantics Componization*. Furthermore the semantics constraints for the application as a whole are called *Integral Semantics Componization*.

2.1. Referential Semantics Componization

Let´s A and B be two Components that are working together and F an application Framework for which we define the following set of rules:

$$\text{a) } R_A^B = \{r_1, r_2, r_3, ..., r_n\}$$

Where **a)** is the set of semantic restrictions between Components A and B.

$$\text{b) } R_A^F = \{s_1, s_2, s_3, ..., s_n\}$$

Meanwhile **b)** is the set of semantic restrictions between Component A and the Framework of the application.

This means that R_A^B and R_A^F are two restriction Sets imposed between A and B and between the Framework and then Component A, we do not care who is imposing any of the restrictions.

Example:
A= Thermometer Component which process temperatures in Celsius and Kelvin Scales.

B= Control Heat Component whose task is to mantain a specific temperature in an oven.
F= Framework in which a Metals Alloy Application is being implemented.

$$R_A^B = \{PlumbAlloyTemperature, IronAlloyTemperaturte, CopperAlloyTemperature\}$$

R_A^B means the necesary temperatures restrictions of fluid melting metals in a Metals Mix Process to produce a specific Alloy.

2.2. Integral Semantics Componization

Let's P a Metals Alloy Application, A and B the Components before defined. Such as:

$$R_I^P = \{t_1, t_2, t_3, ..., t_n\}$$

Where R_I^P represents semantics restrictions set that the application P must be fulfilled.

Example:

P = Metals Alloy Application

Then:

$$R_I^P = \{MetalsAlloyTemperature, SequenceOfMixingMetals, IronThemperatureInAlloy, PlumbThemperatureInAlloy, CopperThemperatureInAlloy\}$$

Which ones represents restrictions that the application impose during the Metals Alloy Process.

3. Proposed Ontology for the Context Semantics Model

An Ontology shows to be a good way to represents Semantics. In this case, we are working on building an Ontology that implements the Context Semantics Model. The first step is to describe an Ontology for a Software Component similar to the

one proposed by Dong, Liu, He, He [2] in Figure 1. Using this Component Ontology we can relate one component to another and also among Components and the Framework.

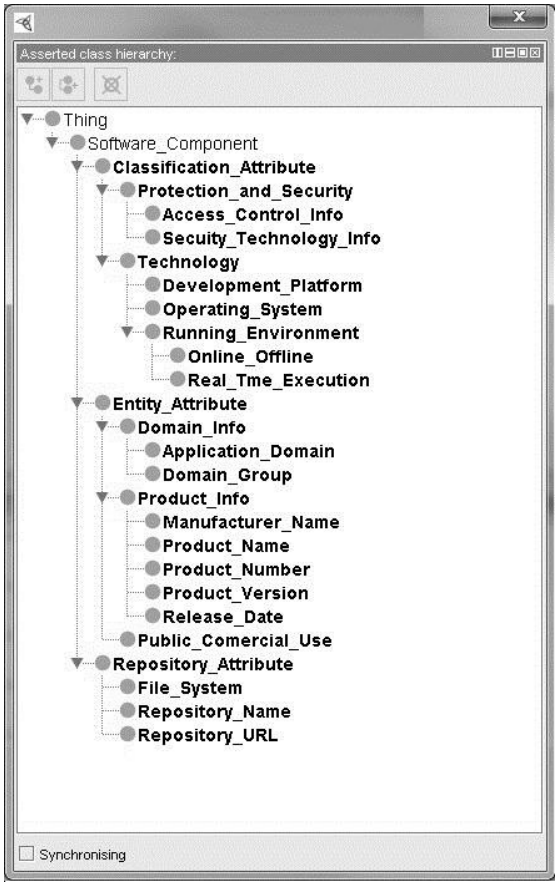


Figure 1: A Software Component Ontology

Using the component methods we can implement the Context Semantics model and generate a complete semantic verification process. We can also consider the semantic context for the specific application where the Software Components are being used.

The aim in this task must be mapping each relation restriction to one relation operation. This should be done by following the pre-conditions and post-conditions imposed by *the interfaces*. Also we need to map the methods in the Method class and all its parameters in the same Ontology. Each element belonging to the relation restricions set in the Context Semantics Model must have a relation between the Method and the Parameters where Components are interacting.

Our proposed contribution as a solution in this work will define relation restrictions that were explained in the Context Semantics Model. By Building relations on the generated Ontology, we establish the operational semantics among methods that each component has defined in its interfaces. We use another Ontology that describe this kind of relations and here it shows where we are going to apply our semantic model.This other Ontology description was done by Linhalis, de Mattos, de Abreu. [3] Figure 2 and is used only to show where our Semantic model will be implemented.

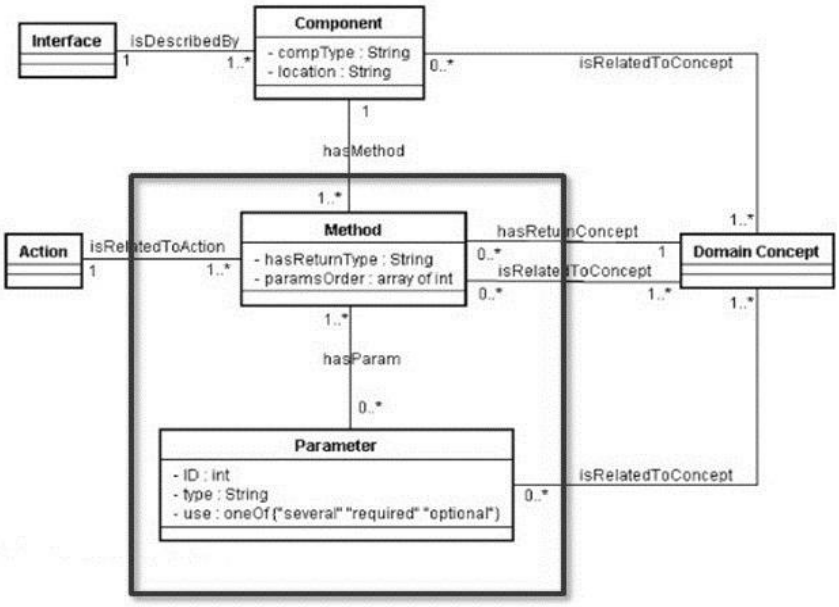


Figure 2: Context Semantics Model is implemented among the Components using Method class and the Parameters class relations

4. Conclusion and Future work

We are building the Context Semantics Model and some Ontology classes in **Protégé** [7] and at the same time we have the abstraction of the Model and the solution domain in a sort to implement relation restriction sets on the semantic context of the application. Using Software component not only requieres sintax and semantics over the components plugins, we are adding a context details because to use components, we need to guarantee additional restrintions when we are using

from one aplication to another. Our work under development is to get a complete Ontology to represent our model doing the relations between the classes and to implement tools to do the Model transformations in the Component-based Software Assembling.
We are developing this work with three main advantages in mind and these are:

- To automate the semantics checks, generating a *glue code* that helps in the implementation of the application.
- To guarantee consistency on the execution of the application, lowering the number of exception errors.
- To get a better performance and realibility of the whole application *per sé*.

Under this perspective, we need to build an environment to run test cases to probe our model. In future papers we will show this results.

References

1. Clemens Szyperski, Dominik Gruntz, Stephan Murer. *Component Software, Beyond Object-Oriented Programming*. Second Edition, Addison-Wesley, 2002.
2. Dang Song, Wudong Liu, Yangfan He and Keqing He. *Ontology Application in Software Component Registry to Achieve Semantic Interoperability*. Wuhan University, China, 2005.
3. Flavia Linhalis, Renata Pontin de Mattos Fortes, Dilvan de Abreu Moreira. *OntoMap: an ontology-based architecture to perform the semantic mapping between an interlingua and software components*. Science Computing and Mathematics Institute (ICMC), University of SÃ£o Paulo (USP), 2009.
4. Thomas R. Gruber. *A Translation Approach to Portable Ontology Specification*. Stanford University, USA, 1993.
5. Ivica Crnkovic, Magnus Larsson. *Building Reliable Component-BasedSoftware Systems*. Editors, Artech House), 2002.
6. Mihai Ciocoiu, Dana S Nau. *Ontology-Based Semantics*. University of Maryland, USA, 1999.
7. Stanford University, University Manchester. *Protege, A free open source Ontology Editor*. http://protege.stanford.edu/, 2006.
8. Xiong Xie, Weishi Zhang. *Research on Software Component Adaptation Based on Semantic Specification*. Department of Computer Science and Technology, Dalian Maritime University, Dalian, China, 2007.

MIKONOS: A Middleware-oriented Integrated Architecture for Clinical Knowledge based on Computational Intelligence Techniques

[1]Giner Alor-Hernández, [1]Guillermo Cortes-Robles, [2]Alejandro Rodríguez-González, [1]Ruben Posada-Gómez, [2]Juan Miguel Gómez-Berbís, [1]Ulises Juárez-Martínez, [1]Alberto Aguilar-Lasserre

[1]Division of Research and Postgraduate Studies, Instituto Tecnológico de Orizaba.
[2]Computer Science Department, Universidad Carlos III de Madrid
Email:{galor, gcortes, rposada, ujuarez, aaguilar }@itorizaba.edu.mx, { alejandro.rodriguez ,juanmiguel.gomez,}@uc3m.es
Paper received on 23/08/10, Accepted on28/09/10.

Abstract- This proposal outlines the development of an infrastructure for the knowledge integration and interoperability in medical entities. The impact of this proposal resides in improving the quality of healthcare services by using semantic Web and computational intelligence techniques. As salient contributions, this work presents the use of semantic Web and Case-Based Reasoning techniques used for medical diagnosis and for developing a medical knowledge memory. Finally, we emphasize our work in comparison with related works in this field and we outline the future work.

Keywords: Case-Based Reasoning; Clinical Knowledge, Medical Diagnosis; Semantic Web.

1 Introduction

Healthcare is a field in which accurate record keeping and communication are critical and yet in which the use of computing and networking technology lags behind other fields. In current healthcare, information is conveyed from one healthcare professional to another through paper notes or personal communication. In an age of electronic record keeping and communication, the healthcare industry is still tied to paper documents that are easily mislaid, often illegible, and easy to forge. When multiple healthcare professionals and facilities are involved in providing healthcare for a patient, the healthcare services provided are not often coordinated [1]. For instance, in low-income countries health care services have become fragmented and organized by a specific health problem. Organization by a specific health problem or specialization usually means people need to visit separate and specialized clinics depending on their health problem [2]. Other aspect to be considered is the interoperability. Data sharing and exchange have always been considered the biggest challenge

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 99-111.

for autonomous heterogeneous medical information systems. It is frequently observed that the knowledge required to solve a healthcare problem is spatially distributed in different locations.

Nowadays, healthcare systems require architectures that offer more than enhanced interoperability and integration potential at local level. In recent years, Service-Oriented Architecture (SOA) has emerged as an architectural paradigm for creating and managing services that can software functionality and access these functions, assets, and pieces of information with a common interface regardless of the location or technical makeup of the function or piece of data [3]. SOA is playing major role in the development of healthcare system which helps to exchange the information between similar and dissimilar medicine applications. Clinical knowledge integration based on SOA can address some of these issues and problems. In this sense, a distributed Integration architecture for clinical knowledge integration through semantic Web and Case-Based Reasoning techniques for developing a knowledge capitalization tool offers to the user a way to automatically store and reuse the experience accumulated in the decision-making process for medical diagnosis. Having this into account, this paper presents a middleware-oriented integrated architecture for Clinical Knowledge Integration based on Computational Intelligence Techniques for medical diagnosis and for developing a medical knowledge memory.

2 Middleware-oriented Integrated Architecture

With an SOA infrastructure, we can represent software functionality as discoverable services on a network. SOAs have been around for many years but the distinctive feature of the SOAs is that they are based on Internet and Web standards, in particular, Web services. Web services are designed to provide interoperability among diverse applications. Nowadays, a growing number of enterprises, commercial and healthcare systems are redefining their processes under this technology. The platform and language independence of the Web services programming interfaces enable the seamless integration of heterogeneous Web based-systems. In this work, we proposed an SOA-based integration architecture. In Fig.1 each component of the architecture has a defined function, which are briefly described next.

Healthcare services registry: In this component, the healthcare services described as Web services descriptions and metadata information are stored. The category, properties, capacities, localization and access information of available services are inside the metadata information. For the knowledge search, classification and acquisition, the use of OWL-S ontologies in the clinical context is outlined. OWL-S is a Web service ontology, which supplies Web service providers with a core set of markup language, constructs for describing the properties and capabilities of their Web services in unambiguous, computer-interpretable form. Ontology defines the terms for using, describing and representing a knowledge field. For the development of ontologies was considered the use of METHONTOLOGY. The main steps of METHONTOLOGY are: (1) specification of the goal, the reach and granularity of the ontology, (2) Conceptualization that helps to organize and to structure the acquired knowledge using independent representation languages of the implementation languages, (3) Implementation that consists in the formalization and implemen-

tation of the conceptual pattern with formal languages such as Ontolingua, RDF / S (Resource Description Framework / Schema), OWL, and (4) Evaluation. This approach used conceptual maps for the development of ontologies for modeling the knowledge domain. The conceptual maps were built with CmapTools. The OntoTermTM system for the management of terminological databases was used [4]. This system allows user the creation of customized terminological database on a knowledge domain. Protégé was used for building and modeling the OWL ontology which was tested in the Jena framework. Jena allows the use of ontologies for developing Web applications. The queries are carried out with SPARQL.

SOAP Message Analyzer: This component determines the structure and content of the documents exchanged in healthcare services involved. This component determines the information involved of the incoming messages by means of XML parsers and tools. A DOM API is used to generate the tree structure of the messages, whereas SAX is used to determine the application logic for every node in the messages. A set of Java classes based on JAXP [5] were developed to build the XML parser.

Discovery Service: It is a component used to discover healthcare services implementations. Given the dynamic environment in medicine, the power of being able to find healthcare services on the fly is highly desirable. A key step in achieving this capability is the automated discovery of healthcare services described as Web services. These Web services can be obtained from suitable service providers and can be combined into innovative and attractive ways. When there is more than one service provider of the same function, it can be used to choose one service based on the user's requirements. Inside the discovery service, there is a query formulator which builds queries based on the domain ontology that will be sent to the healthcare services registry. This module retrieves a set of suitable services selected from the previous step and creates feasible/compatible sets of services ready for binding.

Dynamic Binding Service: It is a component that binds compatible healthcare services described as Web services. The binding of a healthcare service refers to how degree of coupling with other service is. For instance, the technology of one healthcare service provider might be incompatible with that of another even though the capabilities of both of them match with some requirements. In this sense, the module acts as an API wrapper that maps the interface source or target healthcare services to a common interface supported by our approach.|

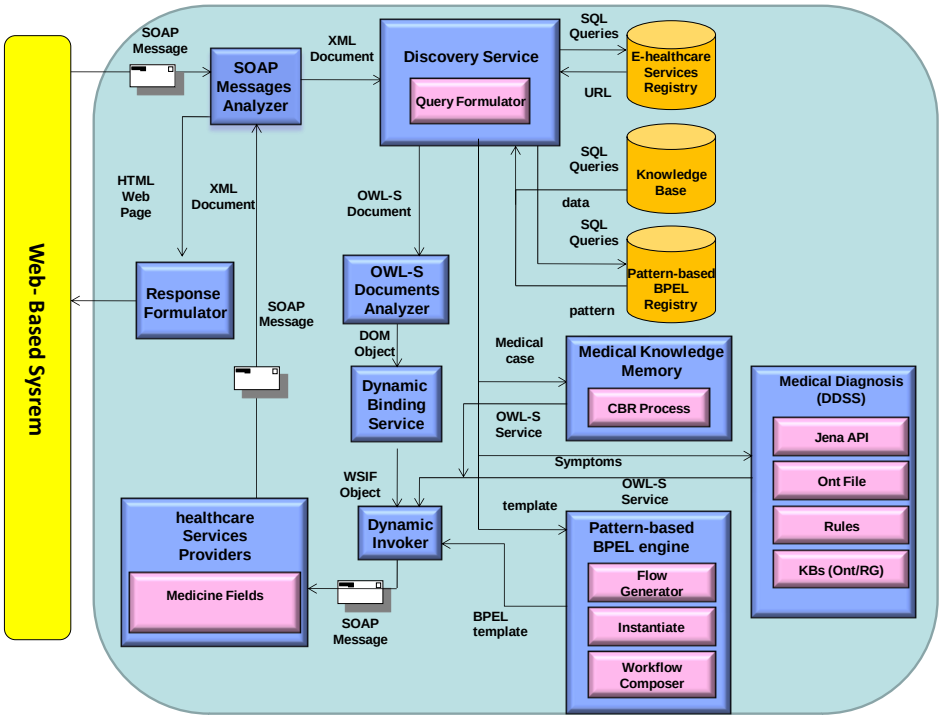


Fig. 1 General architecture for the clinical knowledge integration system

Dynamic Invoker: This component transforms data from one format to another. This component can be seen as a data transfer object which contains the data (i.e., request or response) flowing between the requester to the provider applications of e-healthcare services. We used Web Services Invocation Framework (WSIF) that is a simple Java API for invoking/consuming e-healthcare services described as Web services, no matter how or where the services are provided [6].

Response Formulator: This module receives the responses from an e-healthcare service provider about a requested clinical service. This module retrieves useful information from the responses and builds a XML document with information coming from the service registry and the invocations' responses. This XML document is presented in HTML format using the Extensible Style sheet Language (XSL).

OWL-S Document Analyzer: This internal validates OWL-S documents that describe e-healthcare services. OWL-S documents employ XML Schema for the specification of information items either technical information or healthcare services operations. In this context, this component reports the healthcare services operations, input and output parameters, and their data types in a XML DOM tree.

Workflow Engine: This module coordinates Web services by using a BPEL-based healthcare services language. It consists of building at design time a fully instantiated workflow description where Web services descriptions are dynamically defined at execution time. We have designed and implemented a repository of generic BPEL workflow definitions which describe increasingly complex forms of re-

curring situations abstracted from the various stages from an integration process. This repository contains workflow patterns of interactions involved in integration scenarios for healthcare services. These workflows patterns describe the types of interactions behind each healthcare service, and the types of messages that are exchanged in each interaction.

Healthcare services providers: This component represents information subsystems that have attended the clinical diagnosis by using CBR and Semantic Web techniques. These subsystems are located in different domains of health such as pathology, neonatal, pediatrics, among others.

Medical case-based memory: In this module users have access to a set of solutions derived from real problems and the praxis of an expert panel. The medical case memory has its theoretical foundation over the Case-Based Reasoning (CBR) process. The subjacent approach of the CBR was developed in the Artificial Intelligence (AI) field as a flexible and easy to evolve tool for problem solving. Furthermore, artificial intelligence and more precisely knowledge management approaches try to use past experiences in order to solve new problems.

In the next section, we describe how semantic web technologies and the case-based reasoning process were integrated in our architecture for developing a medical knowledge memory and a medical diagnosis system. These are the more relevant aspects of our approach.

3 Using semantic Web and Case-Based Reasoning techniques for medical diagnosis and for developing a knowledge capitalization tool

Semantic Web Technologies [7] have emerged as an attempt to provide machine-processable metadata to the ever increasing information resources on the Web. These software standards and methodologies may be applied to particular application domains in order to make maximum use of Semantic Web representation specifications such as RDF. Such specifications can define the terminology of a scientific domain as a computer-interpretable ontology, using XML as the syntax for data interchange. In our work, we have integrated a medical diagnosis system based on Semantic Web capabilities namely ODDIN [8]. The system was initially constructed and named Ontology DDx (ODDx) where DDx stands for "differential diagnosis". This work was based on the development of this system in Java language programming, with the use of ontologies as knowledge database or knowledge representation in order to allow inference engines to work with the system. The ontology used by the system was built using the international classification of diseases version 10 [9], diseases implemented by the WHO [10]. Figure 2 represents the ontology built and some of the relationships that exists between classes and properties. The most of the relationships (ObjectTypeProperties) established in the ontology were used to establish a relation between Diseases and Symptoms, or Diseases and Laboratory Tests. However, there are other kind of relationships like the relation between Medicines and Symptoms, in order to set possible reactions (signs or symptoms) that a medicine can provoke. The DataTypeProperties are used to describe the sign, disease or

any other item of the ontology. The typical properties are set, like name, description, etc. One important property is "code" that identify the code of the sign (normally, using ICD codes). The figure 3 shows how the diseases and their relationship with the symptoms or laboratory tests are described.

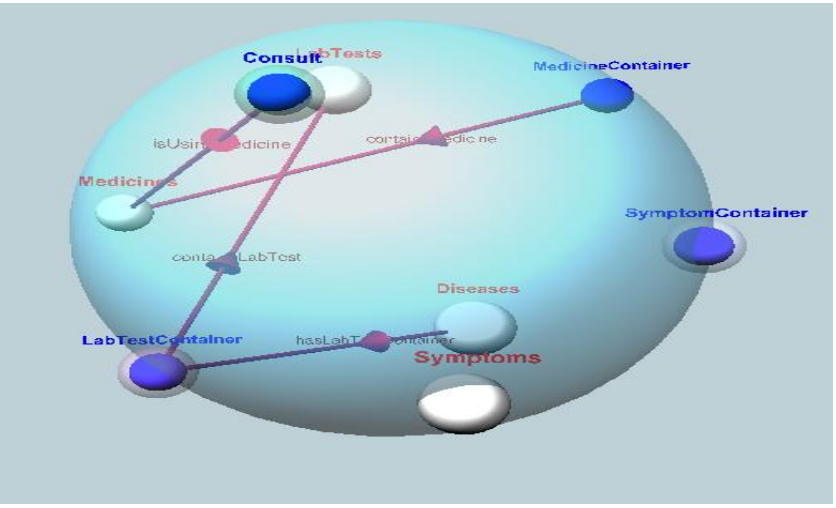


Fig. 2. Representation of the ontology for medical diagnosis.

The main representation of this disease in the ontology is as follows:

- hasSymptom: Used to make relationship between diseases and symptoms.
  - SYM A
  - SYM E
- hasLabTest: Used to make relationship between diseases and laboratory tests.
  - LT 1

All the ObjectTypeProperties connect instances, so, in this case for example for "SYM_A", the connection is as follows:

DIS_X (Instance of DIS_X_Class) hasSymptom SYM_A (Instance of SYM_A_Class)

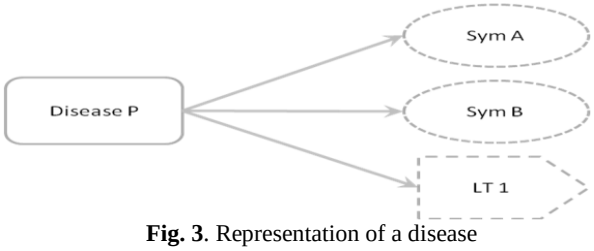


Fig. 3. Representation of a disease

The inference system used to build this system was based in Jena Rules [11]. The description represented in Figure 3 is translated from Description Logics (DL) to this kind of rules. The Inference Architecture is represented in Figure 4. In these cases, forward chaining and backward chaining rules were used.

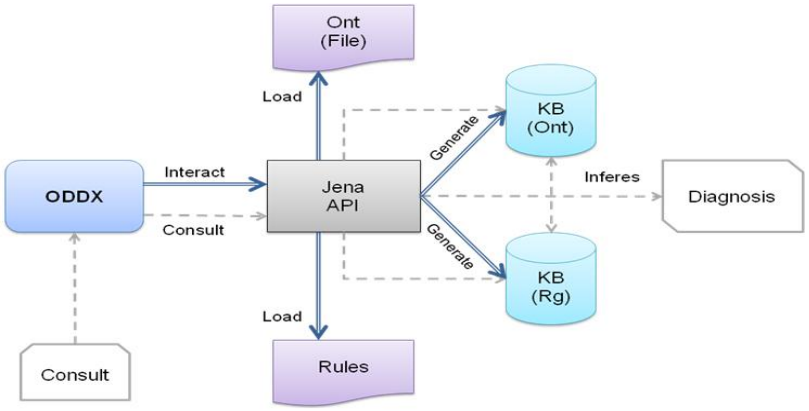


Fig. 4. Representation of the architecture of the inference system

In another hand, we implemented a Case-Based Reasoning (CBR) process for building a medical knowledge memory that facilitates knowledge transfer and training. Because this memory facilitates the storage and transfer of knowledge that could be reused when a similar context is identified, our proposal address a student centered teaching /learning approach and active student participation in the process of medical assessments and treatments. The main idea in CBR is that similar problems have similar solutions. Basically in a CBR system, users search to solve a new problem by establishing some common characteristics between the initial problem (the problem to solve or target problem) and some previous solved problems. Then, the CBR process reutilizes and adapts earlier successful solutions in order to solve the new problem, which in fact imitates everyday human problem solving process. CBR traces its roots to the work of Schank [12] on dynamic memory. In our proposal, we describe the memory-based approach to reasoning, which means that human memory is dynamic because the experiences acquired while solving problems, radically modify the way to face new problems. Thus, these new experiences which inherently contain some lessons learned in a particular context could be used to face new problems. A case is then a solved problem in a particular domain that has been characterized, captured and indexed in a memory for further utilization. A case is usually composed of three elements: the problem description, the solutions applied and its intrinsic result. Eventually it is also added the strategy to deploy or apply the solution. Then an expression of a case is: Case (Pb, Sol (Pb), Re), where Pb represents the problem description, Sol the associated solution for this problem and Re, the obtained result that could be success or a failure.

CBR involves a process where learning is a capital benefit. If a CBR system is intensely used, its case-memory will have more problem situations and then its ca-

pacity to create more solutions will increase. CBR is a very flexible process because as cases are added to the memory, the system should be able to reason in a wider variety of situations and with a higher degree of refinement and success. A CBR system assists users when reasoning with incomplete or imprecise data and concepts. This condition is materialized when measuring similarity between two problems. The difference should be explained and surmounted in order to propose an efficient solution and thus, impelling reasoning. Avoiding repeating all the steps that need to be taken to arrive at a solution, because the solving process starts remembering previous solution. This process also allows capitalize the steps taken to reach one solution in order to be reused for solving other problems. The CBR process has a wide application domain. It can be used for many purposes, such as creating a plan, making a diagnosis, and arguing a point of view. Therefore, the data dealt with by a CBR system are able to take many forms, and the retrieval and adaptation methods will also vary [13]. Thus, a CBR system solves problems by acquiring knowledge and reusing this knowledge in a similar context. This is a four stages process usually called the 4R´s process for retrieve, reuse, revise and retain [14], [15]. These processes have been implemented using the JColibri 2.1framework [16] useful to build CBR systems. The main objective of this system is to offer a knowledge base for pneumonia treatment in the neonatology department. Figure 5 shows the essential components for a case stored in the knowledge memory. This system stores knowledge extracted from a panel of experts and then this knowledge is socialized in a collaborative Web by proposing a comparative interface. In this interface a non-expert could propose a treatment for a solved pneumonia case and then this treatment is compared with the treatment proposed by an expert to measure the similarity. Figure 6 shows this collaborative interface.

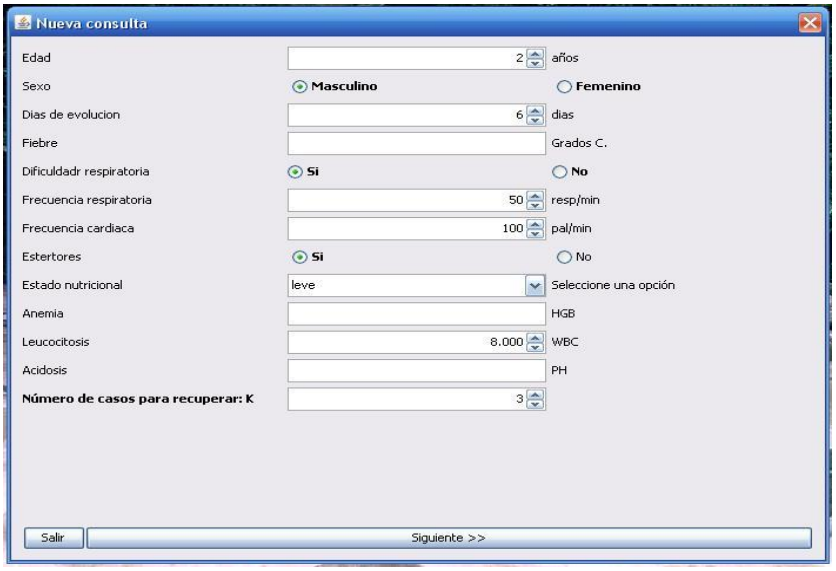


Fig 5. Case description using JColibri

Fig.6. The collaborative interface

We validated our proposal developing a Web-based system to improve the quality of the services that offer the regional hospital of Rio Blanco (HRRB) located in Veracruz, Mexico with the purpose of helping in the process of decision taking, research and specialist's formation and medical residents. It is a medical center which has different medicine fields.

4 Related Works and Discussion

In recent years, different approaches have been proposed for building healthcare systems by using Semantic Web technologies. These initiatives go from the description of integration architectures to the use of emergent IT. Some of these works have obtained outstanding partial results for the clinical knowledge integration. In comparison with our work, these proposals do not cover all the features presented here. For instance, service-oriented architectures in healthcare systems are presented in [1, 17, 18, 19, 20, 21, 22, 23, 24]. These works only propose layered architectures for support of doctors, nurses and patients but they did not implement computational intelligence techniques.

The development of Medical Differential Diagnosis and Therapy systems using computational intelligence has gained momentum over the last years [25]. Approaches in research which apply the use of combined techniques such as the current one include neuro-fuzzy methods [26], the application of genetic algorithms (GAs) for rule selection [27], or the unification of genetic algorithms with fuzzy clustering techniques [28]. Other methods apply a single approach, applying neural networks, GAs, or fuzzy inference systems [29, 30, 31]. A large number of medical diagnosis

systems is also available, including DiagnosMD, DXPlain [32], eMedicine, Isabel [33], IWSMD [34], EasyDiagnosis, ADM [35], and Your Diagnosis. Other types of systems exist, such as those based on the PDP model [36]. Nevertheless, the application of Semantic Web to this concrete field (medical diagnosis in particular) is not fully exploited. Actually, exists a lot of papers about clinical knowledge representation like OBO Foundries [37], Biological Ontologies [38], Ontology-Based Support for Human Disease Study and Medical Ontologies to support human disease research and control [39], and Relations in biomedical ontologies [40]. Other works, talk about the construction of a medical domain ontology to build medical decision support systems [41], or the use of ontologies for the patient modeling in medical management systems [42].

Finally, some works have used Case-Based Reasoning for developing medical diagnosis systems and medical trainers. A case-based analysis of health care data quality problems in a situation, where data of diabetes patient are combined from different information systems is presented in [43]. In [44], a medical reasoning system by using differential diagnosis is described. This style of reasoning consists of the following three kinds of reasoning processes: exclusive reasoning, inclusive reasoning and detection of complications, which corresponds to screening, differential diagnosis, and close examination of each case.

5 Conclusions

In this work was presented an SOA-based approach with workflow-enabled capabilities beyond integration and interoperability within the healthcare organization. The proposed architecture enables clinical knowledge integration through a set of loosely-coupled software components as a proof-of-concept implementation by using semantic Web and computational intelligence techniques for medical diagnosis and for developing a medical knowledge memory. This integration covers several advantages: the system is useful to assist decision taking and to facilitate service selection for a patient; it is a very valuable resource in research activities; it is an axis that support organizational learning because experiences are shared through an organizational memory, increasing efficiency in the learning process of medical residents and students, for mentioning a few.

The proposed architecture follows design principles like interoperability, integration, dynamism, abstraction and scalability. Future research will focus in designing and building a healthcare service bus that will be able to manage rich, complex, and high-value workflows across applications, departments, and geographic locations. A healthcare service bus will be a standards-based platform that will combine messaging, web services, data transformation, and intelligent routing to reliably connect and coordinate the interaction of diverse applications across extended healthcare systems with transactional integrity.

Acknowledgements

This work is supported by the General Council of Superior Technological Education of Mexico (DGEST). Additionally, this work is sponsored by the National

Council of Science and Technology (CONACYT) and the Public Education Secretary (SEP) through PROMEP. Furthermore, this work is supported by the Spanish Ministry of Industry, Tourism, and Commerce under the EUREKA project SITIO (TSI-020400-2009-148), SONAR2 (TSI-020100-2008-665 and GO2 (TSI-020400-2009-127).

References

1. Firat Kart, et al. "Building a Distributed E-Healthcare System Using SOA". IT Professional. March/April 2008. IEEE Prees. pp 24-30.
2. Briggs CJ, et al. Strategies for integrating primary health services in middle- and low-income countries at the point of delivery. Cochrane Database of Systematic Reviews 2006, Issue 1. Art. No.: CD003318. DOI: 10.1002/14651858.CD003318.pub2
3. Mike P. Papazoglou. "Service-Oriented Computing: Concepts, Characteristics and Directions". Proceedings of the Fourth International Conference on Web Information Systems Engineering (WISE'03). IEEE Press.
4. Arianne Reimerink, "Sciscribe: una aplicación de software para redactar y traducir artículos de investigación", Panace. Val. VIII, n° 25, Primer semestre, 2007.
5. Jeff Suttor, et al. "Java API for XML Processing (JAXP) version 1.3". Sun Microsystems, Inc. Final Release, September 1, 2004
6. Matthew J. Duftler, et al. "Web Services Invocation Framework (WSIF)". IBM T.J. Watson Research Center. August 9, 2001
7. Berners-Lee T., et al. The Semantic Web. Scientific American, 284(5), 2001. pp 34-44.
8. Garcia-Crespo, A., et al. "ODDIN: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements". Expert Systems With Applications, 2009.
9. ICD-10. International Statistical Classification of Diseases and Related Health Problems. Available in http://www.who.int/classifications/icd/en/
10. WHO. World Health Organization. Avalaible in http://www.who.int/en/
11. Reynolds, D., Jena Rules. Proceedings of the 2006 Jena User Conference., 2006.
12. Schank R., Dynamic Memory: A Theory of Learning in Computers and People, Cambridge University Press, 1982.
13. Avramenko Y., et al. Case Based Design, Applications in Process Engineering, Springer, 2008.
14. Aamodt A., et al. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications. IOS Press, Vol. 7: 1, pp. 39-59, 1994.
15. Aamodt A., Knowledge Acquisition and Learning by Experience - The Role of Case-Specific Knowledge, Academic Press, 1995.
16. Recio García J. A., "jCOLIBRI: A multi-level platform for building and generating CBR systems", Universidad Complutense de Madrid, Department of Software Engineering and Artificial Intelligence, Ph.D. thesis, 2008.
17. Asadullah Shaikh, et al. "The Role of Service Oriented Architecture in Telemedicine Healthcare System". Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems. IEEE Press. 2009, pp. 208-214.
18. Andreas Hein, et al. "A Service Oriented Platform for Health Services and Ambient Assisted Living". Proceedings of the 2009 International Conference on Advanced Information Networking and Applications Workshops. IEEE Press. pp 531-537.
19. Eugen Vasilescu, et al. "Service Oriented Architecture (SOA) Implications for Large Scale Distributed Health Care Enterprises". Proceedings of the 1st Distributed Diagnosis and Home Healthcare Conference. IEEE Press. 2006, pp 91-94.

20. Mahmoud Barhamgi, et al. “A framework for Data and Web Services semantic mediation in Peer-to-Peer based Medical Information Systems”. Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems. IEEE Press. 2006
21. Wail M. Omar, et al. “Service Oriented Architecture for e-Health support services based on Grid Computing Overlay”. Proceedings of the IEEE International Conference on Services Computing. IEEE Press. 2006.
22. Firat Kart, et al. “A Distributed e-Healthcare System Based on the Service Oriented Architecture”. Proceedings of the 2007 IEEE International Conference on Services Computing . IEEE Press.
23. Kamran Sartipi, et al. “Mined-knowledge and Decision Support Services in Electronic Health”. Proceeding of the International Workshop on Systems Development in SOA Environments. IEEE Press. 2007, pp 1-6.
24. David Budgen, et al. “A Data Integration Broker for Healthcare Systems”. IEEE Computer. IEEE Press. 2007. pp 34-41.
25. Zhao W., et al. “A Model of Intelligent Distributed Diagnosis and Therapy System Based on Mobile Agent and Ontology”. Proceedings of the 8th International Conference on High-Performance Computing in the Asia-Pacific Region. IEEE Press. 2005
26. Nauck, D., et al. ”A neuro-fuzzy method to learn fuzzy classification rules from data”. Fuzzy Sets and Systems, 8(3), 1999. pp. 277-288
27. Ishibuchi, H., et al. “Performance Evaluation of Fuzzy Classifier Systems for Multidimensional Pattern Classification Problems”. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 29 (5), 1999. pp 601–618.
28. Setnes, M., et al. “Fuzzy relational classifier trained by fuzzy clustering”. IEEE Trans. SMC B 29, 1999. pp 619-625.
29. Song, H., et al. “New methodology of Computer Aided Diagnostic System on Breast Cancer”. Proceedings of International Symposium on Neural Networks, 2005, pp 780-789.
30. Jang, J. “ANFIS: Adaptive-Network based Fuzzy Inference System”. IEEE Transactions on Systems, Man, and Cybernetics, 23(3), 1993. pp 665-685.
31. Pena-Reyes, C., et al. “Designing Breast Cancer Diagnostic System via a Hybrid Fuzzy-Genetic Methodology”. Proceedings of IEEE International Fuzzy Systems Conference, 1999, pp 135-139.
32. London, S. “DXplain: a Web-based diagnostic decision support system for medical students”, Medical reference services quarterly, 17(2), 1998, pp 17-28.
33. Ramnarayan, P., et al. “ISABEL: a web-based differential diagnostic aid for paediatrics: results from an initial performance evaluation”, Archives of disease in childhood, 88(5), 2003, pp 408-413.
34. Ibrahim, M., et al. “IWSMD: An Intelligent Web Service based Medical Diagnosis System”, The 6th International Conference on Informatics and Systems 2008, El Cairo Egypt.
35. Le Beux, P., et al. “ ADM : A Diagnostic Aid Knowledge Base for General Practitioners”. IMIA International Conference on Medical Informatics and Medical Education. 1998, pp. 223-229.
36. Saito, K., et al. “Medical diagnostic expert system based on PDP model”. IEEE International Conference on Neuronal Networks, 1988, pp 255-262.
37. Smith, B., et al. “ The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration”. Nature Biotechnology 25(11), 2007, 1251-1255.
38. Lambrix, P., et al. “Biological ontologies”.Semantic Web Revolutionizing Knowledge Discovery in the Life Sciences (pp. 85-89). Berlin: Springer. 2007.
39. Hadzic, M., et al. “Medical Ontologies to support human disease research and control”. International Journal of Web and Grid Services, 1(2), 2005, pp 139-150.

40. Smith, B., et al. “Relations in biomedical ontologies”. Genome Biology, 6(5), 2005, pp 1425-1433.
41. Djedidi, R., et al. "Medical Domain Ontology Construction Approach: A Basis for Medical Decision Support", Twentieth IEEE International Symposium on Computer-Based Medical Systems, 2007, pp.509-511.
42. Bouamrane, M., et al. “Using Ontologies for an Intelligent Patient Modelling, Adaptation and Management System”. Proceedings of the OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008. Part II on On the Move to Meaningful Internet Systems. 2008, pp. 1458-1470.
43. Merja Miettinen, et al. "Information Quality in Healthcare: Coherence of Data Compared Between Organization's Electronic Patient Records".Proceedings of the 21st IEEE International Symposium on Computer-Based Medical Systems 2008.IEEE Press. pp 488-493.
44. Shusaku Tsumoto, et al. "Clinical Decision Support based on Mobile Telecommunication Systems". Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence. IEEE Press.

Adaptación de un Modelo de Desarrollo de Software para la Construcción de Software Usable con Calidad

Fabio Armando García Toribio[1], Juan Manuel Gómez Reynoso[1]

1Universidad Autónoma de Aguascalientes,
Centro de Ciencias Básicas, México.fgarc@live.com.m
Paper received on 25/08/10, Accepted on 05/10/10.

Abstract. Software is built in an intuitive way as a generalization of users; this approach makes the developer to ignore end-users except during analysis phase. Ignoring the user may result in systems that are not easy-to-use, as well as economic losses and high maintenance costs. This research proposes a way to evaluate software, enriched through the use of quality models, such as the proposed by McCall as well as the ISO/IEC 9126. This evaluation is made primarily as an assessment by end-users, since they made the final decision to accept or reject the system. This proposal intends to evaluate whether our approach can improve software development process using such models so that a user satisfaction can be improved.

Keywords: User-Centered Development, Usability, Quality, Quality Factors, Triangle McCall, ISO/IEC 9126.

1 Introducción

El software se ha vuelto un elemento indispensable para toda organización, un término común y entendido aun para aquellos ajenos al área. En la actualidad, es común encontrar sistemas informáticos en casi cualquier lugar, robustos sistemas bancarios, sistemas contables, y administrativos. Cada uno enfocado a un dominio en particular, haciendo de aquellos trabajos complejos actividades simples de realizar. Sin embargo, esto no es tan sencillo dado que un punto fundamental para la correcta aplicación de los sistemas informáticos estriba en el grado de eficiencia en el cual, el usuario puede interactuar y completar las tareas que desea de forma satisfactoria. Sin embargo, las prácticas de desarrollo de software en muchas ocasiones no contemplan a quiénes serán los responsables de operar el sistema, es decir, el usuario final. Todo esto provocando que los sistemas sean difíciles de manejar, y en ocasiones, colapsos totales debido a un software ineficiente o mal diseñado. Debido a todo esto, consideramos indispensable que el software sea sometido a una cuidadosa revisión por parte de los usuarios antes de su entrega definitiva. De esta manera, se

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 113-124.

buscaría cumplir con una meta primordial en el software: Usuarios satisfechos en sus expectativas [1].

Para los desarrolladores del software es importante contar con modelos que aseguren la calidad, afirmando de manera eficiente aquello que los usuarios necesitan. Dichos modelos listan un conjunto de atributos, los cuales todo producto de software con alta calidad debe poseer (ej. confiabilidad, mantenibilidad, entre otros) [2].

2 Conceptos Clave

2.1 Calidad del software

Existen múltiples definiciones de calidad del software. Algunas de ellas son: ISO/DIS 9126 [3]:

"La totalidad de características de un producto de software que conlleva en su habilidad de satisfacer necesidades existentes o implícitas".

Dentro de la literatura computacional existen generalmente dos significados aceptados para el término calidad. El primero es que la calidad significa "conocer los requerimientos". Con esta definición, para tener un producto de calidad, éstos deben ser medibles, y pueden ser conocidos o no. Dentro de este enunciado, la calidad es un estado binario, esto quiere decir que un producto posee calidad o no.

Los requerimientos pueden ser complejos o simples, siempre y cuando sean medibles, puede determinarse en donde los requerimientos de calidad se han cumplido. Esta es la vista de calidad del productor, a través del cumplimiento de las especificaciones y requerimientos. Cumplir con los requerimientos se vuelve un fin en sí mismo [4].

2.2 Modelos de Calidad

Los modelos de calidad de software, se definen de la siguiente manera [6]:

"La examinación sistemática de la capacidad del software que llene requerimientos específicos de calidad". Así mismo, un modelo de calidad es una buena herramienta para evaluar la calidad de un producto de software. Firesmith [7] lo define de una manera más concisa:

"Es un modelo jerárquico (colección por capas de abstracciones relacionadas o simplificadas) para formalizar el concepto de calidad de un sistema en términos de:"

Factores de calidad.- Atributos de un sistema que capturan los mejores aspectos de su calidad (usabilidad, interoperabilidad, seguridad, confianza, etc).

Subfactores de Calidad.- Componente mayor de un factor de calidad que captura un aspecto subordinado de la calidad de un sistema (respuesta, tiempo, etc.).

Criterios de Calidad.- Descripción específica de un sistema que proporciona evidencia a favor o en contra de la existencia de un factor específico de calidad.

Medidas de Calidad.- Medidas que califican los atributos de calidad y los hacen medibles, objetivos e inambiguos (Precisión = número de resultados adecuados / número total de resultados regresados).

2.2.1 Modelo McCall

Uno de los primeros modelos de calidad fue propuesto por McCall Richards, y Walters [8]. Este modelo describe la calidad como nacida de una relación jerárquica entre los factores de calidad, criterios de calidad, y métricas de calidad. McCall propone una clasificación útil de los factores que afectan la calidad los cuales se concentran en tres puntos clave (ver Figura 1).

Fig. 1. Triángulo de McCall (adaptado de [9].)

El modelo establece tres áreas principales que intervienen en la calidad del software [10]:

- Revisión de la calidad del producto de software. Tiene como objetivo realizar revisiones durante el proceso de desarrollo para detectar los errores que afecten a la operación del producto.
- Transición del producto. Requiere de la definición de estándares y procedimientos que sirvan como base para el desarrollo del software.
- Calidad en la operación del producto. Requiere que el software pueda ser entendido fácilmente, que opere eficientemente y que los resultados obtenidos sean los requeridos inicialmente por el usuario.

Pese a que el modelo de calidad de McCall fue concebido a finales de los años 70s, sigue siendo en la actualidad un marco de trabajo reconocido para el establecimiento de estándares de calidad en materia de software.

2.2.2. ISO/IEC 9126

EL ISO/IEC es un estándar para la evaluación de software y uno de los más conocidos. Propone modelos como los artefactos que dan seguimiento a la aseguración de los factores de calidad que son de interés particular [11]. Típicamente la calidad interna del software se obtiene por medio de la revisión de documentos de especificación, modelos de revisión, o por análisis de código fuente. La calidad externa se refiere a las cualidades del software que interactúan con su entorno. En contraste la calidad en uso se refiere a la calidad percibida por un usuario final que utiliza el

software bajo un contexto específico [12]. De tal manera las métricas internas pueden utilizarse para predecir la calidad del producto final. Para modelar la calidad interna y externa ISO/IEC 9126 define el mismo marco. El estándar está basado en seis características: Funcionalidad, Confiabilidad, Usabilidad, Eficiencia, Mantenibilidad, y Portabilidad (ver Tabla 2).

Table 1. Caracteristicas ISO/ISEC 9126-1[14].

CARACTERISTICAS	SUB-CARACTERISTICAS	EXPLICACIÓN
Funcionalidad	Conveniencia	¿ Realiza las tareas deseadas?
	Exactitud	¿Es el resultado esperado?
	Interoperabilidad	¿Puede interactuar con otros sistemas?
	Seguridad	¿Prevé acceso unautorizado?
Confiabilidad	Madurez	¿La mayoría de los defectos se eliminaron en tiempo?
	Tolerancia a Fallas	¿puede manejar errores?
	Recuperabilidad	¿Puede continuar con el trabajo y restablecer los datos después de una falla?
Usabilidad	Entendimiento	¿el usuario sabe cómo utilizar el sistema sin complicaciones?
	Facilidad de Aprendizaje	¿el usuario aprende fácilmente a utilizar el software?
	Operabilidad	¿El usuario utiliza el sistema sin esfuerzo?
	Atractivo	¿La interface luce bien?
Eficiencia	Tiempo de Respuesta	¿Qué tan rápido responde el sistema?
	Utilización de recursos	¿Utiliza los recursos eficientemente?
Mantenibilidad	Análisis	¿Las fallas pueden ser diagnosticadas?
	Cambio	¿Puede ser fácilmente modificado?
	Estabilidad	¿Continúa funcionando si algún cambio es realizado?
	Pruebas	¿Es fácil de probar?
Portabilidad	Adaptabilidad	¿Puede moverse a otros entornos?
	Instalación	¿Es fácil de instalar?
	Conformidad	¿Cumple con estándares de portabilidad?
	Remplazabilidad	¿Pude reemplazar a otro software?

Abran et al [13] argumentan que “Aunque se pensó, no es exhaustivo, estas series constituyen el modelo de calidad de software más extenso creado hasta la fecha”, es también un modelo simple de utilizar para aquellos no considerados como especialistas. Éste estándar incluye la visión del usuario e introduce el concepto de calidad en uso. La presente investigación se centra en el desarrollo de software con calidad, y por tanto, la importancia del usuario toma un papel importantísimo para la elaboración de software exitoso, por ello el modelo de calidad ISO/IEC 9126 es base para esta investigación.

3 Problemática

En la actualidad la mayoría de proyectos de software realizan tareas para implementar planes de calidad. Esto genera una forma ordenada y controlada de llevar a cabo el proceso de desarrollo de software [1]. Sin embargo, esto no determina la calidad en el producto final, debido a que tal afirmación depende del usuario final. Los sistemas de software se construyen de una manera intuitiva pensando en una generalización para el usuario, lo cual generalmente provoca que el desarrollador tienda a dejarlo de lado durante el análisis. No en pocas ocasiones el desarrollador realmente no conoce al usuario fanal y/o ha tenido contacto con ellos. Razones por las cuales, hacen una serie de suposiciones, las cuales no atienden a las necesidades reales y se ven reflejadas en el producto final. Existen aspectos propios del usuario tales como estilo de aprendizaje, herramientas favoritas, e incluso diferencias culturales, los cuales pueden impactar en la eficiencia del sistema en cuestión [15].El resultado pueden ser sistemas de información difíciles de utilizar por los usuarios. Algunos de ellos contienen textos técnicos que el usuario desconoce, en otros simplemente no se lograrán las metas que deseaba realizar al utilizar dicho sistema. Algunos problemas que acarrea el no realizar un diseño centrado en el usuario pueden ser los siguientes [15]:

- Pérdida significativa de tiempo en desarrollo, incluyendo en el sistema apartado que el usuario nunca utilizará porqué desconoce que existen o simplemente no sabe cómo hacer uso de ellos.
- Altos costos en diseño y mantenimiento ya que software mal construido requerirá futuras modificaciones que significan costos extra.
- Pérdidas económicas, debido a que los cambios efectuados en los sistemas para corregir errores sobrepasan los presupuestos establecidos al inicio.

Por lo anterior se puede decir que la calidad de un producto de software no depende únicamente del seguimiento en los procesos de desarrollo sino en el grado de satisfacción del usuario al utilizar dicho producto.

De acuerdo a Pfleeger [16] un sistema puede tener un desempeño excelente al realizar una función, pero si los usuarios no logran entender cómo utilizarlo, el sistema es un fracaso. De lo anterior reafirmarnos la importancia del aseguramiento de la calidad, de acuerdo a Lewis [4] De tal definición se puede decir que tales actividades y funciones de seguimiento no están especificadas, es decir varían según el tipo de proyecto, y personal que da seguimiento, sin embargo es importante destacar que en este seguimiento de calidad el usuario debe ser indispensable, propuesta que establece este trabajo de investigación, ya que la falta de calidad en el software (que determina el usuario final) acarrea varios inconvenientes.

Algunos ejemplos que la poca calidad en el software puede ocasionar son los siguientes [4]:

- Fallas frecuentes en el software.
- Las consecuencias de las fallas de software son inaceptables, desde lo económico a escenarios de vida.

- Frecuentemente el software no está disponible para lo que era su propósito original.
- Las mejoras del software suelen ser muy costosas.
- El costo de detectar y corregir errores es excesivo.

Por ello se han definido modelos de calidad que sirven como referencia para conocer las cualidades necesarias para alcanzar la calidad de un producto de software. Aunque los modelos de calidad definen criterios que evalúan la calidad de cierta forma, es necesario establecer métodos de evaluación cuantitativa que permitan obtener resultados de manera más objetiva, integrando tanto elementos técnicos a evaluar como involucrando al usuario en la determinación de la calidad del software [1].

Finalmente, desarrollar sistemas no es tarea simple ni barata. Asimismo, se ha reportado que el mantenimiento representa aproximadamente el 80% del costo del software [9], [18]. En consecuencia, es de vital importancia minimizar estos costos al detectar todo los posibles fallos que el sistema pueda tener. Además, Pressman [9] explica que generalmente los desarrolladores solo hacen pruebas "suaves" por lo que pueden tender a no ser exhaustivas. Creemos que el usuario final opera el sistema en condiciones normales –o extremas en no pocas ocasiones- por lo que la evaluación por ellos será más real y valiosa para obtener un producto final con calidad suficiente.

4 Desarrollo de la Investigación

Para llevar a cabo la presente investigación primero fue necesario definir un modelo de desarrollo de software base (ver Figura 2), para la realización del método desarrollado, dentro del cual se hace énfasis en el apartado *Pruebas de Calidad* (usuario/desarrollador) con el objetivo que al ser parte integral del proceso de desarrollo, la evaluación iterada por parte de los usuarios así como desarrollador permita alcanzar un producto final exitoso.

Fig. 2. Modelo de desarrollo base (adaptado de [17]).

Posteriormente en base a los modelos de calidad analizados (McCall e ISO 9126), se estableció que factores de calidad (ver Figura 3) deben evaluarse desde el punto de vista del usuario final, y cuáles del lado del desarrollador, así como la forma de medirlos y en qué términos.

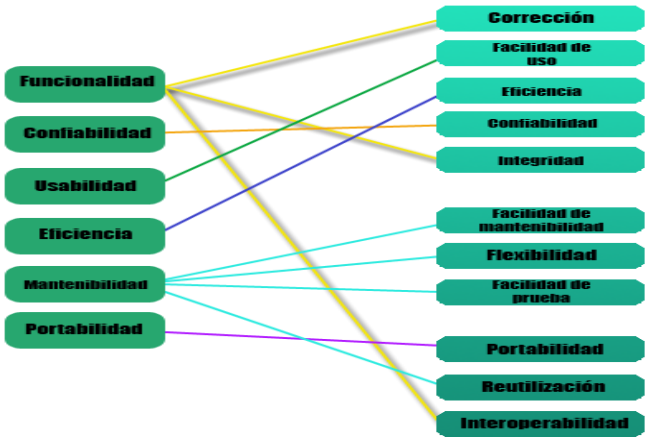


Fig. 3. Equivalencias McCall e ISO/IEC 9126 [1].

Tabla 2. Factores de calidad que se miden en la investigación

ASPECTOS	ATRIBUTO	SUB ATRIBUTO	COMO SE MIDE
Operación del Producto	Confiabilidad	Madurez	USUARIO
		Tolerancia a fallas	
		Recuperabilidad	
	Usabilidad	Comprensibilidad	USUARIO
		Facilidad de aprendizaje	
		Operabilidad	
		atractivo	
	Eficiencia	Tiempo de respuesta	USUARIO
	Funcionalidad	Conformidad	USUARIO
		Exactitud	
		Seguridad (No Aplica)	
Transición del Producto	Portabilidad	Adaptabilidad	USUARIO
		Conformidad	DESARROLLADOR
	Mantenibilidad	Facilidad de análisis	DESARROLLADOR
		Facilidad de cambio	
Revisión del Producto	Mantenibilidad	Facilidad de prueba	DESARROLLADOR
		Estabilidad	

4.1 Prueba Piloto

Una vez definidos los criterios, es necesario diseñar un instrumento de evaluación para medir los factores por parte del Usuario Final, para tal efecto se creó un cuestionario, que consiste de 7 preguntas demográficas y 34 preguntas sobre los factores de calidad deseados, utilizando la escala de Likert. Se utilizaron como base cuestionarios tales como: Questionnaire for User Interface Satisfaction (QUIS) [19], Computer System Usability Questionnaire (CSUQ) [19], System Usability Scale (SUS) [20]. Una vez terminado se realizó una prueba piloto para evaluar la efectividad de la herramienta creada por medio de la prueba de alfa de Cronbach que arrojó los siguientes resultados:

Table 3. Resultados prueba piloto.

FACTOR	ALFA DE CRONBACH	MEDIA	VARIANZA
Operabilidad	0.887	2.44	1.33
Facilidad de Aprendizaje	0.847	2.70	1.34
Atractividad	0.842	2.19	1.08
Comprensibilidad	0.804	2.45	1.60
Funcionalidad	0.810	2.51	1.60
Exactitud	0.905	2.38	1.15
Madurez	0.878	3.20	1.80
Tolerancia a Fallos	0.843	3.28	1.82
Recuperabilidad	0.947	3.18	1.80
Tiempo de Respuesta	0.868	2.51	1.44

Por lo tanto, debido a que los valores de alfa de Cronbach están por arriba de 0.8 se puede decir que la herramienta posee una consistencia interna adecuada, además al observar el histograma de cada una de las preguntas se aprecia que se cuenta con una distribución normal.

4.2 Métricas para el lado del desarrollador

Para las pruebas por parte de los desarrolladores se definieron métricas para medir los factores necesarios, a continuación se muestra como fueron definidas (ver tabla 5).

Table 4. metricas para medir el sistema del lado del desarrollador.

ATRIBUTO	SUB ATRIBUTO	METRICA
Mantenimiento	Facilidad de Análisis	$\alpha = \frac{\sum n.v.}{T.p.}$ $\beta = \frac{\sum e.c.}{T.p.}$

		$\gamma = \frac{\sum c.c.}{T.p.}$ $\delta = \frac{\sum n.c.}{T.p.}$ $f.a. = \frac{\alpha + \beta + \gamma + \delta}{4}$ ***Donde:*** Promedio Nombres de variables y métodos. Promedio Estructura del Comentario. Promedio Calidad del Comentario. Promedio Número de comentarios. ***T.p.***= Total Programadores. ***f.a.*** = Facilidad de análisis.
	Facilidad de Cambio	El índice de madurez de software (IMS) : ***IMS* = [Mt – (Fa + Fc + Fd)]/ Mt** ***donde:*** ***Mt*** = número de módulos en la versión actual. ***Fc*** = número de módulos en la versión actual que se han cambiado. ***Fa*** = número de módulos en la versión actual que se han añadido. ***Fd*** = número de módulos de la versión anterior que se han borrado en la versión actual.
	F. de prueba	Complejidad Ciclomática: **CC = A - N + 2C** ***donde:*** ***CC*** = *La complejidad ciclomática* ***A*** = *El número de aristas del grafo* ***N*** = *El número de nodos en el grafo* ***C*** = *El número de componentes conectados* Si la función posee múltiples puntos de salida, entonces: ***CC = A - N + C + R*** R= número de declaraciones de salida.
	Estabilidad	Eficacia de la Eliminación de Defectos(EED): **EED = E / (E + D)** ***donde:*** ***E****= es el número de errores encontrados antes de la entrega del software* ***D****= es el número de defectos encontrados después de la entrega*
Portabilidad	Conformidad	***Cdev(req,e1)* =** ***Cdes(req) + Ccod(req,e1) + Ctd(req,e1) + Cdoc(req,e1).*** **Donde:** ***Cdev(req,e1)****= Costo desarrollar software* ***Cdes(req)****= costo análisis y diseño* ***Ccod(req,e1)****=costo codificación* ***Ctd(req,e1)****= costo pruebas* ***Cdoc(req,e1)****=* costo documentación ***Cdevp(req,e1) = Cdev(req,e1) + Cpa(req)*** ***Donde:*** ***Cdevp(req,e1)****= costo rediseño.* ***Cpa(req)****= diseño portable* ***Cport(su,e2) = Cmod(su,e2) + Cptd(req,e2) + Cpdoc(req,e2).*** **Donde:** ***Cport(su,e2)****=portabilidad del software en nuevos ambientes* ***Cmod(su,e2)****= modificación manual*

		Cptd(req,e2)= *pruebas* ***Cpdoc(req,e2)=***documentación ***DP(su) = 1 - (Cport(su,e2) / Crdev(req,e2)).*** **Donde:** ***DP(su)=****conformidad en la portabilidad*

4.3 Perfil del usuario

Para realizar el experimento se desarrolló un sistema web orientado al público en general, por esto debido a que es web las características del usuario serán acorde al perfil de los Internautas en México, las características relevantes para definir el perfil del usuario se toman en base los siguientes datos:

- **Edad**

La población de usuarios que hacen uso de Internet en México está conformada principalmente por Jóvenes, de acuerdo con cifras de INEGI [21] 77.3% de los Internautas Mexicanos tiene menos de 35 años; AMIPCI [22] establece que 61% de los jóvenes entre 20 y 24 años hacen uso del ciberespacio.

- **Escolaridad**

Conforme a un estudio realizado por INEGI [21] 7 de cada 10 jóvenes que cursan estudios de nivel superior hacen uso frecuente de la web

4.4 Muestra

Por lo anterior se realiza el estudio con estudiantes de licenciatura de la Universidad Autónoma de Aguascalientes, de la carrera de Ingeniería en Sistemas Computacionales, ya que al ser relacionadas al uso de tecnologías de la información hacen mayor uso de Internet, factor clave para la investigación. Por ello los grupos seleccionados de usuarios cumplen con las características antes mencionadas de forma adecuada.

4.5 Análisis de Datos

Con el fin de evaluar la calidad de cada aspecto y poder tomar decisiones, se considero como Excelente un valor promedio entre 1 y 1.5, 1.5 a 2 como muy buena, 2 a 2.5 mínima aceptable, y superior a 2.5 mala. [23]

En cada iteración se obtiene una evaluación adecuada del producto y en base a esto se realizan los ajustes necesarios al sistema que se evalúa. Ello permitirá mejorar en lo subsecuente hasta lograr el nivel indicado para su liberación. Para la evaluación de los factores se utiliza estadística descriptiva, y posteriormente un análisis de correlación entre los datos obtenidos.

5 Conclusiones

Aunque existen varios trabajos en nuestro país referentes a la calidad en los sistemas, pocos de ellos han intentando hacer una evaluación sistemática y donde el usuario final sea el principal crítico del producto de software entregado. Creemos que de esta forma se están atacado una serie de aspectos relevantes relacionados con la adopción de tecnología, tales como: reducir la resistencia al cambio, elevar la calidad, lo cual reduce costos de mantenimiento, entre otros. Adicionalmente, este trabajo se centra en la evaluación del producto, ya que otros se han centrado en la evaluación del proceso donde se sugiere usar Moprosoft, CMMI, entre otros. Esto no quiere decir que son estrategias rivales, más bien son estrategias complementarias para que al final los productos de software sean producidos bajo estándares y con la calidad esperada del usuario final.

Es importante recalcar que la calidad en el proceso de desarrollo de un producto de software no garantiza la calidad del producto final. Sommerville [18] argumenta que una suposición subyacente de la gestión de la calidad, es que la calidad del proceso de desarrollo afecta directamente a la calidad de los productos derivados.

Referencias

1 A. Castañeda, et al.,"Método de Evaluación de Software Para Mejorar la Calidad del Producto Final," Universidad Autónoma de Aguascalientes, Aguascalientes, Ags.2009.

2 W. A. Ward and B. Venkataraman. 1999, Some Observations on Software Quality. School of Computer and Information Sciences, University of South Alabama.

3 ISO/IEC, "Information technology – software product evaluation – quality characteristics and guidelines for their use," in Technical Report 9126, ed: International Organization for Standardization, 1990.

4 W. E. Lewis, Software Testing and Continuos Quality Improvement, 3 ed.: CRC Press, 2009.

5 R. A. Khan, et al., Software Quality Concepts and Practices: Alpha Science, 2006.

6 S. R. Kalaimagal Sivamuni. 2008, A Retrospective on Software Component Quality Models.

7 D. Firesmith. 2005. Achieving Quality Requirements with Reused Software Components: Challenges to Successful Reuse.

8 J. A. McCall, et al., "Factors in software quality," US Rome Air Development Centre Reports1977.

9 R. Pressman, Ingeniería de Software, 6ª ed.: McGraw Hill, 2005.

1 M. Calero y Coral, "Modelos de Calidad. WQM, PQM, e-Commerce, portlets," Departamento de Informática, Universidad de Castilla-La Mancha, 2005.

11 J. P. Carvallo y X. Franch. 2006, Extending the ISO/IEC 9126-1 Quality Model with Non-Technical Factors for COTS Components Selection.

12 B. Zeiss y D. Vega. 2007, Applying the ISO 9126 Quality Model to Test Specifications. Institute for Informatics.

13 A. Abran, et al., "Usability Meanings and Interpretations in ISO Standards," Software Quality Journal, 2003.

14 SO, "ISO 9126-1:2001 Software engineering-Product quality," in Part 1: Quality Model, ed: International Organization for standardization, 2001.

15 D. Macracken y R. J. Wolfe, User-Centered Website Development: Prentice Hall, 2004.

16 Pfleeger y S. Lawrence, Ingeniería de software: teoría y práctica, 1a ed., 2002.

17 J. M. Gomez-Reynoso y M. R. Brisuela-Sandoval, "Participación de los usuarios en sistemas de información," presented at the Americas Conference on Information System, Toronto, ON, Canada, 2008.
18 I. Sommerville, Ingenieria de Software, 7 ed.: Addison Wesley, 2006
19 T. S. Tullis and J. S. Stetson, "A Comparison of Questionnaires for Assessing Website Usability," 2004.
20 J. R. Lewis y J. Sauro. *The Factor Structure of the System Usability Scale.*
21 INEGI. 2009, ESTADÍSTICAS A PROPÓSITO DEL DÍA MUNDIAL DE INTERNET.
22 AMIPCI. 2009, ESTUDIO AMIPCI 2009 Sobre hábitos de los Usuarios de Internet en México.
23 J. M. Gomez-Reynoso, *et al.*, "Utilizando el Modelo de Calidad McCall y el Estándar ISO-9126 para la Evaluación de la Calidad de Sistemas por los Usuarios," presentado en Proceedings of the sixteenth Americas Conference of Informations Systems Lima, Peru, 2010.

Diseño de un Almacén de datos basado en Data Warehouse Engineering Process (DWEP) y HEFESTO

Castelán García Leopoldo, Ocharán Hernández Jorge Octavio

Facultad de Economía e Informática
Universidad Veracruzana
lcastelan@uv.mx, jocharan@uv.mx
Paper received on 02/07/10, Accepted on 04/10/10.

Resumen. El desarrollo de un almacén de datos se basa en el diseño de un modelo conceptual que incluye tanto los requisitos de información de los usuarios así como las fuentes de datos operacionales. A partir de éste se obtiene un modelo lógico basado en una tecnología de base de datos específica que guía la implementación. Actualmente muchas de las propuestas no definen mecanismos para estructurar de manera sistemática el desarrollo de un almacén de datos, convirtiéndolo en una tarea compleja y artesanal. Para dar solución a este problema, el trabajo expuesto propone el uso de la metodología HEFESTO para definir la arquitectura de datos y DWEP para facilitar el diseño y construcción de un almacén de datos. El propósito es combinar estas metodologías ayudando a reducir los problemas que resultan de seguir un proceso sin comprenderlo.

Palabras clave: Almacén de Datos, DWEP, PU, HEFESTO, UML.

1. Introducción

Almacén de Datos

Los Almacenes de Datos (AD) o *Data Warehouse* son una colección de datos históricos que sirven de apoyo a la toma de decisiones para mejorar un proceso de negocio. Bill Inmon lo define como [1]:

"…un almacén de datos es una colección de datos orientados a temas, integrados, no-volátiles y variables en el tiempo, organizados para soportar necesidades empresariales…".

Un AD se dice que es orientado a temas porque la información se clasifica en base los intereses de la organización. Es integrado cuando sus datos provienen de

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 125-135.

diversas fuentes, es decir, que son producidos por distintos departamentos y aplicaciones, tanto internas como externas. Estos datos deben ser consolidados en una instancia antes de ser agregados al AD. Este proceso se conoce como Extracción, Transformación y Carga de Datos (*Extraction, Transformation and Load* - ETL). La integración de datos resuelve diferentes tipos de problemas relacionados con las convenciones de nombres, unidades de medidas, codificaciones, fuentes de datos múltiples, entre otros. Son no-volátiles porque la información es útil para el análisis y la toma de decisiones solo cuando es estable. Los datos operacionales varían constantemente, en cambio, los datos una vez que entran en el AD no cambian, y por último se dice que son variables en el tiempo debido al gran volumen de información que se maneja cuando se realiza una consulta, los resultados deseados tardarán en originarse, este espacio de tiempo entre la búsqueda de datos y el resultado es del todo normal en este ambiente y es precisamente por ello, que la información que se encuentra dentro del AD se denomina de tiempo variable [2].

Un AD permite tener el manejo y control de la información, así se tiene asegurada una vista única de los datos de fuentes funcionalmente distintas (bases corporativas, bases propias, sistemas externos, etc.). Los usuarios finales no tienen la necesidad de aprender a utilizar diferentes sistemas de acceso y manipulación de los datos. Facilita la comprensión de los datos al transformarlos en información útil y ayuda a la organización a responder preguntas esenciales para la toma de decisiones que le permitan obtener ventajas competitivas y mejorar su posición en el mercado.

Importancia del almacén de datos

Para Inmon la importancia de un AD se define en tres dimensiones [1]:

1. Mejora la entrega de información: información que sea completa, correcta, consistente, oportuna y accesible.
2. Facilita el proceso de toma de decisiones: con un gran respaldo de información se puede tomar decisiones rápidamente; así también, la gente de negocios adquiere confianza en sus propias decisiones y logra un mayor entendimiento de los impactos de éstas.
3. Impacto positivo sobre los procesos empresariales: cuando la gente tiene acceso a una mejor calidad de información, la empresa mejora eliminando los retardos en los procesos que resultan de información incorrecta, inconsistente; logrando así una integración y optimización de procesos empresariales a través del uso compartido e integrado de las fuentes de información.

Problemática de los sistemas basados en almacén de datos

En distintas fuentes [5] [6] se observa que entre el 40% y 50% de los sistemas basados en AD fallan o son abandonados por aspectos que no se tienen considerados desde un principio. Según Larry Poole [7] las razones de por qué estos sistemas fallan son:

1. No cuentan con un líder que entienda el valor del proyecto y esté dispuesto a asignar los recursos apropiados.
2. Los requisitos son pobres, ya que no se involucran a los usuarios en las discusiones.
3. Los diseños son pobres debido a que los requisitos son deficientes y el tiempo de modelado es limitado.
4. No se tiene entrenamiento a usuarios finales para el uso adecuado de la solución, para llevar a buen término la implantación del proyecto.
5. Se cree que con la solución inicial se termina el proyecto descuidando su mantenimiento o crecimiento.
6. Se escogen inadecuadamente las herramientas a utilizar.
7. Muchos proyectos arrancan pensando en una solución final pero sin saber el tiempo y no se estima el trabajo que requiere, o si la solución es compleja.
8. La solución no cumple con los objetivos.

Arquitectura de un almacén de datos

La Arquitectura de un AD está formada por diversos elementos que interactúan entre sí y que cumplen una función específica dentro del sistema, como se muestra en la Figura 1, Los componentes de una AD según Bernabeu son [2]:

- OLTP (On Line Transaction Processing), representa toda aquella información transaccional que genera la organización diariamente y las fuentes externas.
- LOAD MANAGER. Los ETL se encargan de extraer los datos desde los OLTP para manipularlos, integrarlos, transformarlos y posteriormente cargar los resultados obtenidos en el AD, es necesario contar con un sistema que se encargue de ello.
- DW MANAGER. Su finalidad es transformar e integrar los datos fuentes y de almacenamiento intermedio en un modelo adecuado para la toma de decisiones. Permitiendo realizar todas las funciones de definición y manipulación del depósito de datos, para poder soportar todos los procesos de gestión del mismo.
- QUERY MANAGER. Este componente realiza las operaciones necesarias para soportar los procesos de gestión y ejecución de consultas relacionales, propias del análisis de datos, recibe las consultas del usuario, las aplica a la estructura de datos correspondiente y devuelve los resultados obtenidos.
- HERRAMIENTAS Y CONSULTAS DE DATOS. Son los sistemas que permiten al usuario realizar la exploración de datos del AD. Básicamente constituyen el nexo entre el depósito de datos y los usuarios.
- USUARIOS. Son aquellos que se encargan de tomar decisiones y de planificar las actividades del negocio.

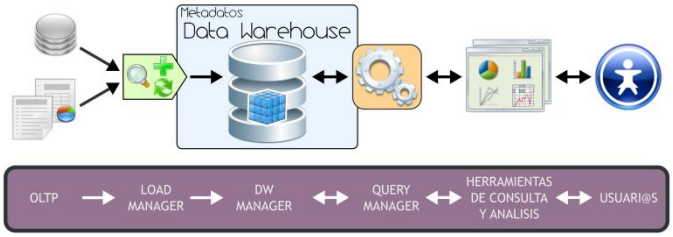


Figura 1. Arquitectura de un Almacén de Datos tomada de la metodología HEFESTO [2].

Esta arquitectura de AD opera de la siguiente manera:

Los datos se extraen de aplicaciones, bases de datos, archivos, entre otros. Esta información generalmente reside en diferentes tipos de sistemas, orígenes y arquitecturas con diferente formatos, los datos son integrados, transformados y limpiados, para luego ser cargados en el AD, la información se estructura en cubos multidimensionales para responder a consultas dinámicas con una buena presentación. Los usuarios acceden a los cubos, utilizando diversas herramientas de consulta, exploración, análisis, reportes, etc.

Un cubo multidimensional, representa o convierte los datos planos que se encuentran en filas y columnas, en una matriz de N dimensiones. Los objetos más importantes que se pueden incluir en un cubo multidimensional son los indicadores, atributos y jerarquías, de esta manera los atributos existen a lo largo de varios ejes o dimensiones, y la intersección representa el valor que tomará el indicador que se está evaluando.

2 Propuesta

Motivación

Los problemas más frecuentes de los sistemas basados en AD se encuentran en la recolección de requerimientos, el análisis y el diseño [3], debido a que no sigue una metodología estándar para su desarrollo.

La metodología denominada proceso de ingeniería para el desarrollo de almacenes de datos (DWEP) [4] la cual está basada en el proceso unificado (PU), abarca los flujos de trabajo de requerimientos, análisis, diseño, pruebas, mantenimiento y revisiones posteriores al desarrollo; sus principales características son que es iterativa, dirigida por casos de uso y se basa en las etapas de desarrollo de PU, utilizando UML como lenguaje para modelado gráfico.

Por otro lado, en el componente del proceso de arquitectura de datos se tiene a HEFESTO [2] una metodología cuya propuesta se fundamenta en una amplia investigación, comparación de metodologías existentes y la experiencia en la elaboración de almacenes de datos. La ventaja principal de esta metodología es que específica puntualmente los pasos a seguir en cada fase a diferencia de otras metodologías que mencionan los procesos, más no explican cómo realizarlos.

Objetivo

Crear un proceso de desarrollo para el diseño de un AD basado en la integración de la metodología proceso de ingeniería para el desarrollo de almacenes de datos (DWEP) y la metodología HEFESTO, partirá de la recolección de requerimientos y necesidades de información del usuario, y concluirá en la elaboración de un modelo conceptual, lógico y físico con la finalidad de poder facilitar el trabajo que implica la elaboración de un AD desde su inicio.

Fases del DWEP y proceso unificado

El PU como se muestra en la Figura 2, es un marco de desarrollo compuesto de cuatro fases, cada una de ellas a su vez dividida en una serie de iteraciones que ofrecen como resultado un incremento del producto desarrollado, que añade o mejora las funcionalidades del sistema en desarrollo [8].

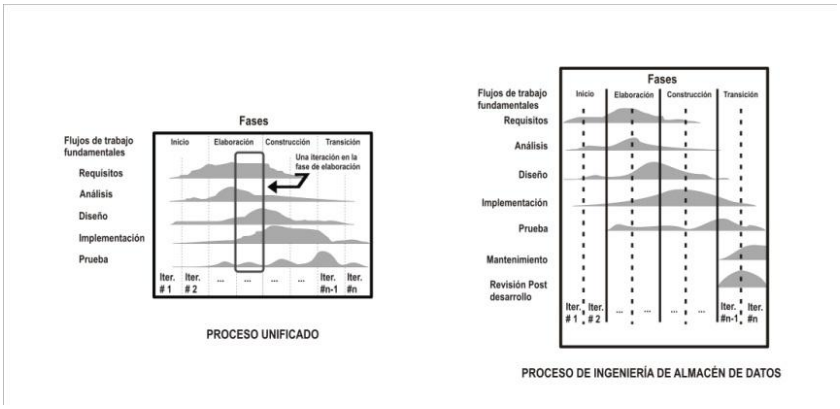


Figura 2. El proceso unificado [8] y proceso de ingeniería de almacenes de datos [4]

***Fase de inicio*:** El objetivo de esta fase es analizar el proyecto para justificar su puesta en marcha, para lograrlo se realiza una descripción general del proyecto, se detectan los riesgos críticos y se establecen la funcionalidad básica del software con una descripción de la arquitectura candidata.

***Fase de elaboración*:** Una vez finalizada la fase de inicio, se pretende formar una arquitectura sólida para la construcción del software. En esta fase se busca establecer la base lógica de la aplicación con los casos de uso definitivos y los artefactos del sistema que lo componen.
***Fase de construcción*:** Se inicia a partir de la línea base de arquitectura que se especificó en la fase de elaboración y su finalidad es desarrollar un producto listo para la operación inicial en el entorno del usuario final.
***Fase de transición*:** Una vez que el proyecto entra en la fase de transición, el sistema ha alcanzado la capacidad operativa inicial. Esta fase busca implantar el producto en su entorno de operación.

Flujos de trabajo aplicados al proceso DWEP

En términos generales para el PU y el DWEP un flujo de trabajo es un conjunto de actividades realizadas en un área determinada cuyo resultado es la construcción de artefactos (un texto, un diagrama, una página Web, código en lenguaje de programación, etc.).

Requerimientos. Durante este flujo de trabajo, los usuarios especifican las medidas y agregaciones más interesantes, el análisis dimensional, consultas usadas para la generación de reportes periódicos y frecuencia de la actualización de los datos. El PU sugiere el uso de casos de uso [9] [4]. Esto ayuda a comprender el sistema y obtener los requisitos y funciones para la solución. Además establece como deben ser las interacciones del sistema.
Análisis. Tiene como objetivo mejorar la estructura y los requisitos obtenidos en la etapa de requerimientos. En esta etapa se documentan los sistemas operacionales preexistentes que alimentaran el AD. El PU propone el uso del diagrama de clase [4] [9].
Diseño. Al final de este flujo de trabajo, está definida la estructura del AD. El principal resultado de este flujo de trabajo es el modelo conceptual del AD.
Implementación. Durante este flujo de trabajo, el AD es construido y se empiezan a recibir datos de los sistemas operaciones, se afina para un funcionamiento optimizado, entre otras tareas.
Pruebas. El objetivo de este flujo de trabajo es verificar que la aplicación funcione correctamente, realizar las pruebas y analizando los resultados de cada prueba.
Mantenimiento. **U**n AD es un sistema que se retroalimenta constantemente. El objetivo de este flujo de trabajo es definir la actualización y carga de los procesos necesarios para mantener el AD.
Revisiones post desarrollo. Esto no es un flujo de trabajo de las actividades de desarrollo, sino un proceso de revisión para la mejora de proyectos a futuro. Si hacemos un seguimiento del tiempo y esfuerzo invertido en cada fase es útil en la estimación de tiempo y de las necesidades para generar los requisitos para desarrollos futuros.

Etapas de la metodología HEFESTO

El objetivo de la metodología HEFESTO [2] es que de manera metódica y sencilla podamos comprender cada paso que se ejecuta para entender el por qué del proceso y no caer en el error de seguir un método sin saber que se está haciendo, guiándose por pasos lógicos relacionados durante todas las etapas del proceso. La metodología HEFESTO, puede ser utilizada en cualquier ciclo de vida que no requiera fases extensas de requerimientos y análisis, con el fin de entregar una implementación que cumpla con una parte de las necesidades proporcionadas por el usuario.

Descripción

La metodología HEFESTO inicia con la recolección de las necesidades de información de los usuarios obteniendo así las preguntas claves del negocio. Luego, se deben identificar los indicadores resultantes y sus perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del AD, se analizarán las base de datos de los OLTP o fuentes externas para señalar las correspondencias con los datos fuentes y seleccionar los campos de estudio de cada perspectiva. Una vez hecho esto, se pasará a la construcción del modelo lógico, tomando en cuenta las jerarquías que intervendrán. Por último, se definirán los procesos de carga, transformación, extracción y limpieza de los datos fuente.

A continuación podemos ver cada uno de los pasos correspondientes a la metodología HEFESTO y que resumen la descripción antes mencionada, sobre los procesos que se siguen.

1. Análisis de Requerimientos
    a. Identificar preguntas
    b. Identificar indicadores y perspectivas de análisis
    c. Modelo conceptual
2. Análisis de los OLTP
    a. Determinación de Indicadores
    b. Establecer correspondencia
    c. Nivel de granularidad
    d. Modelo conceptual ampliado
3. Modelo lógico del Almacén de Datos
    a. Tipo del modelo lógico del Almacén de Datos
    b. Tabla de dimensiones
    c. Tabla de hechos
    d. Uniones

4. Procesos ETL

Pasos y aplicación metodológica

Análisis de Requerimientos. Se identifican los requerimientos del usuario con el fin de entender los objetivos de la organización, haciendo uso de técnicas y herramientas, como la entrevista, la encuesta, el cuestionario, la observación, el diagrama de flujo y el diccionario de datos, obteniendo como resultado una serie de preguntas que se deberán analizar con el fin de establecer cuáles serán los indicadores y perspectivas que serán tomadas en cuenta para la construcción del AD. Finalmente se realizará un modelo conceptual en donde se podrá visualizar el resultado obtenido en este primer paso.
Análisis de los OLTP. Tomando en cuenta el resultado obtenido en el paso anterior se analizarán las fuentes OLTP para determinar cómo serán calculados los indicadores con el objetivo de establecer las respectivas correspondencias entre el modelo conceptual y las fuentes de datos. Luego, se definirán qué campos se incluirán en cada perspectiva y finalmente, se ampliará el modelo conceptual con la información obtenida en este paso.
Modelo lógico del Almacén de Datos. Como tercer paso, se realizará el modelo lógico de la estructura del AD, teniendo como base el modelo conceptual. Para esto, debemos definir el tipo de representación de un AD que será utilizado, posteriormente se llevarán a cabo las acciones propias al proceso, para diseñar las tablas de dimensiones y de hechos. Por último, se realizarán las uniones pertinentes entre estas tablas.
Procesos ETL. El último paso de la metodología HEFESTO es probar los datos, a través de procesos ETL. Para realizar la compleja actividad de extraer datos de diferentes fuentes, para luego integrarlos, filtrarlos y depurarlos, se podrá hacer uso de software que facilita dichas tareas, por lo cual este paso se centrará solo en la generación de las sentencias SQL que contendrán los datos que serán de interés.

Antes de realizar la carga de datos, es conveniente efectuar una limpieza de los mismos, para evitar valores faltantes y anómalos. Al generar los ETL, se debe tener en cuenta cuál es la información que se desea almacenar en el AD, para ello se pueden establecer condiciones adicionales y restricciones. Estas condiciones deben ser analizadas y realizadas con mucha prudencia para evitar pérdidas de datos importantes.

Modelado de proceso

Para aplicar el proceso de desarrollo propuesto en este trabajo se definen tres niveles de abstracción: conceptual, lógico y físico. El nivel conceptual representa las interacciones entre las entidades y las relaciones, el nivel lógico describe con tanto detalle como sea posible los datos, sin tener en cuenta cómo estén físicamente en la

base de datos y el nivel físico incluye la especificación técnica, después de la reglas del negocio para determinar el diseño del AD.

Esta propuesta se encuentra en la definición del nivel conceptual y en una etapa posterior la definición de los demás niveles. El objetivo más importante de este nivel es la representación de las principales propiedades sin tener en cuenta detalles específicos de tecnología alguna de bases de datos, posibilitando así la independencia del modelo respecto de la plataforma en la cuál sea implementado. Una vez que los requisitos de información se han especificado, debe derivarse un modelo conceptual inicial.

En esta etapa inicial de desarrollo (conceptual) se crean las bases del proyecto, definiendo el alcance, plan inicial, la visión del negocio junto con las metas y la justificación del proyecto. Los requerimientos iniciales son capturados a través de casos de uso.

Se define un equipo de trabajo para el plan de proyecto que maneja las dependencias principales y la estrategia general, mientras que los planes más refinados manejan las tácticas propias de las situaciones en cuestión de cada iteración.

Este nivel finaliza con la existencia de un alcance definido, plan de desarrollo, análisis de riesgos donde los clientes concuerden con lo anterior.

En la Figura 3 se representa de forma general el nivel conceptual.

Los pasos para el desarrollo de un AD son los siguientes:

1. Se debe iniciar con la definición del alcance, para evitar grandes problemas en sus fases, se recomienda usar la técnica de descomposición de tareas, y para su representación normalmente se usa la forma tipo organigrama.
2. Se debe redactar un plan inicial que contempla la visión del negocio junto con las metas y la justificación del proyecto.
3. Se deberán definir roles dependiendo de las actividades de cada usuario.
4. Se debe realizar un documento que describa las características de la empresa que contemple datos que la identifiquen, objetivos, políticas, estrategias, definición de los procesos y la relación de las metas con el AD.
5. Como siguiente paso de realiza la identificación de preguntas mediante cuestionarios, entrevistas u observaciones sobre los objetivos del proyecto, con la finalidad de identificar los indicadores y perspectivas de las cuales partirá el análisis de diseño.
6. Una vez que se han establecido las preguntas claves, se debe proceder a su descomposición para descubrir los indicadores que representan lo que se desea analizar concretamente y las perspectivas que se refieren a los objetos mediante los cuales se quiere examinar los indicadores, realizando un detalle de estos.
7. A partir de los indicadores y perspectivas obtenidas en el paso anterior se debe construir un modelo conceptual.
8. Se realiza la representación de los casos de uso que representan los requerimientos del AD.
9. La fuente de datos del AD es el modelo entidad-relación representado en diagramas UML. Al transformar el modelo ER a un diagrama de clases se transforma en el esquema conceptual de la fuente de datos (SCS). En el di-

seño conceptual del AD según DWEP se realiza los diagramas: conceptual de datos (SCS), mapeo de datos (DMS), almacén de datos (DWCS) y el esquema conceptual del cliente (CCS. El esquema conceptual de la base de datos se realiza en tres niveles que consisten en la realizar un detalle de cómo se encuentra integrada.

10. Este nivel finaliza con la redacción de un acta de aprobación del cliente respecto a la definición de indicadores, fuente de datos y el modelo conceptual.

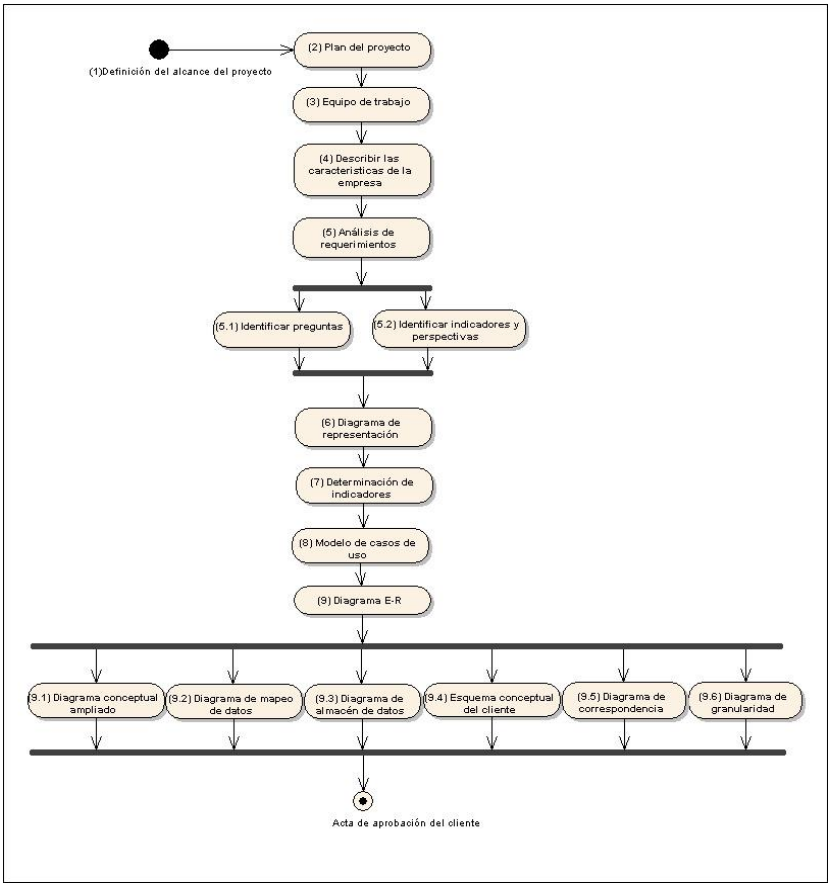


Figura 3. Esquema propuesto de nivel conceptual

Conclusiones

Este trabajo presenta una aproximación de modelo de desarrollo que se realiza, mediante la integración de HEFESTO Y DWEP. Este modelo de desarrollo consta de tres niveles en los cuales se integrada cada una de las actividades a realizar y los

artefactos por cada nivel. La propuesta se centra en los pasos del nivel conceptual debido al tiempo que requiere para el análisis de integración de los artefactos generados por ambas metodologías para la definición del modelo lógico y físico que no ha sido posible definir de una manera clara.

El uso de este modelo posibilita que el desarrollo de un AD esté bien estructurado, permitiendo considerar varios aspectos muy importantes en el desarrollo de este tipo de sistemas, como son la recopilación de requisitos de información y las OLTP, los procesos ETL, el AD como estructura física y la explotación de la información.

La necesidad que existe en las empresas por compartir y obtener un conocimiento sobre el mundo de información que guardan para poder realizar acciones para la toma de decisiones, hace posible el desarrollo de AD que sean guiados por procesos bien definidos y sustentados en metodologías que hacen uso de estándares como UML para el modelado, que garantiza que el diseño de cada elemento sea definido de forma única y poder ofrecer sistemas de calidad basados en una ingeniería de software que esté sustentada en la elaboración de artefactos y seguimiento de procesos que sean claros que permiten entender el por qué realizar dicha actividad al momento de su diseño.

Este modelo de desarrollo será propuesto para la elaboración del sistema de indicadores de la Universidad Veracruzana, con la idea de evaluar cada uno de los niveles considerando los artefactos que son necesarios; conocer los resultados obtenidos para validar si es un proceso de ayuda para el cumplimiento de los objetivos o si es necesario un nuevo planteamiento en su desarrollo.

Referencias

1. Inmon, W. Building the data warehouse. s.l. : Wiley & Sons, 2002.
2. Dario, Bernabeu Ricardo. DATA WAREHOUSING:Investigación y Sistematización de Conceptos. Córdoba, Argentina : Licencia de Documentación Libre de GNU, 2009.
3. B. Husemann, J. Lechtenborger, G. Vossen. Conceptual Data Warehouse Desing, Proceeding of the International Workshop on Design and Management of Data Warehouses. Sweden : StockHolm, 2000.
4. Lujan, S. Data WareHouse Desig with UML, PHD. Thesis Universidad de Alicante. Alicante, España : s.n., 2005.
5. consortiwn, Custer. 41% HAVE EXPERIENCED DATA WAREHOUSE PROJECT FAILURES. [En línea] http://www.cutter.com/research/2003/edge030218.html.
6. Madsen, Mark. A 50% Data Warehouse Failure Rate is Nothing New. [En línea] Mark Madsen. http://it.toolbox.com/blogs/bounded-rationality/a-50-data-warehouse-failure-rate-is-nothing-new-4669.
7. Poole, Larry. 8 Reasons Why Business Intelligence Initiatives Fail. [En línea] www.xyber.net/8Reasons.doc.
8. (OMG)., Object Management Group. Unifie Modeling Language (UML),version 2.0. [En línea] http://www.uml.org/.
9. Mora, J. Trujillo and L. Data WareHouse Desig with UML, PHD. Thesis Universidad de Alicante. Alicante, España : s.n., 2005.
10. Jacobson, Ivar, Booch, Grady y Rumbaugh, James. El proceso unificado de desarrollo de software. Madrid : Addison Wesley, 2000.

Valoración de la Respiración Basada en Modelos GMM

Pedro Mayorga *, Christopher Druzgalski**, O. Hugo González Arriaga ***, and Raúl Ludwig Morelos***

* Instituto Tecnológico de Mexicali, Mexicali, B.C., México.
Electrical/Biomedical Engineering, CSULB, CA, USA
* Electrical/Biomedical Engineering, CSULB, CA, USA
*** Instituto Tecnológico de Mexicali.
Paper received on 29/03/10, Accepted on 25/09/10.

Abstract. Mexicali se ubica en una región árida, experimentando enfermedades respiratorias, asma y alergias, que son debidos principalmente a partículas PM_{10} que afectan especialmente a niños. Este artículo presenta un método para una valoración cuantitativa en la salud de pacientes relacionados con desordenes respiratorios utilizando sonidos del pulmón. Aquí, aplicamos tecnologías tradicionales en el dominio del procesamiento de voz, utilizando señales del pulmón obtenidas con un estetoscopio digital. Técnicas tradicionales para el estudio de enfermedades respiratorias y asma, involucran auscultaciones y espirometría, pero métodos más confiables como el estetoscopio electrónico están ya disponibles; asimismo, métodos cuantitativos de análisis de señales ofrecen mejorar estos diagnósticos. Particularmente, proponemos una metodóloga de evaluación acústica apoyada en los Modelos Mezclados Gaussianos (GMM) la cual resultar en un amplio análisis, identificación y diagnostico de asma, basándose en el análisis en frecuencia de sonidos pulmonares.

Keywords: Asma, Crepitaciones, Sibilancias, Estetoscopio, Modelos Mezclados Gaussianos (GMM)

1. Introducción

Mexicali, México destaca por sus en niveles altos de partículas PM_{10} [1-7]. Una de las enfermedades respiratorias asociadas con niveles altos PM_{10} es el asma [7], la cual se presenta como una obstrucción de las vías respiratorias, tensión pectoral, tos y sibilancias al respirar [1]. La auscultación tradicional tiene muchas limitaciones siendo un proceso subjetivo que depende del oído individual, la experiencia y habilidad para distinguir diferentes patrones acústicos. Además, el estetoscopio tiene una respuesta en frecuencia que atenúa componentes acústicas pulmonares, donde el oído humano no es muy sensible [2-5].

Durante una auscultación, es esencial tener un buen historial clínico mediante un cuestionario efectuado a los padres del niño [6-14]. En este contexto, es vital una

©E.Cuatecontzi Cuahutle, A.Cortés Fernández, J. F. Ramírez Cruz, J. H. Sossa Azuela. (Eds.). Advances in Intelligent and Information Technologies. Research in Computing Science 50, 2010, pp.139-148.

buena valoración de sonidos adventicios por parte de los padres o por el médico que atiende al paciente. También, es primordial que la familia reconozca el significado de sonidos adventicios, ya que frecuentemente los padres usan una misma palabra para describir distintos sonidos. En práctica clínica, la auscultación depende de la experiencia y aptitudes sensoriales del médico. La literatura asocia las sibilancias con el asma y enfermedades pulmonares obstructivas crónicas [2-14]. Un estudio demuestra que los padres no suelen identificar bien las sibilancias [3-12]. Las herramientas digitales son mejores al identificar niños con sonidos adventicios de aparición temprana, a fin de dar un tratamiento adecuado [4-25], y fortalecer el diagnostico medico. Estudios realizados al oído humano muestran deficiencias en la auscultación, dependiendo de la edad del médico y de su entrenamiento [25]. De aquí, que técnicas como reconocimiento de patrones y procesamiento digital prometen ser más precisas y ayudan a consolidar los diagnósticos médicos. La comunidad Europea ha financiado diversos proyectos concertados como el intitulado *Computerized Respiratory Sound Analysis* (CORSA), cuyo objetivo principal fue establecer protocolos para la investigación y practica clínica en el campo de los sonidos respiratorios [4].

2 Técnicas Acústicas para la Valoración de Asma

Algunos expertos en asma argumentan que los sonidos adventicios no son un elemento suficiente para diagnosticar asma, pero estos siguen siendo fundamentales al diagnosticar patologías respiratorias. La espirometría es una serie de pruebas respiratorias controladas para medir capacidad y volumen pulmonar. Durante la valoración se pide al paciente que efectué varios soplidos de práctica, adiestrándolo para que realice una inspiración máxima, seguida por una espiración forzada rápida, repitiéndolo hasta completar tres ensayos [1]. El estetoscopio es útil para detectar sonidos distintivos; un estetoscopio electrónico convierte ondas acústicas a señales eléctricas, permitiendo escucharlas óptimamente. En otra técnica, se coloca un micrófono en la pieza para el pecho, pero presenta interferencia por ruido ambiental; un método más efectivo incluye un cristal piezoeléctrico en la cabeza de un eje metálico haciendo contacto con un diafragma. Actualmente, la comunidad científica y organismos como CORSA e *International Lung Sounds Association* (ILSA) promueven el uso de métodos digitales en la detección y tratamiento de enfermedades respiratorias.

3 Frecuencias Características y Herramientas para la Evaluación de Patologías Respiratorias

Hay dos clasificaciones de sonidos pulmonares (LS): los LS traqueales escuchados sobre la tráquea, de alta intensidad y ancho de banda entre de 0 y 2 kHz; además están los LS vesiculares, presentes a la altura del pecho, retirados de las vías aéreas centrales, con una frecuencia entre 0 y 600 Hz [2-5]. Para ampliar el rango espectral de las señales, se deben procesar mediante un micrófono o estetoscopio electrónico

con respuesta casi plana en el rango de 20Hz a 5 kHz. Las tasas de muestreo pueden ser tan bajas como 4 kHz y tan altas como 22.05 kHz. La duración de los LS varía entre 20 y 50 ms, implicando una tasa de muestreo de 10 kHz, con un bloque de señal de 256, 512 o 1024 muestras necesarias para la FFT [4].

3.1 Sonidos Adventicios

Crepitaciones: Sonidos explosivos de carácter transitorio, con una duración menor a 20 ms y espectro amplio, en un rango de 100 a 2000 Hz [6, 8, 9].

Graznidos: Sonidos cortos aspiratorios, raramente superando los 400 ms [6].

Sibilancias: Con frecuencia dominante entre 100-2000 Hz y duración entre 80-250 ms [6, 9, 11-14]. Útiles como clasificador epidemiológico [6, 9, 11-13].

Silbidos: Son sibilancias de volumen alto, caracterizados por un pico frecuencial prominente en 1 kHz [6].

3.2 Análisis de Sonidos de la Respiración

Los LS traqueales declinan rápidamente en intensidad después de 850-900 Hz; el sonido del corazón debería filtrarse mediante un pasa-altas con frecuencia de corte en 50-60 Hz, con el fin de lograr mejores resultados en la modelación. Cabe notar que los LS normales están contenidos debajo de 600 Hz.

Para efectos de suprimir las señales del corazón y aumentar la eficiencia, se aplicó un filtro Butterworth digital, cuya frecuencia de corte en pasa-altas está entre 50 y 120 Hz, obteniendo una nueva señal LS sin componentes del corazón ni componente en DC (Fig. 1). La Figura 2 muestra las formantes o picos principales en el dominio de la frecuencia, resaltando las zonas más energéticas de la señal. Esto es muy importante ya que la metodología de modelación implica el uso de un banco de filtros donde posteriormente se cuantifica la energía resultante de cada uno de ellos. Asimismo, la figura 2 evidencía una contribución energética del corazón.

3.3 Base de Datos RALE

RALE fue desarrollada en la universidad de Winnipeg, Canadá; contiene un repositorio de archivos obtenidos de registros de pacientes que presentaban respiración normal, crepitaciones, sibilancias y otros. RALE contiene más de 50 registros etiquetados, lo que permite crear modelos acústicos para estos casos; igualmente, contiene otros 24 registros sin etiquetar, lo que permite probar el sistema y evaluar un aprendizaje.

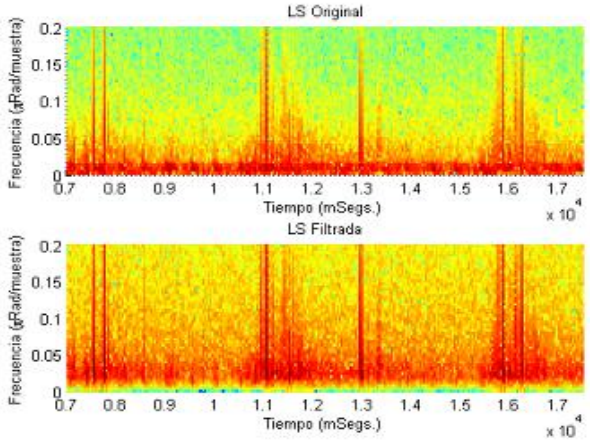


Fig. 1. Espectrogramas tiempo-frecuencia de la señal original (sup.) y filtrada con un Butterworth pasa-altas (inf.).

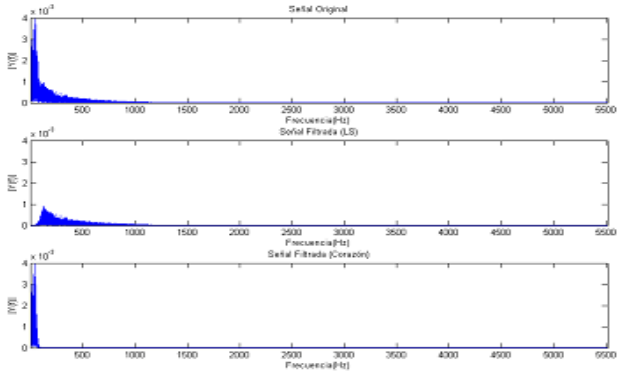


Fig. 2. Diagramas frecuencia-amplitud de la señal original LS (superior), la señal filtrada con pasa-altas (al medio) y filtrada con un pasa-bajas (corazón, inferior).

4 Metodología Sustentada en Modelos Acústicos GMM para el Reconocimiento de Sonidos Respiratorios

En un sistema de reconocimiento, la base de datos se divide en dos particiones: una de ellas para efectuar el cálculo de los modelos acústicos (aprendizaje o entrenamiento) [16]; la partición restante se emplea para estimar la eficiencia del sistema. Al no contar con un número apropiado para ambas etapas, se opto por evaluar con

validación cruzada (VC), la cual es una modalidad empleada cuando la cantidad de datos no es abundante [22].

4.1 Fase de Entrenamiento y Generación de Modelos Acústicos GMM

Un sistema de reconocimiento útil para un medico, debe de contar con modelos acústicos de las patólogas (o diccionario) a fin de emitir un diagnostico de la salud del paciente, como en el caso de asma. Por esta razón, una base de datos debe tener una gran cantidad de registros y lograr una hipótesis confiable sobre una nueva lectura de LS, especialmente en el caso de niños.

Los sonidos son parametrizados haciendo un preénfasis con filtros FIR (Respuesta al Impulso Finita), se le aplica una ventana Hamming cada 10ms con longitud de 30ms; se aplica el algoritmo FFT (Transformada Rápida de Fourier) por trama y de ahí se obtiene el modulo que multiplicamos por las escalas de Mel o de Bark, lo que se denota como vector acústico MFCC. Un modelo GMM (Modelo Mezclado Gaussiano) esta caracterizado por sus medias, covarianzas y ponderaciones; cada caso será representado por un modelo GMM (λ).

En la fase de entrenamiento se calculan los modelos acústicos para cada caso o patología, formando el llamado diccionario de modelos acústicos (en ingles *codebook*). Una señal de una misma patología tiene que ser grabada con múltiples pacientes para que sea representativa. Luego, dependiendo del caso, se le extraen los vectores característicos MFCC (Coeficientes Cepstrales en Frecuencia Mel) con una cierta longitud (por ejemplo, d=13), formando una clase acústica. Una vez que se tienen todos los vectores MFCC para un caso o patología, estos son empleados por el algoritmo de Máxima Expectación (EM) para calcular el correspondiente modelo acústico; una explicación mas exhaustiva se puede ver en [15-18]. Una alternativa muy eficiente para propósitos de inicialización, consistiría en efectuar primero el proceso con cuantización vectorial.

4.2 Fase de Evaluación

El método GMM calcula los modelos $\lambda_i = \{m_i, \mu_i, \Sigma_i\}$ valiéndose del algoritmo EM. La media μ_i representa el promedio de todas las observaciones (o vectores MFCC) y la matriz de covarianza Σ_i modela la variabilidad de las características en una clase acústica, siendo más eficiente si se centra en un rango de edades.

En nuestro caso, M es el numero de densidades para un modelo y mi es la ponderación para cada densidad Gaussiana dentro de una mezcla o modelo. Con el propósito de optimizar los modelos, se ejecuto el cálculo de los modelos con un número de densidades que variaron de 1 a 20, seleccionando el mejor compromiso entre eficiencia y numero de densidades en la mezcla. Utilizando la formula de Bayes y eliminando de esta *p(x)* por mantenerse constante en el proceso de maximización, se obtiene la formula fundamental en el reconocimiento automático de la señal, la cual proporciona la mejor hipótesis (Ecuación 1).

$$C = \arg_{1 \le i \le I} \max p(X \mid \lambda_i) p(\lambda_i) \qquad \text{Ec. (1)}$$

Tomando en cuenta que el total de grabaciones de sonidos para una patología o para respiración normal arrojan un número enorme de vectores MFCC y suponiendo independencia estadística entre ellos, la generalización de la regla para seleccionar la mejor hipótesis conduce a la Ecuación 2.

$$\prod_{t=1}^{T} p(\vec{x}_t \mid \lambda_i) > \prod_{t=1}^{T} p(\vec{x}_t \mid \lambda_r) \qquad \text{Ec. (2)}$$

En el contexto científico de procesamiento de voz y de reconocimiento de patrones [15-18], al término $\prod_{t=1}^{T} p(\vec{x}_t \mid \lambda_i)$ se le conoce como función de similitud. Pero como cada grabación arroja cuantiosos vectores acústicos, es necesario simplificar evitando desbordes computacionales, por lo cual es mejor presentar el *log* de la función de similitud (lo cual es válido, ya que es una función monótona y no cambia la relación > o <). A la expresión resultante se le conoce como la regla de decisión de máxima similitud (Ecuación 3)

$$L(\lambda_I) = \sum_{t=1}^{T} \log p(\vec{x} \mid \lambda_i) \qquad \text{Ec. (3)}$$

Esta última expresión, es la que se utiliza al decidir cuál es la hipótesis más probable. En otras palabras, la señal de entrada se asocia con el modelo acústico más probable dentro del diccionario. En nuestros experimentos, la evaluación con validación cruzada (VC) fue aplicada debido a que se contaba con una cantidad limitada de registros por caso (4-7 registros distintos por caso o señal adventicia) [22]. Las particiones se efectuaron seleccionando 3 registros para crear el modelo acústico y un registro para realizar la evaluación. Estas configuraciones se fueron alternando hasta completar 4 evaluaciones por caso.

5 Resultados

Las evaluaciones se efectuaron de dos maneras: una de ellas (denotándola "referencia", o REF) calculando los modelos acústicos con un conjunto de cuatro señales por caso y evaluando con estas mismas, examinando si el sistema es capaz de reconocer las señales con las cuales fue calculado el modelo acústico; la segunda modalidad fue efectuada mediante VC, ya que es la más objetiva. Los primeros experimentos efectuados con LS fueron para determinar el tamaño mínimo de GMM, extrayendo de RALE datos en forma de vector MFCC, construyendo un diccionario de modelos por caso, (tabla 1). Otro experimento consistía en determinar la longitud mínima necesaria del vector MFCC como se muestra en las tablas 1 y 2. Aquí se va-

rio el tamaño del vector y se considero la inclusión o no de sus derivadas. Los coeficientes MFCC están relacionados con la energía resultante de una serie de filtros aplicados, distribuidos linealmente hasta el orden de 1000 Hz y logarítmicamente a partir de ahí. La primera y segunda derivada es útil para ver la dinámica o evolución en las señales.

Los resultados con mejor comportamiento se obtuvieron aplicando 11 Gaussianas (GMM11), pero no son substanciales; sin embargo, fue evidente que modelando con vectores de 4 o más coeficientes los resultados mejoran, salvo algunas irregularidades (para GMM9, estabilizándose con GMM10). Esto último es razonable ya que con más de tres coeficientes incluimos las bandas más significativas en el caso de señales adventicias.

6 Discusión

En reconocimiento de voz o de locutor, se debe contar con una gran cantidad de datos (*i.e.* alrededor de 24 grabaciones de varios segundos por locutor, con frases fonéticamente equilibradas y alrededor de 200 locutores) a fin de capturar la estructura fina del aparato fonador en el modelo GMM.

Tabla 1. Tamaño mínimo aceptable en modelos GMM (numero de densidades) y vector MFCC (numero de coeficientes), así como modelos GMM con mejores resultados (GMM9, GMM10, GMM11). Llaves: #Den.= Numero de densidades; Coef. = Numero de Coeficientes; %Recon = Porcentaje de reconocimiento.

Experimento	#Den. mínimo (sin silbido)		#Den. mínimo (incluye silbido)		#Coef. sin Δ's		GMM9: %Recon.		GMM10: %Recon.		GMM11: %Recon.	
Evaluacion:	REF	VC	REF	VC	REF	VC	REF	VC	REF	VC	REF	VC
Normal	3	6	4	6	3	7	92.3	48	92.3	44.2	92.3	46.1
Crepitancia	3	6	4	6	3	7	100	90.3	100	98	100	98
Asma	3	6	4	6	3	7	100	50	100	51.9	100	50
Sibilancias	3	6	4	6	3	7	84.6	25	84.6	25	84.6	26.9

Tabla 2. Eficiencia de reconocimiento con vector MFCC incluyendo o no derivadas (deltas). Valores promediados con modelos GMM de 1-20 densidades.

Experimento:	13 Coef.		13 Coef.+13Δ.		13 Coef.+13$\Delta\Delta$.	
Tipo de evaluación:	REF	VC	REF	VC	REF	VC
Normal	100	46.2	95	48.7	95	51.2
Crepitancia	100	90	100	87.5	100	86.2
Asma	100	61.2	100	53.7	100	50
Sibilancias	100	30	100	31.2	100	32.5

Buscando reforzar los modelos acústicos, se aumento el corpus de señales LS normales, cuyos resultados experimentales se ven en la tabla 3, estandarizando aquí

las grabaciones, correspondiendo a 4 puntos de grabación, de la zona del pulmón bajo posterior y utilizando 28 grabaciones LS normales de esta base de datos protocolaria. El resto corresponden a RALE, las cuales son, por tipo: 8 de respiración normal, 4 de crepitancias, 5 de asma y 7 de sibilancia. Se llevaron a cabo numerosos experimentos, variando el tamaño de GMM y de los Vectores MFCC para la construcción del modelo. La tabla 3 muestra el resultado experimental y su eficiencia de reconocimiento.

Si se desea reconocer el habla, se requiere una cantidad de grabaciones de la misma frase con una diversidad de locutores, buscando que los modelos reflejen

Tabla 3. Eficiencia de reconocimiento con vector MFCC, utilizando el corpus de grabaciones protocolario.

Parámetros experimentales	Señal adventicia	No. de Grabaciones disponibles	No. de Grabaciones reconocidas
Mezclas Gaussianas = 10	Normal	36	34 / 36
Vectores MFCC = 13	Crepitancia	4	4 / 4
	Asma	5	5 /5
	Sibilancia	7	7 / 7

Características acústicas de los fonemas [16]. Desafortunadamente, no se conto con un numero favorable representar convenientemente la estructura fina del sistema respiratorio; RALE tiene un número pequeño de grabaciones por señal LS, que no corresponde a un mismo grupo de edad ni a los mismos puntos del cuerpo. A pesar de las limitantes, los experimentos muestran más del 50% de aciertos, evidenciando el potencial de la metodología.

En la mayoría de las evaluaciones, los modelos menos bien comportados fueron para sibilancia y normal, lo que hace pensar que respiración normal tiene un patrón menos definido. Requeriríamos más grabaciones y disponer el experimento por grupos de edades, ya que las características anatómicas y diferencia de edades impactan las características frecuenciales de las señales. Esta mejora fue constatada a través de los resultados, mostrados en la tabla 3.

Otra observación, es que la inclusión de silbidos LS en el diccionario, repercute en la necesidad de más datos para calcular los modelos y mantener la misma eficiencia. En el caso de sibilancias, solo se obtuvieron modelos acústicos bien comportados cuando se utilizaban vectores MFCC con un número de coeficientes mayores a tres. Similar al reconocimiento de voz, las características espectrales de la señal se encuentran en bandas superiores, correspondiendo a coeficientes superiores dentro de los vectores MFCC.

La señal de crepitancia, fue la más factible de reconocer; parece que tienen una banda de frecuencia característica con una buena contribución energética, distinguiéndola del asma, normal, sibilancias y silbidos. Es decir, los picos característicos de crepitaciones son estadísticamente suficientes para distinguirlas. Luego, repercute en un peso importante en la varianza y media, repercutiendo en un modelo GMM más robusto.

Las señales de asma son muy complejas, el concepto mismo de asma diverge y evoluciona; según los expertos, establecer asma a partir de sonidos adventicios no es suficiente. Como nuestros registros son limitados y heterogéneos, no muestran estadísticas concluyentes ni bandas energéticas distintivas. Sin embargo, obtuvimos valores de 50% o más en todos los casos (tablas 1 y 2), lo cual evidencia la utilidad del uso de estas técnicas para valoración de asma.

7 Conclusión

Llevamos a cabo experimentos con señales de la respiración normales y adventicias, aplicando vectores acústicos MFCC y modelos acústicos GMM. Dichas metodologías facilitaran la auscultación en casos de niños o en situaciones donde la persona que ausculte tenga sus sentidos auditivos debilitados y/o falta de pericia. Equivalentemente, prometen ser una alternativa confiable evitando la ambigüedad en la detección de señales adventicias por parte del médico o de los padres del paciente. En nuestros experimentos, variamos el tamaño de los modelos acústicos buscando un buen compromiso en términos de eficiencia y costo computacional. Los mejores resultados se obtuvieron con modelos de 9, 10 y 11 mezclas Gaussianas. Asimismo, constatamos que MFCC y GMM tienen potencial en la valoración de enfermedades en las vías respiratorias. Nuestro sistema arrojó 52.5% de eficiencia de reconocimiento al aplicar validación cruzada. Sin embargo, la eficiencia de reconocimiento de referencia fue del 98.75%, dejando de manifiesto que la metodología de reconocimiento de patrones amerita ser aprovechada en problemas de salud.

En un futuro, se pretende mejorar estas técnicas aplicando pre-procesamiento, tal como un filtrado para suprimir el ruido y el corazón en algunos registros. Definitivamente, ser pertinente extender el número de registros de la base de datos y llevar acabo análisis más concienzudos de las bandas importantes y de los coeficientes MFCC y explorar otros vectores acústicos. Además, deberíamos objetivar los experimentos a rangos de edad más delimitados, conduciendo a modelos más robustos y específicos de patologías o respiración normal.

Referencias

1. Miguel A. Serra Valdes, “Evaluacion de la Terapia Inhalada en Pacientes con Asma Bronquial Persistente”, http://www.revistaciencias.com/publicaciones
2. Leontios J. Hadjileontidais, “Biosignal and compression Standards, M-Health Emerging Mobile Health systems, Topics in Biomedical Engineering”, Springer 2006 International Book Series, pp. 277-292, ISBN 0387-26558-9.
3. Volker Gross, Anke Dittmar, Thomas Penzel, Frank Schutter, and Peter Von Wichert, “The Relationship between Normal Lung Sounds, Age, and Gender”, Am J Respir Crit Care Med Vol 162. pp 905909, 2000, www.atsjournals.org
4. A.R.A. Sovijrvi, J. Vanderschoot, J.E. Earis, “Standardization of computerized res- piratory sound analysis”, Eur Respir Rev 2000; 10: 77, 585, ISSN 0905 9180.5. J.E. Earis, B.M.G. Cheetham, “Current methods used for computerize respiratory sound analysis”, Eur Respir Rev 2000; 10: 77, 586590, ISSN 0905 9180.

5. A.R.A. Sovijrvi, L.P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi, J.E. Earis, “Characteristics of breath sounds and adventitious respiratory sounds”, Eur Respir Rev 2000; 10: 77, 591596 ISSN 0905 9180.
6. Efraın Carlos Nieblas Ortiz, Margarito Quintero Nunez, “Gestion Fronteriza para la Generacion electrica en la Region California, Estados Unidos-Ba ja California, Mexico”, Revista: Region Y sociedad Vol. XVIII, No. 37, 2006, ISSN 1870-3925.
7. M. Rossi, A.R.A. Sovijrvi, P. Piiril, L. Vannuccini, F. Dalmasso, J. Vanderschoot, “Environmental and subject conditions and breathing manoeuvresfor respiratory sound recordings”, Eur Respir Rev 2000; 10: 77, 611615, ISSN 0905 9180.
8. G. Charbonneau, E. Ademovic, B.M.G. Cheetham, L.P. Malmberg, J. Vanderschoot, A.R.A. Sovijrvi, “Basic techniques for respiratory sound analysis”, Eur Respir Rev 2000; 10: 77, 625635, ISSN 0905 9180.
9. Druzgalski, C., Potentials and Barriers of Extensive Auscultatory Databases, 27th International Conference on Lung Sounds, September 12 - 14, 2000, Stockholm, Sweden.
10. Druzgalski, C., Distributed Analysis of Signal Integrity Impediments in Respira- tory Acoustic Signatures. WC2003 - The World Congress on Medical Physics and Biomedical Engineering. August 21-30, 2003. Sydney, Australia.
11. Druzgalski, C.; Shenoy, N.; Kumar, S.; Enhanced Pulmonary Function Testing and Segmental Respiratory Performance Evaluation. The World Congress on Medical Physics and Biomedical Engineering 2006(WC2006), Aug. 27 B Sept. 1, 2006, Seoul, Korea.
12. Hans Pasterkamp, “State of Art: Respiratory Sounds Advances beyond the Stethoscope”, American Journal of Respiratory and Critical Care Medicine, Vol. 156. pp. 974987, 1997.
13. Henk J.W. Schreur, Jan Vanderschoot, Aeilko H. Zwinderman, Joop H. Dijkman and Peter J. Sterk, “Abnormal Lung Sounds in Patients with Function Asthma dur- ing Episodes with Normal Lung”, DOI 10.1378/chest.106.1.91, Chest 1994;106;91-99, ISSN:0012-3692.
14. Pearce D., “An Overview of ETSI Standards Activities for Distributed Speech Recognition Front-Ends”, AVIOS 2000: The Speech Applications Conference, San Jose, CA, USA, May 22-24, 2000.
15. Mayorga-Ortiz P., Besacier L., Lamy L. and Serignat J.F., “Audio Packet Loss over IP and Speech Recognition”, ASRU IEEE 2003 (Automatic Speech Recognition & Understanding), St. Thomas, Virgin Islands, USA, Nov. 1-Dec. 4, 2003, pp. 607-612.
16. Reynolds D. A, “A Gaussian Mixture Modeling Approach to Text-Independent speaker Identification”, Ph. D. Thesis, Georgia Institute of Technology, August 1992. 18. Andrew R. Webb, “Statistical Pattern Recognition”, John Wiley & Sons, Second Edition, 2002.
17. 19. Marco Antonio Reyna Carranza, Margarito Quintero Nu ñez, Kimberly Collins, “Correlation Study of the Association of PM10 with the Main Respiratory Diseases in the Populations of Mexicali, Baja California and Imperial County”, California, Revista Mexicana de Ingenierıa Biomedica, Volume 2 Numero 1 2005 M.
18. 20. S. Cortes, R. Jane, “Analisis de Sibilancias en Pacientes Asmaticos durante Res- piracion Espontanea”, XXV Jornadas de Automatica, Ciudad Real, Espa ña, 8-9 Septiembre, 2004.
19. 21. Andrew Bush, “Diagnostico de Asma en Ni ños Menores de Cinco”, Primary Care Respiratory Journal (2007), 16 (1) : 7-15., 1471-4418 2006 General Practice Airways Group. All rights reserved doi:10.3132/pcrj.2007.00001.
20. 22. Martinez W.L. and Martinez A.R., “Computational Statistics Handbook with Math Lab”, Second Edition, Chapman & Hall/CRC, 2008, ISBN 1-58488-566-1.

Sistema de Adquisición y Análisis de Señales Electroencefalográficas

C. Bustillo-Hernández, V. Rebollar-González, J. Figueroa-Nazuno,

Centro de Investigación en Computación
Instituto Politécnico Nacional
Unidad Profesional Adolfo López Mateos
chbustillo004@cic.ipn.mx, jfn@cic.ipn.mx
Paper received on 23/07/10, Accepted on 05/09/10.

Resumen. El Electroencefalógrafo juega un papel muy importante en la Medicina moderna, debido a que la información que provee puede ser usada para la detección y diagnóstico de patologías del cerebro. En los últimos años se han desarrollado métodos de análisis de señales de gran potencia y utilidad, capaces de ofrecer nuevos indicadores útiles en la determinación de patologías y estados fisiológicos. En este trabajo se presenta la implementación de un sistema de Electroencefalografía en hardware y software, donde se desarrolla la parte electrónica de forma completa a muy bajo costo y por otro lado, se acopla una etapa de análisis de señales con técnicas modernas muy poderosas para el estudio de la actividad eléctrica cerebral.

Palabras Clave: Análisis No Lineal, EEG, Electroencefalógrafo, Sistema

1 Introducción

El estudio de la anatomía y fisiología del cerebro humano es una de las tareas de investigación más importantes en el área de Neurofisiología; esto ha implicado el desarrollo de diversas técnicas que permitan obtener datos precisos acerca de su funcionamiento, una de las más utilizadas para su análisis es la Electroencefalografía, que permite explorar el cerebro mediante el registro de actividad eléctrica sin afectarlo o modificarlo, ya que es un procedimiento no invasivo.

Sin embargo el proceso de adquisición de dichas señales (que se caracterizan por ser de muy bajo voltaje), su procesamiento, visualización y análisis sigue siendo uno de los campos de investigación de diversas áreas y laboratorios con el objetivo de diseñar e implementar nuevos instrumentos y/o dispositivos, con mucha mayor precisión y velocidad, para adquirir estas señales en forma económica y rápida.

En este trabajo se presenta la implementación de un sistema de electroencefalografía (hardware y software), que permite realizar el registro y análisis de señales eléctricas del cerebro. Debido a que existe la necesidad de crear dispositivos de alta precisión accesibles que permitan obtener, ya sea los mismos resultados y/o inclusi-

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 149-158

ve mejores que los instrumentos comerciales, dado que éstos últimos resultan ser sumamente costosos. La implementación de electroencefalógrafos electrónicos basados en computadora es una propuesta técnicamente interesante, importante y de gran utilidad social.

Por otro lado, las técnicas de análisis que han sido utilizadas en Electroencefalografía como Fast Fourier Transform (FFT), desde 1932 por Hans Berger[1], no han tenido un avance significativo ya que se desconocen los nuevos métodos de análisis, derivados de distintas áreas, que puedan proponerse para obtener un diagnóstico más preciso.

2 Descripción del Sistema

2.1 Adquisición y Procesamiento de la Señal Electroencefalográfica.

Para la adquisición de la señal electroencefalográfica (EEG) se utilizaron electrodos con una aleación de cloruro de plata[1] y se implementó un circuito analógico que consta de las siguientes etapas: preamplificación, amplificación y filtrado. Debido a que la amplitud de la señal adquirida se encuentra en un rango de 1 a 100 μV_{pp} , incluyendo estados normales y patológicos [1], es necesario amplificar varias veces la señal antes de ser procesada. Sin embargo, este tipo de señal es sumamente sensible al ruido de 50 o 60Hz, que es el que se encuentra en instalaciones eléctricas comunes.

En la etapa de preamplificación se utilizó un circuito de protección eléctrica [1], el cual asegura que el paciente no sufra de alguna descarga electrostática y se configuró un amplificador instrumental con alta impedancia, haciendo posible medir la diferencia de voltaje entre dos electrodos. Esto asegura que un porcentaje de ruido no contamine la señal de interés.

Aunado a lo anterior, se implementa un amplificador denominado *driver right leg* (DRL), utilizado también en electrocardiografía (ECG), que permite atenuar el ruido de 50/60Hz en las entradas del amplificador instrumental. A continuación la señal pasa por un filtro pasa-altas que removerá los voltajes en corriente directa (DC) que se encuentren en la señal.

Posteriormente se amplifica la señal y se pasa por un filtro pasa-bajas a 50Hz tipo Bessel-Butterworth de tercer orden. Finalmente la señal obtenida es amplificada con un factor de 7808 veces por el circuito implementado y mediante una interfase muy sencilla se conecta a la PC para poder visualizar los datos y aplicar diferentes técnicas de análisis.[1]

2.2 Descripción de la Técnicas.

Las técnicas de análisis y procesamiento de señales basadas en métodos clásicos como la transformada de Fourier, en el sentido estricto, requiere que la señal sea infinita, estacionaria y lineal. Estas condiciones no se encuentran en señales reales dado que se caracterizan por tener tiempos finitos, no son lineales ni estacionarias.

E1 desarrollo de herramientas en software para el análisis de señales, es una de las necesidades básicas para diferentes áreas de investigación, siendo necesario emplear técnicas distintas, que no sean restrictivas en aplicaciones específicas y que el análisis realizado sobre la temática proporcione una mayor información.

Una característica muy importante dentro de los fenómenos descritos en el plano X-Y es que pueden tener muchas propiedades, las cuales no todas pueden ser visibles en el espectro de potencia. Sin embargo, en los últimos años han surgido nuevas técnicas de procesamiento y análisis que proporcionan una mayor información de la señal algunas provienen de áreas como: Matemática Moderna, Teoría de Sistemas Dinámicos No Lineales, Estadística Moderna y otras disciplinas que han desarrollado sus propias herramientas para la caracterización de sus fenómenos, por ejemplo en la Economía Moderna.

De acuerdo a lo anterior, en este sistema se implementan diferentes tipos de procesamiento y análisis de señales con técnicas modernas que puedan obtener diferentes características en diversos fenómenos, ya sea cuantitativamente o gráficamente. Además permiten hacer diferentes análisis de fenómenos que pueden ser en forma individual y/o en la comparación varias señales. La mayoría de estas técnicas están basadas en la teoría de sistemas dinámicos no lineales. A continuación se describe brevemente cada una de las técnicas implementadas.

1. **Cross Correlation**. La correlación es la relación que existe entre objetos, fe-nómenos o señales. El proceso para correlacionar dos señales discretas se denomina co-rrelación cruzada (cross-correlation.)

El algoritmo propuesto en esta técnica busca la relación de correspondencia entre dos señales utilizando operaciones de adición, multiplicación y estadísticas. En los resultados se puede observar gráficamente como se relacionaron ambas señales. [4]

2. **Dynamic Time Warping**. Es una transformación que permite la expansión y compresión de una secuencia, para la alineación local con respecto a otra secuencia, con el objeto de minimizar la distancia base, que comúnmente es la distancia euclidiana. DTW permite comparar y obtener un valor de semejanza entre secuencias que se encuentran desfasadas una con respecto a la otra, o cuando una de ellas presenta o ausenta segmentos con respecto a la otra. [5]
3. **Coupling Direction.** Permite determinar la interacción bidireccional o unidireccional de dos señales y cuantificar su grado de acoplamiento mediante un índice de sincronicidad. Se puede comparar dos sistemas o fenómenos distintos que presenten comportamiento caótico, ruidoso y estructuralmente diferentes.

   Algunas de las características de esta técnica son: a) revela la intensidad de interacción entre sistemas acoplados y b) caracteriza las relaciones "causales", esto es: si el segundo sistema depende fuertemente del primero, siendo el eje principal de la interacción, entonces la evolución del segundo depende del primero. En base a lo anterior, la "predicción" del segundo sistema puede ser improvisada a partir de los datos previos del primer sistema. [6,7]

4. **Alineación de Señales**. Esta técnica permite la comparación de señales que presentan diferente longitud y son similares en forma. Se obtiene como resultado del análisis una representación general de las series, es decir, una señal canónica.
   El algoritmo implementado permite que se pueda encontrar la similitud entre las señales de forma más directa. Esta técnica aborda el problema de comparación entre dos o más series donde existen cambios con respecto al tiempo, y es una aproximación al problema de alineamiento entre señales. [3]

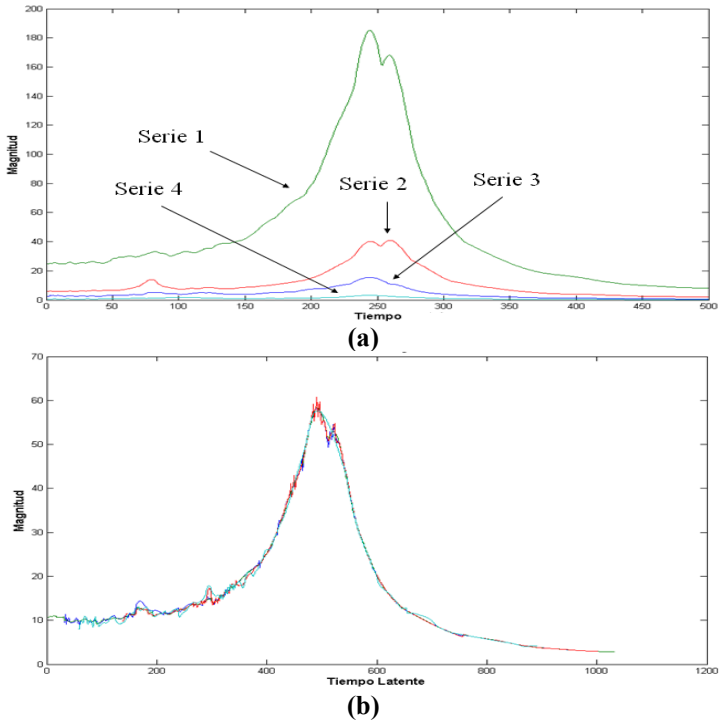


Fig. 1. Datos correspondientes a Espectros de Respuesta de una señal, a) Series desalineadas y no escaladas, b) Series Alineadas y Escaladas después de 10 iteraciones.

5. **Intrinsic Mode Function (IMF).** Un método alternativo a los análisis de Fourier para el procesamiento de datos no estacionarios provenientes de sistemas no lineales; es el referido como la Transformada Hilbert-Huang (HHT). La HHT se compone de la descomposición empírica en modos (EMD) y el análisis espectral de Hilbert. Cualquier registro en el dominio del tiempo, por más complicado que éste sea, puede ser descompuesto a través de la EMD en un reducido número de funciones de modos intrínsecos (IMF) que admiten la transformación Hilbert. Una IMF representa un modo de oscilación simple, similar a una componente

armónica de Fourier, pero mucho más general. La EMD explora la variación temporal en la escala de tiempo característica a cada conjunto de datos por lo que se adapta fácilmente a los procesos no lineales y a los datos no estacionarios. El análisis espectral de Hilbert define las frecuencias temporales (dependientes del tiempo) de los datos a través de la transformación de Hilbert de cada componente IMF.

La EMD es un método de análisis muy poderoso para datos no estacionarios. La técnica descompone los datos en funciones de modos intrínsecos IMF´s tal que la señal original sea igual a la suma de las IMF´s y un residuo.

El algoritmo básico tiene como objetivo identificar las IMF´s que representan los modos de oscilación, embebidos en los datos de la señal, a partir de dos condiciones simples: 1) En el conjunto de datos, el número de veces que atraviesa el eje cero es igual o difiere a lo más en uno, 2) En cualquier punto el valor medio de la envolvente definida por los máximos locales y la envolvente definida por los mínimos locales es cero (esta condición modifica un requerimiento global por uno local y es indispensable para asegurar que la frecuencia instantánea no presente fluctuaciones no deseadas, como las inducidas por formas de onda asimétricas). [8]

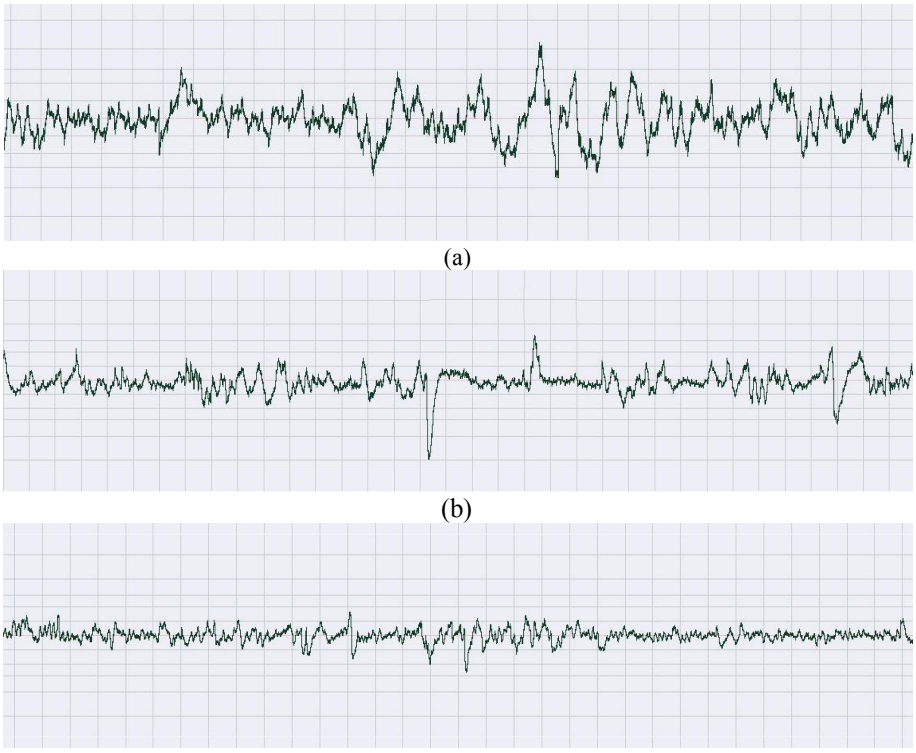
6. **Detección de Picos**. Su implementación esta basada en un algoritmo que determina los picos y valles de una señal deseada sin importar la presencia de ruido. Se basa en el principio de que un punto alto se encuentra entre dos bajos; la ventana de detección de picos se puede variar con la finalidad de ajustarse a las condiciones necesarias de la señal.
7. **Cross Sample Entropy**. Este método se basa en el análisis estadístico llamado Sample Entropy [1], con el cual se obtienen valores de sincronía y una cantidad especifica de alineamientos en una época determinada. Dada una época (E) queda establecido en un rango de tiempo determinado, esto es, donde n es el total de datos de la señal y todo esto esta basado en la Entropía de Shannon.[2]
8. **Pseudo Spectral**. Es una técnica alternativa para la obtención del espectro de una señal a través de un procedimiento basado en la codificación simbólica y en el análisis de strings de pulsos digitalizados obtenidos directamente de la señal; con lo cual se puede obtener un gráfico de pseudoespectro de la serie de tiempo analizada.

   Otra característica importante de esta técnica es que no requiere satisfacer las tres condiciones estrictas que se mencionaron al principio de la FFT, se puede obtener una representación o firma fácil de interpretar, y al igual que en el espectro obtenido por Fourier, hay pérdida de información, sin embargo se gana mucho en rapidez y simplicidad computacional.
9. **Recurrence Period Density Entropy (RPDE).** Esta técnica se utiliza para la caracterización de series de tiempo donde existen secuencias iguales repetidamente. Este procedimiento es similar al método de autocorrelación lineal e información mutua, excepto que esta medida de re-

petitividad se obtiene en el espacio de fase del sistema, de modo que etrae información basada en la dinámica subyacente del sistema. Una de sus ventajas es que no hace consideraciones de linealidad, distribución gaussiana de los datos o comportamiento determinista que pueda presentar el sistema. [9]

3 Ejemplos de Pruebas Experimentales

En esta sección se presentan ejemplos de las pruebas experimentales con las que se obtuvo señales electroencefalográficas, así como los diferentes análisis que se hicieron. Los resultados se obtuvieron de sujetos de sexo femenino con edades de 15, 22 y 23 años, en estado de vigilia; pudiéndose captar señales tipo alfa y beta.



(a)

(b)

(c)

Fig. 2. Señales electroencefalográficas obtenidas de sujetos femeninos de 15 y 22 años respectivamente con el sistema implementado con frecuencias de: a) 38Hz , b) 28 y 47Hz. y c) 30y 50Hz.

En la Fig. 3 se muestra la señal característica cuando se encuentran los ojos abiertos y los ojos cerrados, obtenido de un sujeto masculino de 33 años en estado

de vigilia. El resultado que se presenta es el adquirido con un equipo comercial de Electroencefalografía.

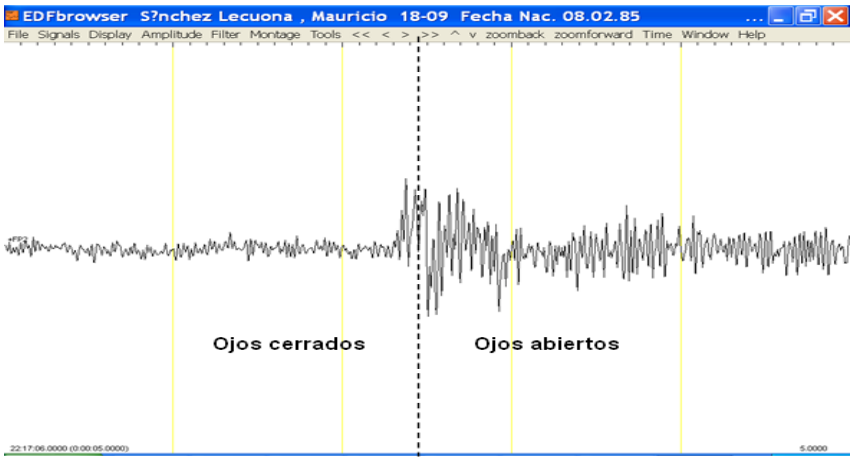


Fig. 3. Señal captada entre 30 y 50Hz. Montaje bipolar, electrodos frontales (Fp1 y Fp2)

En la Fig. 4 se presenta la misma condición obteniendo señal electroencefalográfica de electrodos frontales (Fp1 y Fp2) con el sistema propuesto.

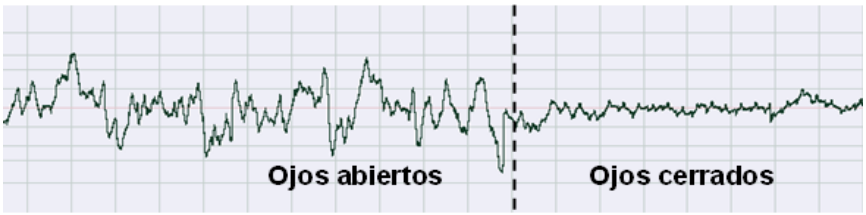
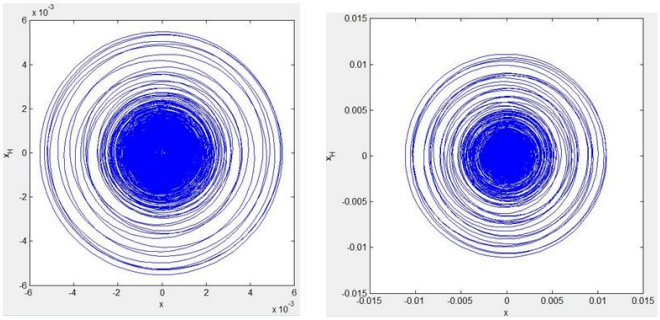


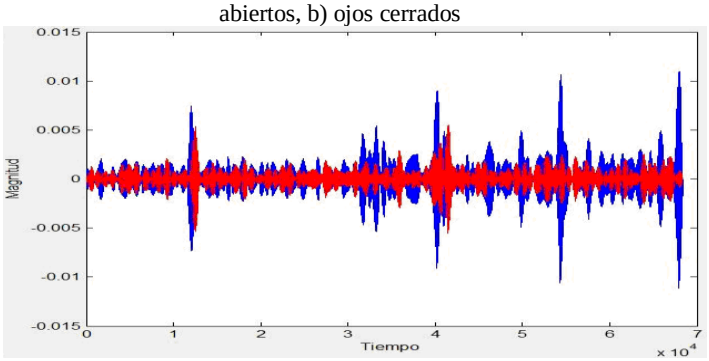
Fig. 4. Señal captada entre 30 y 50Hz, sujeto femenino 15 años.

Los valores obtenidos para la caracterización de los dos estados que se presentan en la Fig.4 con la técnica RPDE, para los datos cuando los ojos están abiertos fue H_{NORM}=0.68702 y ojos cerrados H_{NORM}=0.7434, esto indica que el fenómeno para ambos casos es complejo, ya que para señales periódicas el valor H_{NORM}=0 y en las estocásticas $H_{NORM} \approx 1$ [9]. Por otro lado, con la técnica Coupling Direction se logra observar mucho mejor las diferencias entre los dos estados: a) mediante gráficas de fase de los datos (Fig. 5) y b) verificando la Fig. 7 b) donde se observa el desacoplamiento de los datos en tiempo.

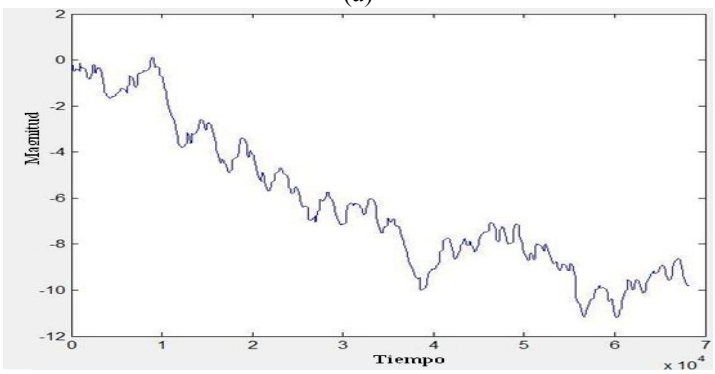


(a) (b)

Fig. 5. Gráficas de fase de los datos obtenidos con la técnica Coupling Direction a) ojos abiertos, b) ojos cerrados



(a)



(b)

Fig. 6. Resultados gráficos, a) datos filtrados y b) acoplamiento de ambas señales.

4 Discusión

En este trabajo, como se puede observar, esta compuesto de dos partes; la primera es la implementación electrónica directa de un electroencefalógrafo digital mediante componentes electrónicos modernos y técnicas de filtrado eficientes, esta

parte es un sistema cerrado que se encuentra entre los electrodos de registro de EEG que se colocan en el sujeto y una PC, en donde los datos obtenidos con dicha implementación electrónica pasan a la computadora. En esta segunda parte, se pueden hacer diferentes estudios utilizando las técnicas de Cross-Correlation, Dynamic Time Warping, Coupling Direction, Alineación de Señales, Intrinsic Mode Function, Detección de Picos, Cross Sample Entropy, Pseudo Spectral y Recurrence Period Density con las cuales, como se indica en la introducción, se pueden hacer análisis de los datos en forma diferente y más poderosa que sólo el simple registro de la señal electroencefalográfica o con técnicas clásicas como Fourier.

Por lo tanto, las contribuciones que se presentan son en dos aspectos, tanto en la implementación electrónica, económica y directa así como las técnicas de análisis de señales moderno que pueden facilitar el diagnóstico y la investigación de las señales de EEG.

5 Conclusiones

El registro de señales electroencefalográficas requiere de una instrumentación compleja y de técnicas de procesamiento adecuadas de análisis para una mejor utilización en el diagnóstico.

El sistema implementado en Matlab, muestra la utilidad para la obtención de datos, ya que permite observarlos, almacenarlos y puede emplearse para automatizar diagnósticos en análisis intensivos de señales.

Los resultados obtenidos con este sistema proporcionan más información sobre el comportamiento de la señal de EEG y además permite detectar propiedades muy complejas en su estructura, útiles para el diagnóstico.

Referencias

1. C. Bustillo Hernández.: Electroencefalograma Modular para su Uso en Pc´s. Tesis Licenciatura, Escuela Superior de Cómputo, IPN (2009)
2. T. Zhang, Z. Yang, J. H. Coote.: Cross-sample entropy statistic as a measure of complexity and regularity of renal sympathetic nerve activity in the rat. Exp Physiol, 92 (4) 659–669.
3. J. Listgarten, R. M. Nealy, S.T. Roweis, Andrew Emiliz.: Multiple Alignment of Continuous Time Series. Advances in Neural Information Processing Systems, MIT Press, Cambridge 17 817-825.
4. Chugani Mahesh, Shuler Charles.: Digital Signal Processing: A hands Approach. Mc. Graw Hill. (2000)
5. Angeles Yreta, J. Figueroa-Nazuno, K. Ramírez-Amaro.: Búsqueda de Semejanza entre Objetos 3D por Indexado. Reunion de Otoño de Comunicación, Computación Electrónica y Exposición Industrial ROC&C, IEEE Sección México, pp. 112-117., Acapulco, Guerrero, (2005).
6. Bezruchko, V. Ponomarenko.: Characterizing direction of coupling from experimental observations,Chaos 13 (1) pp.179-185
7. M. Rosenblum, L. Cimponeriu, A. Pikovsky.: Coupled oscillators approach in analysis of bivariate data, Handbook of Time Series Analysis. Wiley-VCH pp. 159-180 (2006)

8. H. Solís-Estrella, M. Ortega, F. Correa, K. Ramírez-Amaro, A. Angeles-Yreta, J. Figueroa-Nazuno y S. García.: Descomposición Empírica en Modos: una interpretación sísmica. XV Congreso Nacional de Ingeniería Sísmica, CCS84 (2005)
9. M. Little, P. McSharry, I. Moroz , S. Roberts.: Nonlinear, Biophysically-Informed Speech Pathology Detection, IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006 Proceedings, Toulouse, France. pp. II-1080-II-1083. (2006)

Computación basada en reacción de partículas en un autómata celular hexagonal

Rogelio Basurto Flores*
Paulina Anaid León Hernández †
Miguel Olvera Aldana‡
Genaro Juárez Martínez‡

Centro de Estudios Avanzados del IPN
Depto. de Microcomputadoras, BUAP
Escuela Superior de Cómputo, IPN
* rbasurto@computacion.cs.cinvestav.mx
† pleon@computacion.cs.cinvestav.mx
‡ Director de tesis
Paper received on 31/08/10, Accepted on 27/09/10.

Resumen. Se implemento computación a través de compuertas lógicas mediante reacción de partículas dentro de un autómata celular hexagonal.

1 Introducción

En el año 2005 Andrew Wuensche descubrió un autómata celular hexagonal, que representa un sistema de reacción - difusión basado en una reacción química de tres sustancias: *activador, inhibidor y sustrato*; el autómata celular fue bien recibido dado que presentaba *gliders,* partículas que pueden viajar por el espacio de evoluciones; gracias a estas partículas y a las colisiones entre ellas se logróllegar a la construcción de las compuertas lógicas básicas, mostrando así su universalidad lógica; estos resultados fueron presentados en [2]; debido al comportamiento que presentóse le da el nombre de regla Beehive. Su estudio demostró que a pesar de que presentaba una lógica universal aún tenía detalles a superar, como la ausencia de "glider-guns" estáticos partículas básicas para la computación en autómatas celulares [1], estos resultados fueron un trabajo presentado en [3], que presenta una nueva regla derivada de Beehive. Esta nueva regla es llamada "Spiral rule" (regla espiral), en adelante regla Spiral, y tiene la ventaja de presentar glider-guns estacionarios y móviles, gracias a estoy al antecedente de la regla Beehive se sospecha que la regla Spiral podría soportar una lógica universal a través de compuertas lógicas.

Para obtener los resultados de este trabajo se hizo uso del simulador que se puede encontrar en [18].

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 159.-168

2 Spiral rule

Un autómata celular se define omo l modelo de un *sistema complejo* que itera a través del tiempo, con una regla isotrópica de transición para el cambio de estados de las partículas individuales que lo componen, llamadas células, siendo el conjunto de estados que pueden tomar las células finito.

Tabla 1: Matriz de evolución de la regla Spiral

i \	0	1	2	3	4	5	6	7
0	0	1	2	1	2	2	2	2
1	0	2	2	1	2	2	2	
2	0	0	2	1	2	2		
3	0	2	2	1	2			
4	0	0	2	1				
5	0	0	2					
6	0	0						
7	0							

De manera formal un autómata celular está compuesto por la tupla:

- Q es un conjunto finito de estados.
- d, un número entero positivo que representa la dimensión del autómata
- e, es la configuración inicial de las células
- $f: S \rightarrow S$, es la función de transición local

En el particular caso de la regla Spiral se tiene un conjunto $Q = \{0, 1, 2\}$ representados mediante los colores blanco, rojo y negro respectivamente; la dimensión $d = 2$ y además, es importante mencionar que la forma de las células es hexagonal y debido a ello la vecindad utilizada será como la presentada en la figura 1, con 6 vecinos inmediatos; e representa la configuración inicial, es decir, los estados de las células para el tiempo $t = 0$; y finalmente, la función de transición estará dada mediante una matriz de transición, donde las columnas representan el número de células en estado 1 y las filas el número de células en estado 2, dicha matriz de transición es presentada en la tabla 1.

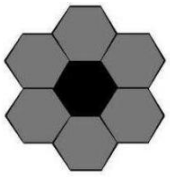


Figura 1: Vecindad hexagonal.

Lo anterior se ve directamente reflejado al hacer una evolución paso a paso de una configuración inicial sencilla como la mostrada en la figura 2, donde se presenta únicamente una célula en estado 1.

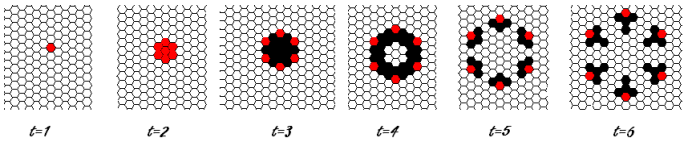


Figura 2: Evolución de una configuración inicial con sólo una célula en estado 1.

3 Dinámica y partículas básicas

La regla de evolución, representada con la matriz 1 permite observar la interacción entre células, lo que a través de las evoluciones hará que se formen estructuras de forma inherente llamadas partículas. Dichas partículas son *gliders*, *glider-guns* y *eaters*; un *glider* es una partícula que se puede mover a través del espacio de evoluciones; los *gliders-gun* son partículas más complejas que después de un determinado número de evoluciones lanzan un *glider*; el *eater* es una partícula estática capaz de eliminar otras partículas. En la figura 3 se muestran los tipos básicos de *gliders* que presenta el autómata, de ellos se podrán obtener diferentes datos relevantes al momento de utilizarlos con fines computacionales:

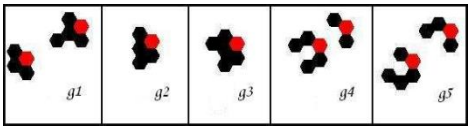


Figura 3: Tipos de gliders.

- *Peso:* se refiere a la cantidad de células que forman una partícula; de tener más de una fase (o forma) se toma la mayor como su peso.
- *Período:* es el número de evoluciones necesarias para que una partícula vuelva a tener su estructura inicial.
- *Desplazamiento:* se refiere al número de células que avanza el glider por período, se define su dirección en base al plano cartesiano.
- *Velocidad:* está determinada como P /D, es decir, Período/Desplazamiento.

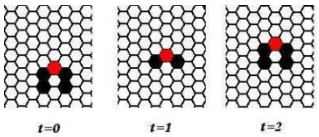


Figura 4: Glider g4.

Para explicar estos conceptos se puede tomar como ejemplo el *glider* g_4. Su peso es 5, pues tiene 5 células que lo conforman en su fase más grande; su período es 2, esto puede apreciarse de la figura 4, donde se observa

que le toma 2 tiempos regresar a su forma inicial. En la misma imagen se observa el desplazamiento que tiene el glider, donde avanza una célula cada tiempo y, sí para completar un período necesita de dos tiempos, entonces tiene un desplazamiento igual a 2; finalmente, la velocidad del glider g_4 se obtiene dividiendo su período entre su desplazamiento, es decir, $\frac{2}{2} = 1$. De esta manera se han analizado los 5 gliders, la información obtenida se muestra en la tabla 2.

Tabla 2: Características de gliders

Partícula	Peso	Velocidad	Período	Desplazamiento
g_1	5	1	2	2
g_2	5	1	1	1
g_3	6	1	1	1
g_4	5	1	2	2
g_5	5	1	2	2

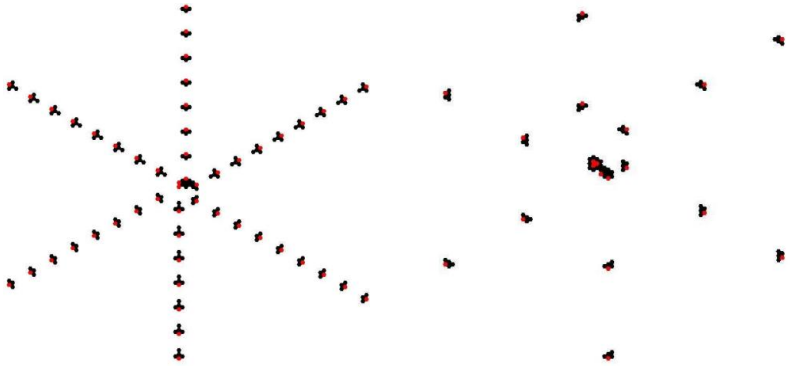


Figura 5: Glider-gun G1

Figura 6: Glider-gun G

La creación de gliders es muy importante, y para ello la regla Spiral tiene dos tipos de gliders-gun básicos, también llamados "guns", estos son mostrados en las figuras 5 y 6. La información que se puede obtener de estos guns está condensada en la tabla 3. De la tabla, se define la frecuencia como el número de gliders que produce en un período, y el período se define de la misma manera que para los gliders.

Tabla 3: Características de glider-guns.

Glider Gun	Glider que produce	período	Frecuencia
G_1	g_1	6	6
G_2	g_2	22	6

Otra partícula importante es aquella capaz de eliminar los gliders, se le denomina eater. Existen dos tipos básicos de eaters en la regla Spiral, estos se presentan en la figura 7. Sin importar cuantas veces y en que ángulo

choque un glider con un eater, este será destruido.

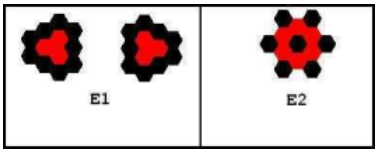


Figura 7: Tipos de eaters.

4 Computación en la regla Spiral

Mediante la manipulación de las partículas básicas que presenta la regla Spiral es posible implementar computación, y se demuestra mediante compuertas lógicas.

La característica de los glider-guns de este autómata que les permite lanzar gliders en 6 direcciones puede llegar a ser una ventaja, pues se podrían procesar 6 señales al mismo tiempo, no obstante, por ahora sólo se considerará un solo flujo; los flujos de gliders que no se utilicen serán eliminados mediante un eater E1. Un glider-gun limitado en 5 de sus 6 flujos se puede apreciar en la figura 8. Esto mismo se puede extender para el glider-gun G2.

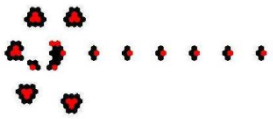


Figura 8: Glider-gun G1 con 5 flujos eliminados.

De la misma manera que en Life [8], en la regla Spiral se representan unos lógicos ‘1’, con la presencia de gliders y ceros lógicos ‘0’, con la ausencia de los mismos. Así, utilizando el glider-gun G_1 se puede representar una cadena constante de información con 1’s. La manera de cambiar esta cadena y poder hacerla más diversa es mediante el glider-gun G_2. Al lanzar los gliders a una frecuencia más baja es posible modificar el flujo de gliders del G_1 para generar cadenas de unos y ceros, un ejemplo de esto se muestra en la figura 9.

La sincronización entre glider-guns es una de las bases primordiales para la creación de compuertas lógicas, no solo en la regla Spiral, sino en cualquier autómata, pues la computación está basada en las colisiones entre las partículas y la reacción que estas colisiones producen [2],[9],[11], [12]. Después de observar la regla Spiral se notó una similitud entre las colisiones existentes entre gliders, dichas colisiones se muestran en la figura 10. En la imagen se observan tres colisiones diferentes, la colisión del inciso A tiene como reacción el cambiar de dirección del glider proveniente del suroeste; en el inciso B se observa la aniquilación de ambos gliders;

el resultado de la colisión del inciso C es la eliminación de un solo glider, mientras el otro sigue su curso normalmente. Estas y otras colisiones se utilizan como función para la construcción de las compuertas lógicas.

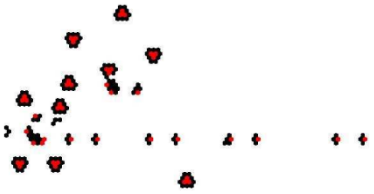


Figura 9: Flujo de gliders modificado.

Otra característica más a considerar son las partículas "excedentes", es decir, gliders que se generan y no son útiles para la computación propuesta, dichos gliders son eliminados mediante eaters. Finalmente, para construir una compuerta lógica y que su resultado sea fácilmente verificable es necesario construir un flujo de entrada que contenga los bits necesarios para comprobar la tabla de verdad de la compuerta implementada.

Dado que no es posible predecir el comportamiento general del autómata, se realizaron una serie de pruebas empíricas para poder encontrar la implementación de las compuertas lógicas; esto con ayuda de un simulador que permitía la manipulación de los estados mediante una interfaz gráfica que fue desarrollado para este fin.

Las compuertas implementadas mediante la regla Spiral son: AND, OR y NOT así, con estas compuertas la regla Spiral posee una lógica universal.

A lo largo de la búsqueda se logró encontrar otras compuertas más, estas son: NOR, XOR y XNOR.

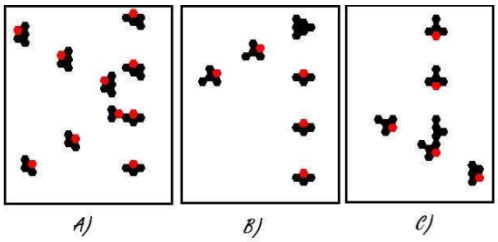


Figura 10: Tipos de colisiones entre gliders.

EL orden cronológico de creación es el siguiente: NOT, AND, NOR y XNOR. Debido a que se tenían las compuertas NOR y XNOR se hizo uso de la compuerta NOT para de esta manera crear las compuertas OR y XNOR.

Tabla 4: Tabla de verdad de la compuerta NOT.

A	S
0	1
1	0

Tabla 5: Tablas de verdad para las compuertas AND, OR, NOR, XOR y XNOR.

Entradas		Salidas				
A	B	AND	OR	NOR	XOR	XNOR
0	0	0	0	1	0	1
0	1	0	1	0	1	0
1	0	0	1	0	1	0
1	1	1	1	0	0	1

La interpretación en el autómata celular de las compuertas lógicas se presenta a continuación:

- La compuerta NOT está formada por dos glider-gun G_1 y un glider- gun G_2 , este último se utiliza para modificar la señal de entrada de la compuerta; se observa la compuerta en la figura 11, donde A es el glider-gun que tiene el flujo de entrada y S es el flujo de salida de la compuerta. La señal de entrada es: 1100110; por lo que su flujo de salida es: 0011001.

- En la compuerta AND, que se puede ver en la figura 12, se utiliza un glider-gun G_1 por cada señal de entrada, de igual manera, para modificar el flujo de gliders se requiere un glider-gun G_2 por cada flujo de entrada; el flujo A es: 1111010; para la entrada B se utiliza:1101100; el resultado de aplicar la operación AND se muestra con la salida S y es: 1101000.

- Para construir la compuerta OR se partió de la NOR, misma que requiere tres glider-guns G_1, dos para las entradas A y B, y uno que forme parte del proceso de transformación de los gliders para generar el resultado; también se utilizan cuatro G_2 para modificar los flujos de entrada, dos por cada flujo. Las cadenas de bits que representan los flujo de los gliders de entrada son: 1100100 para la entrada A y 1101110 para la entrada B, siendo el resultado: 1101110; posteriormente se paso a utilizar la compuerta NOT, así obteniendo la compuerta OR. En la figura 13 se puede observar la compuerta OR.

5 Discusión, resultados y trabajo a futuro

Se ha analizado la regla Spiral y las colisiones entre partículas, así como sus reacciones. Gracias a estas construcciones se ha demostrado que es posible implementar computación dentro de la regla Spiral para autómatas celulares hexagonales; también, se ha demostrado la lógica universal de la regla al tener las tres compuertas lógicas básicas: AND, OR y NOT. Derivado de estas construcciones se logró implementar otras compuertas que serán de ayuda en la búsqueda de construcciones más complejas. Por ejemplo, un objetivo siguiente puede ser la construcción de un medio sumador utilizando la compuerta XOR y AND que se tienen actualmente. Además de utilizarlas para cons-

truir dispositivos más complejos, simular una función computable completa o algún otro sistema activador, inhibidor y refractario; incluso cualquier otro sistema no-lineal con dicha dinámica. Dentro de la regla Spiral existen partículas que aún no han sido utilizadas con fines computacionales, en este sentido queda abierta la posibilidad de encontrar más partículas, o tomar algunas de las ya encontradas para utilizarlas en nuevas construcciones, o alguna construcción derivada de las ya existentes.

Los espacios de evolución utilizados durante las pruebas y construcciones mostradas en el presente trabajo son de 160 × 160 y 240 × 240 células, lo que hace pensar que al realizar construcciones derivadas de las actuales, será necesario utilizar espacios de evoluciones más amplios, lo que conlleva un mayor procesamiento y una visualización menos agradable; es por eso, que se propone como trabajo a futuro utilizar un clúster de visualización. Finalmente, resta pensar en los diferentes tipos de análisis estadísticos que pueden ser mostrados por el simulador; por ejemplo, un análisis de densidad de estados, o de partículas "vivas" dentro del espacio de evoluciones.

Referencias

1. Andrew Adamatzky (Ed.) (2002), *Collision-Based Computing*, Springer.
2. Andrew Adamatzky, Andrew Wuensche and B. De Lacy Costello (2005), "Glider-based computing in reaction diffusion hexagonal cellular automata", *Journal Chaos, Solutions & Fractals*.
3. Andrew Adamatzky and Andrew Wuensche (2006), "Computing in Spiral Rule Reaction-Diffusion Hexagonal Cellular Automaton," *Complex Systems* 16 (4).
4. Andrew Adamatzky and Christof Teuscher, *From Utopian to Genuine Unconventional Computers*, Luniver Press, 2006.
5. Andrew Ilachinski, *Cellular Automata: A Discrete Universe*, World Scientific Press, 2001.
6. Andrew Wuensche (2004), "Self-reproduction by glider collisions; the beehive rule," *International Journal Pollack et. al*, editor, Alife9 Proceedings, MIT Press.
7. Andrew Wuensche (2005), "Glider Dynamics in 3-Value Hexagonal Cellular Automata: The Beehive Rule," *International Journal of Un- conventional Computing*.
8. Elwyn R. Berlekamp, John H. Conway y Richard K. Guy, *"Winning Ways for Your Mathematical Plays"* Volume 4, Second Edition, 927- 961, A K Peters.
9. Genaro J. Martinez, Andrew Adamatzky, Harold V. McIntosh, and Ben De Lacy Costello, "Computation by competing patterns: Life rule B2/S2345678" desde *Automata 2008: Theory and Applications of Cellular Automata* páginas 356–366, Luniver Press, 2008.
10. Genaro J. Martínez, Adriana M. Méndez, y Miriam M. Zambrano,Un subconjunto de autómata celular con comportamiento complejo en dos dimensiones, Escuela Superior de Cómputo, Instituto Politécnico Nacional, México, 2005, http://uncomp.uwe.ac.uk/genaro/Papers/Papers on CA.html.
11. Genaro J. Martinez, Andrew Adamatzky, and Ben De Lacy Costello,(2008) "On logical gates in precipitating medium: cellular automaton model" Physics Letters A 1 (48), 1-5.

12. Genaro J. Martinez, Andrew Adamatzky, and Harold V. McIntosh (2006), "Phenomenology of glider collisions in cellular automaton Rule 54 and sso-ciated logical gates", Chaos, Fractals and Solutions 28, 100-111.
13. Harold V. McIntosh, *One Dimensional Cellular Automata*, Luniver Press, 2009.
14. Howard Gutowitz (1991), *Cellular Automata, Theory and Experi- ment*, The MIT Press.
15. Stephen Wolfram (2002), *A New Kind of Science*, Champaign, Illinois, Wol-fram Media Inc.
16. Tommaso Toffoli and Norman Margolus (1987), *Cellular Automata Machines*, The MIT Press, Cambridge, Massachusetts.
17. Von Neumann, John (1966), *"Theory of self-reproducting automata"*, Urbana and London, University of Illinois.
18. Simulador de la regla Spiral para Windows; http://uncomp.uwe.ac.uk/genaro/Papers/Thesis.html

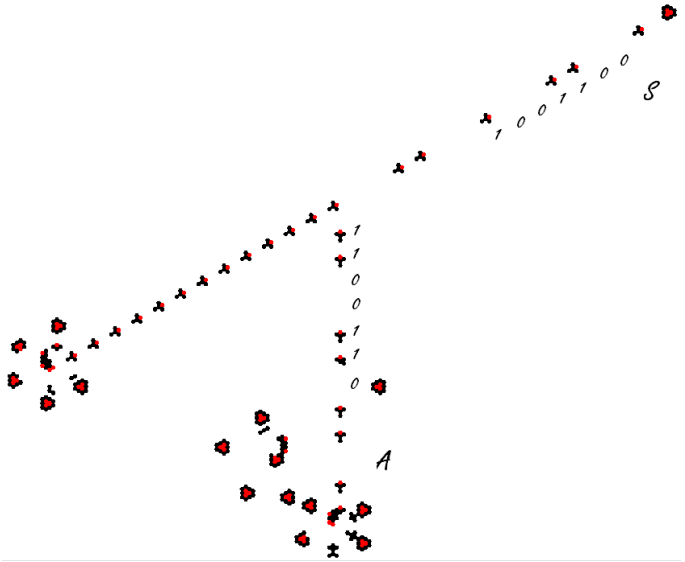


Figura 11: Compuerta NOT en la regla Spiral.

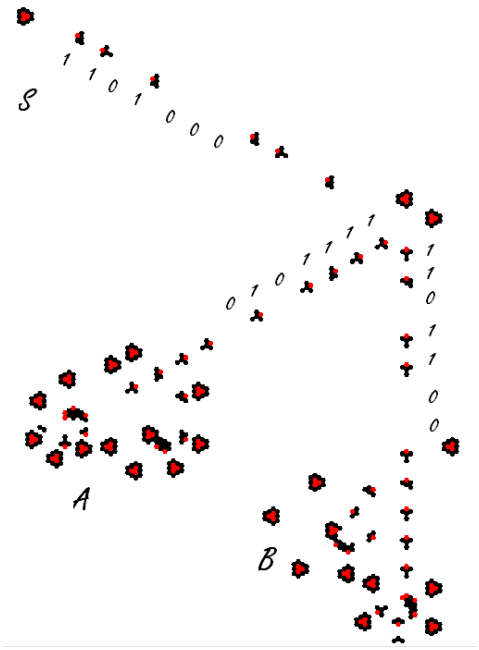


Figura 12: Compuerta AND en la regla Spiral.

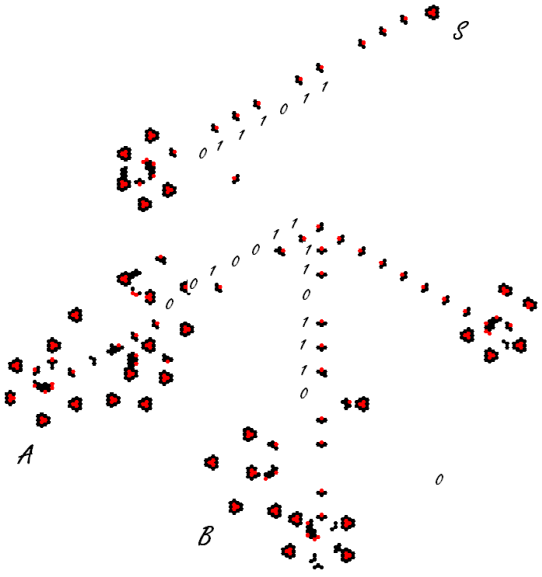


Figura 13: Compuerta OR en la regla Spiral.

A Centralized Self-Healing Architecture as Support to Web Services Applications

Francisco Moo-Mena, Juan Garcilazo-Ortiz, Luis Basto-Díaz, Fernando Curi-Quintal, and Fernando Alonzo-Canul

Universidad Autónoma de Yucatán – Facultad de Matemáticas
Periférico Norte km 33.3, Merida, Mexico
{mmena, gortiz, luis.basto, cquintal}@uady.mx, fernandoalonzo.0@gmail.com
Paper received on 14/08/10, Accepted on 26/09/10

Abstract. Web Services (WS) technology proposes tools to implement complex distributed systems. It allows creating networks of heterogeneous and cooperative WS on applications. Due to the complexities related to the dynamics of this kind of system (new WS could come in and come out of the system), applications are exposed to several failure points in both components and connections. Current efforts in WS literature offer a supporting infrastructure for this problem; most of them are oriented to repairing, mainly on application's workflow level. In this paper, we expose our ideas about an infrastructure based on Quality of Service (QoS) analysis of WS-based applications. Our approach considers a complementary strategy, supported by the adaptability of the application's architecture. This strategy is planned to prevent and respond to QoS degradation in WS-based applications, trying to enhance their performance. Our solution is oriented to discover new approaches for the three main stages identified in the related literature for self-healing infrastructure: Monitoring, Diagnosis and Recovering. In order to show our strategy effectiveness, we implemented it in the access to a Digital Library based on WS.

Keywords: Self-Healing System, QoS, Box-Plot Diagram, Web Services Application, Digital Library.

1 Introduction

Currently, evolution of WS technology offers a great potential in order to develop complex applications in different contexts, such as enterprises, industry, government, or home. Implementing these applications requires creating a network of cooperative WS. Despite the fact that WS technology deals with component's heterogeneity, these WS networks present the inconvenience to be open (exposed to everyone) and highly dynamics (new WS can come in and others come out). To deal with both conditions is an actual challenge.

In a general context, this set of problems can be extrapolated to the complex distributed applications development. In this sense, IBM presented, early this decade, an initiative named Autonomic Computing [1], which approaches these problems proposing four axis of study:

©E.Cuatecontzi Cuahutle, A.Cortés Fernández, J. F. Ramírez Cruz, J. H. Sossa Azuela. (Eds.). Advances in Intelligent and Information Technologies. Research in Computing Science 50, 2010, pp. 171-181

1. *Self-configuring*, deals with components and systems reconfiguration by defining high-level politics. *Self-healing*, deals with auto-detection, diagnosis, and recovering in hardware and software problems.
2. *Self-optimizing*, deals with parameter auto-adjusting at service level.
3. *Self-protecting*, deals with system protection against attacks and intrusions.

Our work focuses mainly on self-healing systems whose critical aspects are: a) mechanisms for system's health maintenance process, b) system's failure detection, and c) system's recovery processes [2].

In general, self-healing process consist of three stages, [3]:

1. *Monitoring*. This stage is responsible for monitoring and recording information about the status of system, activity in order to maintain the reliability level and quality of service.
2. *Diagnosis*. Information collected is examined, and the cause of malfunction or failure is determined. In this stage, a system assessment is done using a reference model.
3. *Recovery*. This stage is responsible for maintaining or restoring the correct state of the system.

Attending to these problems, based on IBM's initiative, several projects have emerged aiming to create a supportive infrastructure for applications. Some examples of these efforts are [4], [5], [6].

This document describes an approach proposing a strategy for WS-based applications with auto-adaptation capabilities. This strategy is based on implementation of architectural reconfiguration techniques, such as WS duplication and substitution defined on previous work [7], [8]. Decision about when to apply an architectural configuration action is determined by defining, monitoring and diagnosing QoS parameters by using statistical techniques based on box-plot diagrams. The main goal in diagnosis techniques is to prevent QoS degradation when executing WS-based applications.

Typically, recently developed WS applications are oriented to business and services areas, for example, travel agency systems and "production chain" process oriented systems, such as client/supplier/producer systems. Thus, the implementation of our strategy on a digital library application represents another innovative proposal.

The term "digital library" has been used to define a large digital information storage accessed through computer networks [9]. As in traditional libraries, a digital library serves as a knowledge archive of different subjects. Since information can be filed in different formats, a digital library would contain text, audio, image, animation and video files. Digital libraries posses features like creating digital documents, indexing and searching, management and access control, personalization, information distribution, suitable interface to deliver results, etc.

The rest of this document is organized as follows: the second section presents some important works with regard to self-healing systems. The third section defines our general strategy for designing and implementing a self-healing infrastructure for WS-based applications. The fourth section shows some important results of this work. Finally the last section describes the conclusion and future perspectives for this work.

2 Related Work

In the context of Autonomic Computing some important works following a self-healing approach are:

The Bio-Networking Architecture [10] is a paradigm as well as a middleware that enables the construction and deployment of scalable, adaptive, and survivable/available applications. In this work, applications are constructed using a collection of autonomous mobile agents called cyber-entities, and the Bio-Net Architecture platforms (execution environments and support services for the cyber-entities). It is based on principles and mechanisms used by biological systems to adapt to changing environmental conditions, like colonies of ants and bees.

The Hydra project [11] is an Integrated Project that develops middleware for Networked Embedded Systems. The Hydra project is co-funded by the European Commission. It is a middleware that allows developers to incorporate heterogeneous physical devices into their applications by offering easy-to-use-Web service interfaces for controlling any type of physical device irrespective of its network technology such as Bluetooth, RF, ZigBee, RFID, WiFi. Hydra incorporates means for Device and Service Discovery, Semantic Model Driven Architecture, P2P communication, and Diagnosis.

The CODA project (Complex Organic Distributed Architecture) [12], includes a means monitoring and controlling objectives to allow the enterprise to evolve with certain degree of autonomy. The architecture includes concepts and principles of Self-organization, Self-regulation toward an intelligent architecture. CODA is organized into five functional layers, each with a storage component and an intelligent component:

1. *Operation:* This layer manages data and connects simple linear data from databases in different locations.
2. *Monitoring Operation:* This layer ensures internal monitoring.
3. *Monitors Monitor:* in this layer operations are monitored in terms of an external monitoring.
4. *Control:* layer that has the ability to learn about the behaviour, trends and predictions.
5. *Command:* the highest-level layer that is responsible for recognizing threats and opportunities.

In the area of Web services and within this category is the WS-DIAMOND (Web-Service Diagnosability, Monitoring & Diagnosis) project [13], which considered the study and implementation of methodologies and solutions by creating self-healing Web services able to detect abnormal operation such as the inability to respond to service's requests or failure to achieve a level of QoS stipulated as a requirement. As well as to restore service from failures by restructuring or reconfiguring network services. WS-Diamond also provides methodologies for the design, support and service running mechanisms to ensure monitoring, diagnosis and fault recovery at runtime.

3 Self-healing Architecture

Our architecture is oriented to Web service's interaction (interaction between two WS). There are always two main actors, the consumer (service

requester) and the provider, the main scenario where the architecture works is divided into two phases; the first one is used to collect the information regarding to Web service's interaction and is represented by an interceptor, and the second one is in charge of processing the data collected and applying the corresponding self-healing mechanism.

3.1 Interceptors

Interaction between two WS is very simple; there is a consumer and a provider (see figure 1). We added a third actor, the interceptor located in both the WS consumer and the WS provider in order to register the following records:

- T1: *Service Request's start time*. Time at WS consumer sends the request.
- T2: *Service Request's end time*. Time at request arrives to WS provider.
- T3: *Service Response's start time*. Time at WS provider sends the response.
- T4: *Service Response's end time*. Time at WS consumer receives the response.

All the four monitoring records (T1, T2, T3, and T4), help us to define the QoS parameters such as Computation Time, Requesting Time, Responding Time and Communication Time, those parameters are calculated as follows:

- QoS1:= T3 - T2 : *Computation Time.*
- QoS2:= T2 - T1 : *Requesting Time.*
- QoS3:= T4 – T3 : *Responding Time.*
- QoS4:= (T4 - T1) - QoS1 : *Communication Time.*

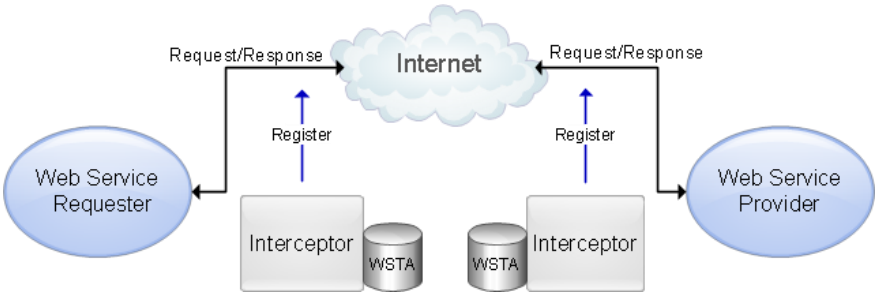


Fig. 1. Interactions between Web services.

In order to use our self-healing architecture the WS consumer needs to implement our Client API, which provides all the interfaces required to send a service request to provider and receives its response. With Client API, client uses the self-healing benefits in a transparent way. Consumers use the Client API in the following way:

1. Create a Client instance.
2. Create a Request instance.
3. Define Web service and operations desired on request instance.
4. Associate the request instance to client instance.
5. Invoke client instance and execute function.
6. T1 is registered and sends transaction to provider.

7. Provider interceptor registers T2 measure.
8. Provider processes Request.
9. Interceptor registers T3 measure.
10. Response arrives to Consumer.
11. T4 measure is registered.

3.2 Self-healing Components

The self-healing core architecture (see figure 2) shows the main components: monitoring, diagnosis, recovering and interceptors.

Also, there are three elements used as information repositories, two of them related to QoS. The Parameter Repository (PaRe) stores all the QoS measures registered by interceptors. The Service Level Parameters (SLP) represents the optimal measures to be accomplished in every Web service's interactions, that is, the contract agreement. The last information repository component is the WS Table Access (WSTA). Conceptually, it has the service information: identification, description and the WS in charge of response to the service request.

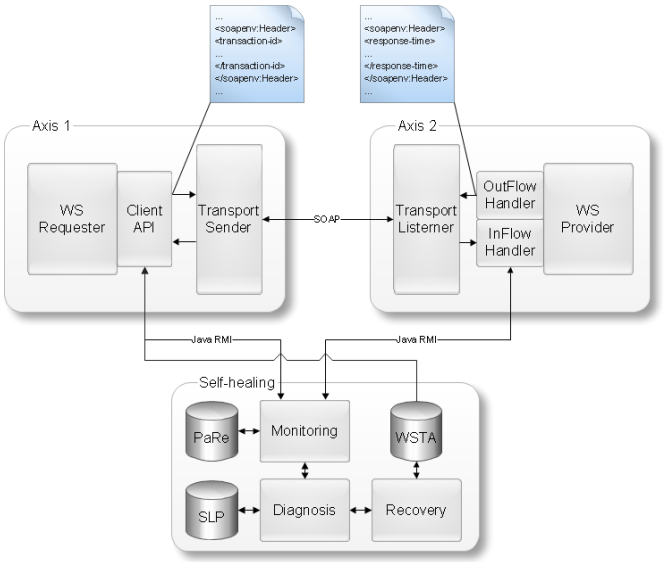


Fig. 2. Self-healing core architecture.

3.3 Monitoring Module

Monitoring module is the only interface between interceptors and architecture's core, being its main objective to collect data between interceptors and store them into PaRe. Interaction begins when a consumer sends a request service to provider and ends when the consumer receives a request response. At the same time if a failure is detected, immediately monitoring module will be alerted, (see figure 3).

Monitoring module implementation uses RMI as a way to receive data from interceptors and sends data to persistence databases through Hibernate [14].

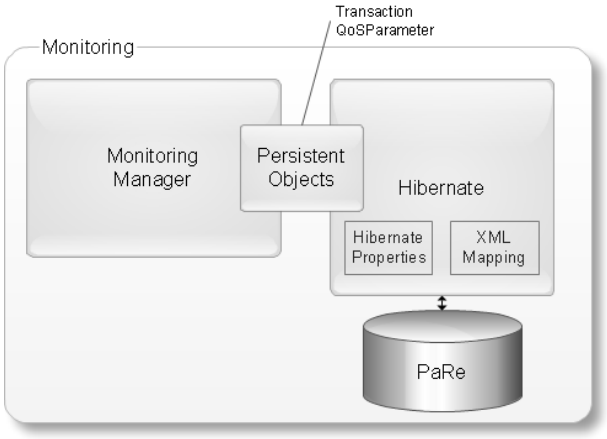


Fig. 3. Monitoring module.

3.4 Diagnosis Module

As we mentioned, monitoring module collects QoS data and stores them into PaRe (Parameter Repository). All data collected are sent to Diagnosis Module (referenced by DIMOD at the rest of paper) if any of the following scenarios occurred:

1. A Web service does not response.
2. Monitoring module collects one hundred transactions.

Based on SLP information, data collected by monitoring module are analyzed by DIMOD, in order to find out architecture behaviour indicating a degraded state; it is presented when DIMOD detects an amount of outliers. We used a statistic model as a main tool to know when application behaviour presents any of the defined states (Optimal, Good or Degraded). In order to define the model, a BoxPlot [15], [16] method was suggested; it is based on a graphic box, which represents a data distribution based mainly on three measures: inferior, median and superior quartiles. On a BoxPlot diagram outliers are very easy to identify being outside of the box. Table 1 shows the variables considered and used by BoxPlot, all of them are based on information collected by monitoring module.

Table 1. Variables considered by BoxPlot.

Variables	Description
N	Total of transactions processed.
x_i	External border for QoS_i, $x_i = Q_3 + 3(IQR)$.
m_i	Amount of outliers values, where $QoS_i >$
p_i	% of outliers values, $p_i = m_i * 100 / n$.

Previous table shows N as the total of transactions that diagnosis model analyzes. Every transaction has four QoS parameters (QoS1, QoS2, QoS3 y QoS4), used to calculate external borders.

Table 2. QoS level agreement.

Criteria	**pi = 0%**	**pi < 5%**	**pi > 5%**
QoS level	**Optimal**	**Good**	**Degraded**

According to data expressed on table 2, a very optimistic scenario is defined, when no outlier is registered, we do not expect this scenario all time. With the percentage of outliers values less than 5% and greater than 0% a good scenario is defined, our main goal is keep a QoS level between good and perfect values. Finally a degraded state is defined when that percentage is greater than 5%, in this case the recovering module will apply a reconfiguration strategy in order to guarantee the QoS agreement.

As mentioned previously, the monitoring module enables the diagnosis whenever one hundred transactions are completed or when an error occurs, and is necessary an immediate reconfiguration action.

The diagnosis process verifies the compliance with the criteria described above, calculating the percentage of outliers, from SLP values, for QoS parameters of the transaction whose identifier is in the range defined by startIndex and endIndex. While all the QoS parameters meet the limits defined by the SLP, the diagnosis module will not raise any recovery action and the range of transactions will increase, setting a new endIndex.

If a QoS parameter does not meet the required limit, then the diagnosis module alerts the recovery module to perform the appropriate reconfiguration action and defines a new range of diagnosis, ruling out the QoS parameters of transactions previously analyzed.

If the monitoring module records that a transaction was not completed successfully because a service provider is out of line, it notifies immediately to the diagnosis module, and to the recovery module, using the method activeSusbtitution. A substitution action is applied immediately to redirect the request to another provider.

3.5 Recovery Module

When diagnosis module detects a QoS degradation, recovery module is alerted in order to stabilize the application. When a WS is down, a new one replaces it immediately. This is considered a panic error scenario on our self–healing architecture.

We defined two recovery actions and their implementation will depend on the degree of degradation of the QoS, based on the percentage of outliers reported by the diagnosis module (see table 3).

Table 3. Recovery actions.

Degradation degree	**$5\% < p_i \leq 10\%$**	**$p_i > 10\%$**
Recovery action	Duplication	Substitution

Duplication actions divide all WS requests between the current WS and a new one that offers the same operation, in order to meet the service quality agreement. This action will keep active these two services in the WSTA instance.

The action of substitution replaces the current service for a new one. It also includes deletion of the previous service from the WSTA instance.

All service information is stored on WSTA table, containing more than one service, with similar operations. The recovery module actives a service, as a result of any reconfiguration actions described above. This module also uses Hibernate as a middleware to manage persistence information.

4 Results

In order to test our approach based on BoxPlot diagrams, we implemented a system with all the necessary components to monitor application performance, diagnose application's health and apply reconfiguration approaches into architecture. We worked with the PDLib digital library application [17]. We adapted this application in order to enable WS interfaces.

There are many services available on a digital library, we chose listDocumentation service, used to show all documents stored in a database, as a main service in our test. We launched 10,000 requests to the application, trying to obtain enough data to validate our diagnosis module implementation.

After execution of testing transactions, we got the results shown in the following three tables referenced as table 4. Values are expressed in milliseconds.

Table 4. Experimental results.

	Frequency	Mean	Median	Mode
QoS1	10000	27.9598	22.0	20.0
QoS2	10000	26.0621	20.0	17.0
QoS3	10000	7.3885	6.0	5.0
QoS4	10000	33.4531	26.0	26.0

	Standard deviation	Minimum	Maximum	Range
QoS1	28.9272	14.0	1061.0	1047.0
QoS2	32.2210	16.0	1154.0	1138.0
QoS3	11.9261	4.0	377.0	373.0
QoS4	33.5988	21.0	1160.0	1139.0

	First Quartile	Second Quartile	Third Quartile	Interquartile Range
QoS1	20.0	22.0	23.0	3.0
QoS2	17.0	20.0	21.0	4.0
QoS3	5.0	6.0	6.0	1.0
QoS4	23.0	26.0	27.0	4.0

Experimental results in table 4 show QoS limit values for each quartile. Based on third quartile measures (Q3) we calculated external borders, xi = Q3 + 3(IQR), for each QoS parameter. Table 5 shows extreme borders values.

Table 5. QoS external borders values.

	xi = Q3 + 3(IQR)	External Borders
QoS1	x1 = 23.0 + 3(3.0)	x1 = 32.0
QoS2	x2 = 21.0 + 3(4.0)	x2 = 33.0
QoS3	x3 = 6.0 + 3(1.0)	x3 = 9.0
QoS4	x4 = 27.0 + 3(4.0)	x4 = 39.0

Each value bigger than its external border is considered as an outlier. Based on table 5 outliers values for QoS1 will be those bigger than 32 ms, bigger than 33 for QoS2, and so on. All values greater than its corresponding external border are considered as unexpected values.

Following figure 4 is a BoxPlot diagram showing QoS results and we can see that most of the measurements are skewed at the right and near of the third quartile. Few measurements are high for the four QoS parameters having 9.46%, 9.15%, 7.67%, and 10.69% respectively of unexpected values. These values are stored in PaRe in order to define the acceptable limits.

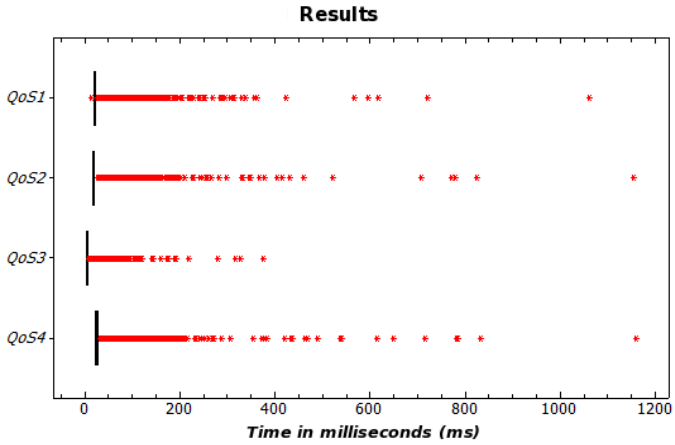


Fig. 4. BoxPlot diagram for QoS registers.

If an application does not accomplish with the QoS level agreement, diagnosis module will detect QoS degradation and recovery module will be alerted in order to stabilize the application. When a WS is down, immediately, is replaced by a new one. This is considered a panic error scenario on our self–healing architecture.

Another test measured the self-healing impact on application performance. So we repeat the previous test with (TestCase1) and without (TestCase2) self-healing architecture having that the values for each QoS parameters vary only in a few milliseconds, which means that the cost-benefit of using our architecture

is favourable as they have all the benefits of self-healing and the cost is relatively low. The biggest difference we can notice is in the parameter QoS1, with the increase of 1.0901 ms on Mean, because in this time period are carried out most of the additional processes required to ensure the QoS.

5 Conclusion

Each module in our self-healing architecture carries out specific tasks and has well-defined interfaces through it communicates with the other modules. The novel points of our diagnosis model are the inclusion of the QoS parameters and the construction of a statistic model based on BoxPlot diagrams. Based on that, we apply reconfiguration actions. Results showed how diagnosis module obtains right conclusions about system health, sending correct information to recovery module.

In order to measure performance and stability of our architecture, two test scenarios were designed by supporting functionality of a digital library application. The first scenario consisted of applying the architecture to the digital library system and calculating statistical measures from the results. The second scenario is similar to the first one, but the main difference is that the complete architecture was deactivated in order to compare the obtained results previously and to calculate the additional processing time that is generated when using our architecture.

The accomplishment of these tests threw favourable results, since when doing the comparison between the obtained results from both scenarios we could observe that when deploying our architecture it affected the application results in an insignificant way, and the benefits that can be obtained are very important. This means that there are compensation between self-healing features implemented against cost and performance.

As a future work, we consider extend our statistic model exploring other diagnosis approaches like the definition and implementation of semantic techniques based on ontologies. Furthermore, diagnosis module will determine precisely failures, and recovering module will be able to apply a better reconfiguration action.

With respect to the digital library application, the development of new interfaces is considered in order to provide more services related to this application.

Finally, developing test scenarios where architecture modules are lesser centralized is currently considered.

Acknowledgments. This project is developed under financial support from PROMEP and UADY.

References

1. Kephart, J., and Chess, D. (2003). The vision of autonomic computing. Computer, 1(36):41–50.
2. Ghosh, D., Sharman, R., Rao, H. R., & Upadhyaya, S. (2007). Self-healing systems - survey and synthesis. ACM Digital Library, 42, 2164-2185.

3. Ben Halima, R., Drira, K., and Jmaiel, M. (2008). A QoS-Oriented Reconfigurable Middleware for Self-Healing Web Services. IEEE International Conference on Web Services, ICWS '08. Pages 104-111, Beijing, China.
4. Garlan, D. and Schmerl, B. (2002). Model-based adaptation for self-healing systems. In Proceedings of the first workshop on self-healing systems, WOSS '02, pages 27–32, New York, NY, USA. ACM Press.
5. Wile, D. and Egyed, A., (2004). An externalized infrastructure for self-healing systems. In Proceedings of the Fourth Working IEEE/IFIP Conference on Software
6. Gurguis, S. A., Zeid, A., (2005). Towards autonomic web services: achieving self-healing using web services. In DEAS '05: Proceedings of the 2005 workshop on design and evolution of autonomic application software, pages 1–5, NY, USA, ACM Press.
7. Moo-Mena, F. J., Drira, K., (2007). Reconfiguration of web services architectures: a model-based approach. In Proceedings of the Twelfth IEEE Symposium on Computers and Communications (ISCC'07), pp. 357–362, Aveiro, Portugal. IEEE Computer Society.
8. Moo-Mena, F. J., Drira, K., (2007). Modeling architectural level repair in web services. In Proceedings of the 3rd International Conference on Web Information Systems and Technologies (WEBIST'2007), pages 240–245, Barcelona, Spain.
9. Association of Research Libraries (1995). Definition and purposes of a digital library. http://archive.ifla.org/documents/libraries/net/arl-dlib.txt
10. Wang, M., Suda, T. (2000). The bio-networking architecture: A biologically inspired approach to the design of scalable, adaptive, and survivable/available network applications. Tech. Rep. 00-03, University of California, Irvine, California, USA.
11. Eisenhauer, M., Rosengren, P., and Antolin, P. (2009). A Development Platform for Integrating Wireless Devices and Sensors into Ambient Intelligence Systems. 6th Annual IEEE Conference on Communications Society. Pages 1-3, Rome, Italy.
12. Ribeiro, G. and Karran, T., (2001). An Object-Oriented Organic Architecture for Next Generation Intelligent Reconfigurable Mobile Networks, doa, pp.0031, Third International Symposium on Distributed Objects and Applications (DOA'01), 2001.
13. Console, L., et al. (2008). At your service: Service Engineering in the Information Society Technologies Program - WS-DIAMOND: Web Services - DIAgnosability, MONitoring, and Diagnosis. Editor E. di Nitto, A-M. Sassen, P. Traverso, and A. Zwegers, MIT Press.
14. King, G,, Bauer, C., Rydahl-Andersen, M., Bernard, E., and Ebersole S. (2010). Hibernate Reference Documentation. Copyright © 2004 Red Hat, Inc.
15. Scheaffer, R.L. and McClave, J.T (1994). Probability and Statistics for Engineers. Duxbury Prees, 4th Edition.
16. Mendenhall, W., Beaver, R. J., and Beaver, B.M (2008). Introduction to Probability and Statistics. Barnes & Noble, 559 p.
17. Alvarez, F., García, R., Garza, D., Lavariega, J., Gómez, L., and Sordia, M., (2005). Universal access architecture for digital libraries. In Proceedings of the Conference of the Centre For Advanced Studies on Collaborative Research, pages 17–20, Richmond-Hill, Ontario, Canada.

Corpus de Voz en Español Mexicano para Experimentación en Reconocimiento Automático de Locutor

José-Martín Olguín-Espinoza[1] y Pedro Mayorga-Ortiz[2]

[1] Universidad Autónoma de Baja California, Mexicali, B.C., México
molguin@uabc.edu.mx

[2] Instituto Tecnológico de Mexicali, Mexicali, B.C., México
pedromayorga@hotmail.com

Paper received on 29/03/10 , Accepted on 20/09/10.

Resumen. Un elemento indispensable en la construcción de un sistema de Reconocimiento Automático de Locutor (RAL) es el corpus de voz, el cual es necesario para la experimentación y evaluación. Las condiciones que debe cumplir un corpus para RAL es que contenga grabaciones de frases fonéticamente equilibradas de varios locutores usando distintos medios de grabación y a lo largo de varias sesiones. En este trabajo se presenta la metodología y resultados de la construcción de un corpus de voz para RAL con locutores de español mexicano, así como la comparación con estudios previos.

Palabras Clave: Reconocimiento Automático de Locutor, Corpus de Voz**.**

1 Introducción

Debido a que la voz es el instrumento de comunicación preferida de los humanos, utilizar Reconocimiento Automático de Locutor (RAL) es de gran importancia [10]. Las modalidades del RAL se subdividen en dos tareas [9]: la primera consiste en determinar si el emisor es quien dice ser, esto se conoce como Verificación Automática de Locutor (VAL); la segunda consiste en determinar quién es el emisor, conocida como Identificación Automática de Locutor (IAL). Esto es, para la VAL el locutor provee la señal de voz e información adicional (como un número de identificación personal o PIN) y el sistema determina si la voz corresponde a ese locutor. En el caso de la IAL, sólo se tiene como entrada la señal de voz, siendo la salida del sistema de reconocimiento el identificador del locutor a quien corresponde esa señal. Recientemente Patil y Basu [16] añaden a estas tareas la Clasificación Automática de Locutor (CAL), la cual trata de agrupar a los locutores de acuerdo a una característica en común, como el acento o manera de pronunciar ciertos fonemas. Esta tarea es importante cuando se trata de identificar por ejemplo la zona geográfica de donde proviene cierto locutor (regionalismos en el habla), por lo que en este caso en

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 185 -194.

particular es indispensable contar con modelos específicos para cada región geográfica que se desea ubicar.

Para la construcción de los sistemas de RAL es necesario tener señales de voz de los locutores que serán reconocidos, esta señal se tiene que procesar con la finalidad de extraer los elementos que permitan caracterizar de manera única a cada locutor (o conjunto de locutores). Al conjunto de señales de voz de varios locutores se le denomina Corpus de Voz y para tener sistemas RAL robustos, es necesario que el corpus cumpla con varios aspectos entre los que destacan: a) Incluir todos los fonemas del lenguaje; y b) Preservar la distribución fonética del lenguaje [17].

En relación al español mexicano, la literatura encontrada es escasa y orientada al Reconocimiento Automático del Habla (RAH), por lo mismo los corpora encontrados han sido construidos con este fin [17],[11], razón por la cual la contribución principal del presente trabajo es la construcción de un corpus en español mexicano para experimentación con sistemas de RAL.

2 Corpus de Voz

Para propósitos del presente trabajo, los corpora de voz para experimentación se pueden clasificar en dos tipos: los orientados a RAH y los orientados a RAL. Los primeros deben permitir analizar las diferencias fonéticas del idioma independientemente del locutor del cual provenga la señal, para lograr esto es necesario grabar las mismas frases para una gran cantidad de locutores, no es tan importante el número de sesiones por locutor, lo que importa es caracterizar los fonemas para un amplio rango de locutores. Por otro lado, los corpora para RAL deben ser multisesión para que el sistema pueda discriminar los elementos de variabilidad intralocutor a lo largo del tiempo y resaltar la variabilidad interlocutor para poder caracterizar a las personas. Esto impone una mayor complejidad en la construcción de este tipo de corpora porque es necesario llevar un seguimiento del locutor a lo largo de la etapa de grabación, debido al tiempo que debe transcurrir entre una sesión y otra, el cual puede ser de semanas a meses.

2.1 Corpora para RAL en Idiomas Distintos al Español

Polycost. Frases en inglés de 134 locutores de 13 países europeos (de habla inglesa y habla no inglesa), más de 5 sesiones por cada locutor grabadas por medio de aparatos telefónicos a través de líneas digitales ISDN [8].

YOHO. Diseñado para evaluar sistemas de verificación de locutor en situaciones dependientes del texto. Consiste de 138 locutores (106 hombres, 32 mujeres) que grabaron frases compuestas de dígitos en idioma inglés [2].

Switchboard I-II. Un corpus en inglés con más de 500 locutores, las grabaciones fueron realizadas mediante un sistema telefónico automático que conectaba a un participante con otro y grababa la conversación. Un subconjunto de este corpus es utilizado en las evaluaciones de reconocimiento de locutor que regularmente organiza el Instituto Nacional de Estándares y Tecnología de Estados Unidos (NIST) [18].

SIVA. Base de señales en italiano, contiene 691 locutores (335 hombres, 336 mujeres), grabado en múltiples sesiones utilizando líneas telefónicas [3,8].

XM2VTSDB. Corpus multimodal que incluye cerca de 300 locutores ingleses en grabaciones de audio y video a o largo de cuatro sesiones. Es un producto en constante evolución cuyos orígenes fueron M2VTS en francés y XM2VTS en inglés [13].

2.2 Corpora para RAL en Español

AHUMADA. Corpus en español ibérico, consiste de 28 locutores hombres. Fue desarrollado en el contexto de un proyecto de investigación forense [14]. En lo que respecta al alcance del presente trabajo de investigación, este es el único corpus encontrado para RAL, que ha sido desarrollado en un país de habla hispana.

Otro corpus en español ibérico encontrado en la literatura es el denominado *Albayzin*, pero está orientado al reconocimiento del habla [4,21]

2.3 Corpora de Voz en Español Mexicano

Hasta donde tienen conocimiento los autores, no existe un corpus para RAL que haya sido desarrollado en México, los siguientes corpora encontrados están orientados a reconocimiento del habla:

DIMEX-100. Desarrollado en el Instituto de Investigaciones en Matemáticas Aplicadas de la Universidad Nacional Autónoma de México (IIMAS-UNAM). Es un corpus en español mexicano orientado al reconocimiento del habla, incluye grabaciones de 100 locutores del centro del México, incluye un estudio detallado de los fonemas y sus correspondientes alófonos para el español mexicano [17].

TLATOA. Desarrollado por la Universidad de las Américas en Puebla (UDLA). Consiste de grabaciones de 550 adultos pricipalmente del centro de México, está orientado a experimentación con reconocimiento del habla [11].

3. Distribución Fonética del Español Mexicano

En [1,6] se establece la ventaja de asignarles peso a los fonemas para mejorar el reconocimiento, por lo tanto es importante que un corpus de voz contenga al menos los fonemas más utilizados en el lenguaje nativo de los locutores. Por esta razón, el primer paso para la construcción del corpus es la definición de una o más frases que los locutores deben grabar. Sin embargo, para definir estas frases es necesario conocer la distribución fonética del lenguaje y en base a ésta construir las frases fonéticamente equilibradas, tal y como se reporta en la construcción del corpus AHUMADA [14]. Pérez [15] presenta la distribución fonética del español latinoamericano obtenida mediante el análisis de grabaciones de audio correspondientes a noticieros chilenos, Villaseñor-Pineda [20] la obtuvo a partir de fuentes de texto de páginas web, ambos trabajos concuerdan en que la diferencia entre el español ibérico y el latinoamericano (chileno y mexicano) es la frecuencia de uso de los fonemas /a/ y /e/; en el español ibérico se utiliza más el fonema /a/ que el fonema /e/ y en el

español latinoamericano es al contrario, tiene una mayor frecuencia de uso el fonema /e/ que el fonema /a/.

3.1 Recolección de Datos

Para este trabajo se utilizó un enfoque similar al presentado por [20], esto es, se utilizaron páginas web como única fuente de información. La principal diferencia consistió en el tipo de fuentes de datos utilizadas y el volumen total de información recolectada (se requirió alrededor de 90% menos datos). Las páginas web fueron seleccionadas de acuerdo a los siguientes criterios:

1. Páginas web en español de periódicos mexicanos de diferentes áreas geográficas. Para esto se utilizó la distribución de zonas propuesta por una editorial de periódicos mexicana cuyo portal de Internet publica los diarios que tiene a lo largo de todo el país (7 zonas: Norte, Golfo, Noroeste, Centro, Bajío, Sur, Este).
2. Con la finalidad de capturar las palabras utilizadas específicamente en esa zonas geográficas (regionalismos), sólo se utilizaron las páginas correspondientes a las noticias locales, las cuales se asume fueron escritas por reporteros locales.

En total se recabó información de 39 periódicos en línea , durante un lapso de 20 días, acumulando un total de 157MB de texto. Se generaron 6 bases de datos, una por región, en cada una se vació la información recabada separando las palabras y su correspondiente número de ocurrencias. La información fue almacenada por separado para poder hacer un análisis por áreas geográficas y así poder identificar diferencias en el uso de los fonemas entre zonas.

3.2 Análisis de Datos y Resultados

Los fonemas analizados corresponden a la definición básica en [15], la cual consiste de 22 fonemas básicos para el español y sus correspondientes alófonos, aunque para este trabajo no se consideran estos últimos. Se aplicó un algoritmo de detección de fonemas a la información recabada en cada una de las zonas, los resultados obtenidos no mostraron diferencias significativas de la frecuencia de los fonemas entre las diversas zonas. Al final se agregó toda la información y se obtuvieron las frecuencias finales, en la Tabla 1 se muestran los porcentajes finales.

3.3 Validación con Trabajos Previos

Los resultados obtenidos en la distribución fonética muestran el mismo comportamiento reportado en [15,20] para el español latinoamericano, el análisis de correlación entre lo encontrado y esos trabajos se muestra en la Tabla 2.

Tabla 1. Frecuencia encontrada para cada uno de los Fonemas.

Fonema	Frecuencia	Fonema	Frecuencia	Fonema	Frecuencia
/i/	7.63%	/e/	13.85%	/a/	12.69%
/o/	9.37%	/u/	2.96%	/p/	2.81%
/t/	4.60%	/k/	4.07%	/b/	2.05%
/d/	5.41%	/g/	0.44%	/f/	0.78%
/s/	9.66%	/j/	0.87%	/ch/	0.18%
/ll/	0.45%	/m/	2.66%	/n/	7.21%
/ñ/	0.16%	/r/	6.59%	/rr/	0.18%
/l/	5.37%				

Tabla 2. Coeficiente de correlación con trabajos previos.

	Coeficiente de correlación
Villaseñor-Pineda (2003)	0.9960
Pérez (2003)	0.9974

4 Desarrollo del Corpus

Una vez establecida la distribución fonética, se diseñaron las frases fonéticamente equilibradas y se realizaron las grabaciones.

4.1 Protocolo de Grabación

Una vez establecida la distribución fonética se definió el protocolo de grabación, el cual incluyó las siguientes características:

1. Por cada locutor, grabar 3 frases fonéticamente equilibradas (4, 6 y 10) segundos respectivamente y un texto adicional no balanceado (aprox. 60 segundos).
2. Realizar tres sesiones por locutor.
3. Utilizar dos tipos de dispositivo de adquisición: micrófono y aparato telefónico.
4. Tres modalidades: sistema telefónico analógico, grabación directa de micrófono y sistema telefónico de VoIP.

4.2. Software y Hardware utilizado

Para las grabaciones telefónicas se configuró el software Asterisk, el cual es un servidor de comunicaciones de voz para líneas telefónicas analógicas y líneas de VoIP, bajo el sistema operativo Linux. El hardware utilizado fue una computadora

con procesador Pentium 4 de 3GHz de velocidad, 2GB de RAM, una tarjeta de puertos telefónicos Digium modelo 1TDM422EF. Las grabaciones de VoIP se realizaron a través de una red local de datos inalámbrica 803.11g utilizando el softphone Zoiper y un micrófono tipo cardiod. Para las grabaciones analógicas y las de VoIP se configuró un sistema de contestación automática en el servidor Asterisk, el cual consistió de un menú interactivo para establecer el número de locutor, tipo de medio, número de frase y número de sesión. Todas las grabaciones telefónicas se almacenaron en el sistema de archivos del servidor Linux. Para las grabaciones directas se usó un micrófono de escritorio tipo cardiod conectado por puerto USB a una computadora Pentium 4 de 3GHz, 2GBde RAM con sistema operativo Windows XP SP3 y el software Audacity para procesar las señales de audio.

4.3 Sesiones de Grabación

Todas las grabaciones se realizaron en un cuarto especialmente acondicionado para tal fin, con dimensiones de 3x3 mts. y las paredes cubiertas con espuma fonoabsortora de poliuretando con cuñas anecoicas para disminuir la reverberación acústica. Las sesiones se llevaron a cabo a lo largo de 18 meses, participaron 50 personas, 31 hombres y 19 mujeres. 23 hombres completaron dos sesiones y 18 completaron las tres sesiones, en el caso de las mujeres, 14 completaron dos sesiones y 13 completaron las 3 sesiones. En total el corpus contiene 31 locutores con las tres sesiones completas. El tiempo transcurrido entre sesiones varía de 1 a 5 semanas. El tiempo promedio que cada locutor le dedicó por sesión fue de 20 minutos.

4.4 Organización del Corpus

El corpus está organizado por directorios, uno para cada locutor, el nombre del directorio es el identificador del locutor e.g. L01 indica al locutor 01. Cada archivo del corpus se nombró utilizando la siguiente nomenclatura:

L<ID de locutor><Tipo de Grabación>F<ID de Frase>S<ID de Sesión>

<ID de locutor>: Número consecutivo asignado a cada locutor con fines de seguimiento e identificación en el corpus.

<Tipo de Grabacion >: Identificador del tipo de adquisición de la señal, MI para micrófono, T1 para teléfono analógico, T2 para VoIP.

<ID de Frase >: Identificador de la frase, puede tomar los valores de 1, 2 , 3 ó 4.

<ID de Sesión >: Identificador de la sesión en que fue grabada esa señal, puede ser 1, 2 ó 3.

Por ejemplo, el archivo *L01MIF2S3* corresponde a la grabación directa de micrófono del locutor 01, frase 2, sesión 3.

4.5 Experiencias Obtenidas durante la Grabación

Durante la etapa de grabación se identificaron dos aspectos a considerar en los proyectos de este tipo:

1. Cuando se construyeron las frases fonéticamente equilibradas, no se consideró importante que fueran frases coherentes, lo único que se buscaba es que cumplieran con la distribución de frecuencia de fonemas, se trató de seguir la idea de las frases de XM2VTSDB [13], las cuales son una secuencia de palabras sin significado. Sin embargo los locutores tendían a confundirse fácilmente en las grabaciones, por lo que se optó por cambiar las frases por otras más coherentes, esto disminuyó el número de errores que el locutor cometió por sesión y por lo tanto disminuyó el tiempo promedio por sesión.
2. Dado que los participantes no tienen claro el objetivo de las grabaciones, es necesario estar constantemente explicando los alcances de la investigación, ya que sienten cierta incomodidad al no tener certeza del uso que se les dará a sus datos. La certeza es indispensable para conseguir de ellos las sesiones subsecuentes.

5 Evaluación del Corpus

Para la evaluación del corpus se implementó un sistema de IAL basado en Modelos Mezclados Gaussianos (GMM) [19], utilizando las herramientas open-source utilizadas en [7], la arquitectura se muestra en la Fig. 1.

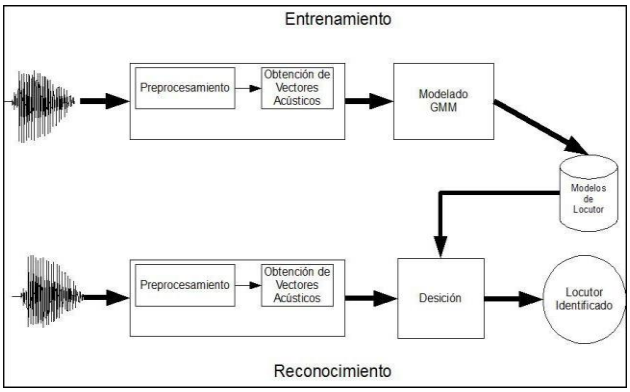


Fig. 1. Arquitectura del Sistema de IAL utilizado para la evaluación del Corpus.

Para la construcción de vectores acústicos se utilizó una ventana de 30 ms con un traslape de 10 ms, se aplicó un banco de filtros de 26 canales para obtener un vector de 41 elementos (13 coeficientes MFCC, 13 valores de la primera derivada, 13 de la segunda derivada y 2 coeficientes de energía) además se aplicó sustracción de medias cepstrales (CMS) para atenuar el ruido del medio. Para el modelado de locutores primero se construyó un modelo impostor con 32 mezclados gaussianos con matriz de covarianza diagonal, el entrenamiento se llevó a cabo con el algoritmo Expectación-Maximización (EM); cada modelo cliente se obtuvo mediante adaptación máxima a posteriori (MAP).

5.1 Evaluación

El corpus se particionó en dos conjuntos, uno de entrenamiento y otro para prueba, quedando de la siguiente forma:

1. Frases 1, 2 y 3 de las sesiones 1 y 2 (aprox. 40 segundos de señal).
2. Frase 4 de la sesión 3 (aprox. 60 segundos de señal).

La evaluación del corpus se realizó sólo con las señales de los locutores hombres, con la finalidad de poder comparar los resultados con otros corpora disponibles, los cuales sólo contienen señales de locutores masculinos.

El resultado de las pruebas se muestra en la Tabla 3. El mejor reconocimiento se obtuvo con las grabaciones de micrófono, tal y como se esperaba ya que este tipo de datos son los que mayor calidad presentan en términos del medio de adquisición, en segundo lugar los de VoIP los cuales también fueron adquiridos por medio del mismo micrófono. Por otro lado, el menor reconocimiento corresponde a los datos adquiridos con un aparato telefónico, como era de esperarse, ya que este tipo de medio es el que tiene menor calidad de captura.

Tabla 3. Resultados de la evaluación del corpus.

	Micrófono	Teléfono	VoIP
% Reconocimiento	95.24	71.43	90.48
% Identificación errónea	0.00	0.00	0.00
% Falsos Positivos	0.00	28.57	9.42
% Falsos Negativos	4.76	0.00	0.00
TOTAL	100.00	100.00	100.00

5.2. Comparación con AHUMADA

Además de la evaluación anteriormente presentada, se realizó otra con el mismo sistema ASR pero utilizando los datos del corpus en español ibérico AHUMADA para micrófono y teléfono ya que este corpus no contiene señales de VoIP. Las particiones del conjunto de datos para AHUMADA se eligieron de manera semejante a las utilizadas con nuestro corpus:

1. Diez frases fonéticamente equilibradas grabadas en dos sesiones distintas (Aprox. 40s de señal).
2. Lectura de un texto de una tercera sesión, un texto distinto al usado en el entrenamiento (Aprox. 60s de señal). De la misma manera que con nuestro corpus, sólo se utilizaron señales de locutores masculinos. La comparación de resultados se presenta en la Tabla 4. Se observa que para el caso de la señal de micrófono, el reconocimiento de los datos de AHUMADA fue ligeramente mayor, mientras que nuestra señal telefónica obtuvo un mejor reconocimiento.

Una posible explicación para la gran diferencia en el reconocimiento de la señal telefónica es que las condiciones de grabación en nuestro caso fueron más cuidadas

que lo reportado por [14] y otra posibilidad radica en la diferencia de calidad del aparato telefónico.

Tabla 4. Comparación de resultados de reconocimiento usando el corpus AHUMADA.

	Micrófono	Micrófono AHUMADA	Teléfono	Teléfono AHUMADA
% Reconocimiento	95.24	100.00	71.43	47.62
% Identificación errónea	0.00	0.00	0.00	4.76
% Falsos Positivos	0.00	0.00	28.57	33.33
% Falsos Negativos	4.76	0.00	0.00	14.29
TOTAL	100.00	100.00	100.00	100.00

6 Conclusiones

En el presente trabajo se construye un corpus de voz en español mexicano orientado al reconocimiento de locutor, es importante contar con una base de señales de este tipo ya que algunas tareas de reconocimiento, especialmente la clasificación, requieren de señales de voz que contengan características fonéticas específicas de los habitantes de cierta ubicación geográfica. También se muestra que es posible obtener la distribución fonética del español mexicano analizando texto de páginas web de periódicos locales de varias regiones de México. Se comprobó que la distribución fonética no varía de región en región. Por otro lado, se establece la importancia de que las frases que se vayan a grabar tengan cierto grado de coherencia, aún y cuando expresen ideas incompletas, ya que los locutores tienden a equivocarse menos cuando las frases a pronunciar son menos complejas.

El corpus construido permitirá continuar con investigaciones en reconocimiento y clasificación de locutores de la zona noroeste de México.

Referencias

1. Auckenthaler, R., Parris, E. S., Carey, M. J.: Improving a GMM speaker verification system by phonetic weighting. In Proceedings of the Acoustics, Speech, and Signal Processing, 1999. Volume 01 March 15 - 19, (1999).
2. Campbell, J.P. Jr.: Testing with The YOHO CD-ROM voice verification corpus. In Proceedings of the international conference on acoustics, speech and signal processing, ICASSP'95, Detroit, 341-344 (1995)
3. Campbell, J.P. Jr., Reynolds, D.A.: Corpora for the evaluation of speaker recognition systems. In Proceedings of international conference on acoustics, speech and signal processing, ICASSP'99. 2, 829-832 (1999).
4. Casacuberta, F., García, R., Llisterri, J., Nadeu, C., Pardo, J.M., Rubio, A.: Desarrollo de Corpus para Investigación en Tecnologías del Habla (ALBAYZIN). Procesamiento del Lenguaje Natural. 12. 35-42 (1992)
5. Faúndez-Zanuy, M.: State-of-the-art in speaker recognition. IEEE A&E Systems Magazine, 20(5). 7-12 (2005)

6. Faltlhauser, R., Ruske, G.: Improving Speaker Recognition Performance Using Phonetically Structured Gaussian Mixture Models. In EUROSPEECH-2001, 751-754 (2001)
7. Fauve, B., Matrouf, D., Scheffer, N., Bonastre, J., Mason, J.: State-of-the-Art Performance in Text-Independent Speaker Verification Through Open-Source Software. IEEE Transactions on Audio, Speech, and Language Processing. 15 (7). 1960-1968 (2007)
8. Hennebert, J., Melin, H., Petrovska, D., Genoud, D.: Polycost: A Telephone-Speech Database For Speaker Recognition. Speech Communication. 31 (2-3), 265-270 (2000)
9. Juang, B.H., Chen, T.: The past, present, and future of speech processing. Signal Processing Magazine, IEEE. 15(3), 24-48 (1998)
10. Keshet, J., Bengio, S.: Automatic Speech and Speaker Recognition: Large Margin and Kernel Methods. Wiley. (2009)
11. Kirschning, I.: TLATOA: Developing Speech Technology & Applications for Mexican Spanish. In 2nd. Intl.Workshop on Spanish Language Processing and Language Technologies. September 14-15, 2001, Jaén, Spain. (2001)
12. Li, Q., Reynolds, D.A.: Corpora for the evaluation of speaker recognition systems. In Proceedings of the Acoustics, Speech, and Signal Processing, ICASSP'99, 829-832. March 15-19, (1999)
13. Messer, K., Matas, J., Kittler J., Jonsson, K.: XM2VTSDB: The Extended M2VTS Database. In Proceedings of the 2nd International Conference on Audio and Video Based Biometric Person Authentication (AVBPA'99). March 22-24, 1999. Washington DC, USA. (1999)
14. Ortega-García, J., González-Rodríguez, J., Marrero, V., Díaz-Gómez, J., García-Jiménez, R., Lucena-Molina, J., Sánchez-Molero, J.: AHUMADA: A Large Speech Corpus in Spanish for Speaker Identification and Verification. Speech Communication. 3. 255-264. (2000)
15. Pérez. H. E.: Frecuencia de fonemas. Revista Electrónica de la Red Temática en Tecnologías del Habla, 1, Marzo 2003. http://lorien.die.upm.es/~lapiz/e-rthabla/numeros/N1/N1_A4.pdf (2003)
16. Patil, H. A., Basu, T.K.: Development of speech corpora for speaker recognition research and evaluation in Indian languages. International Journal of Speech Technology. 11(1). 17-32 (2008)
17. Pineda ,A. L., Castellanos, Cuétara, J., Galescu, L., Juárez, J., Llisterri, J., Pérez, P., Villaseñor, L.: The Corpus DIMEx100: Transcription and Evaluation. Language Resources and Evaluation (DOI: 10.1007/s10579-009-9109-9) . (2009)
18. Przybocki, M. A., Martin, A. F.: NIST Speaker Recognition Evaluation Chronicles. In Procedings of The Speaker and Language Recognition Workshop Odyssey 2004. http://www.nist.gov/speech/publications/papersrc/ody2004NIST-v1.pdf
19. Reynolds, D.A., Rose, R.C.: Robust text-independent speaker identification using Gaussian mixture speaker model. IEEE Transactions on Speech and Audio Processing. 3 (1). 72-83 (1995)
20. Villaseñor-Pineda, L., Montes-y-Gómez, M., Vaufreydaz, D., Serignat J.: Elaboración de un Corpus Balanceado para el Cálculo de Modelos Acústicos usando la Web. XII Congreso Internacional de Computación CIC-2003. ISBN:970-36-0098-0. pp 198-200. Mexico City, México. (2003)
21. Zamalloa, M., Bordel, G., Rodríguez, L.J., Peñagarikano, M., Uribe, J.P.: Selección y Pesado de Parámetros Acústicos Mediante Algoritmos Genéticos para el Reconocimiento del Locutor. In Proceedings of IV Jornadas en Tecnologías del Habla. 349-354. November 8-10, 2006. Zaragoza, Spain. (2006) Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195--197 (1981)

DWT Foveation-Based Multiresolution Compression Algorithm

J. C. Galan-Hernandez, V. Alarcon-Aquino, O. Starostenko, J. M. Ramirez-Cortes[1]

Department of Computing, Electronics, and Mechatronics
Universidad de las Americas Puebla
Sta. Catarina Martir, Cholula, Puebla. C.P. 72810. MEXICO
Email: {juan.galanhz, vicente.alarcon}@udlap.mx
[1]Department of Electronics
Instituto Nacional de Astrofisica, Optica y Electronica
Tonantzintla, Puebla MEXICO
Paper received on 06/08/2010, Accepted on 20/09/2010.206

Abstract – Discrete Wavelet Transform (DWT) foveated compression can be used in real-time video processing frameworks for reducing the communication overhead. Such algorithms lead into high rate compression results due to the fact that the information loss is isolated outside a region of interest (ROI). The fovea compression can also be applied to other classic transforms such as the commonly used discrete cosine transform (DCT). An analysis has then been performed showing different error and compression rates for the DWT-based and the DCT-based foveated compression algorithms. Simulation results show that with foveated compression high ratio of compression can be achieved while keeping high quality over the designed ROI.

Keywords — Foveation, wavelets, wavelet transforms, discrete cosine transform, compression.

1. INTRODUCTION

Digital images and video are essentially multi- dimensional signals and are thus quite data intensive [1]. Video and image compression can help in reducing the communication overhead between computing nodes and with wavelet compression, high ratios of compression can be achieved [2]. However, such algorithms also cause loss of information on video frames. In applications where a region of interest (ROI) can be isolated, foveation can be used in order to constraint the information loss only on those areas outside of the ROI. With such model, several applications can classify an area as a ROI such as medical video processing framework searching for melanomas, and perform a lossy compression over the video frames, leaving the ROI intact for later processing. Previous works in [2, 3, 4] show different foveation methods assessing the final quality of the image against the original image using

J. F. Ramírez Cruz, J. H. Sossa Azuela. (Eds.).
Advances in Intelligent and Information Technologies.
Research in Computing Science 30, 2010, pp. 197.-206

Human Vision System criteria. Foveation also guarantee certain quality from the human system vision perspective granting the output results with enough quality for a user to inspect the ROI without noticing the loss of quality unless a close inspection outside the ROI. In the work reported in this paper, an analysis of DWT-based foveated compression algorithms is carried out. The remainder of this paper is organized as follows. In Section II an overview of foveated compression is given. Section III describes the proposed approach. Section IV presents simulation results, and Section V reports conclusions and future work.

2. FOVEATED COMPRESSION

Wavelet transforms involve representing a general function in terms of simple, fixed building blocks at different scales and positions. These building blocks are generated from a single fixed function called mother wavelet by translation and dilation operations. Thus, wavelet transforms are capable of "zooming-in" on short-lived high frequency phenomena, and "zooming-out" on long-lived low frequency phenomena [5].

A. Wavelets and the Discrete Wavelet Transform

The purpose of wavelet transforms is to represent a signal into the frequency-time domain. To perform this task two functions are required, namely, a wavelet and a scaling function. If a set of mother wavelets and scaling functions is orthonormal it is called an orthonormal bases and is defined as follows [6]

$$\{\varphi_{l_0,n}\}_{0\le n\le 2^{l_0}} \cup \{\psi_{j,n}\}_{j<l_0,0<n<} \quad (1)$$

where each ψ is a translated copy of at scale j:

$$\psi_{j,n}(t) = \sqrt{2^{-j}}\psi(2^{-j}t - \quad (2)$$

and each φ is a translated copy of the scaling function at scale :

$$\varphi_{l_0,n}(t) = \sqrt{2^{-j}}\varphi(2^{-j}t - \quad (3)$$

In two dimensions three mother wavelets are used [6], $\Psi^v_{j,}$, $\Psi^h_{j,}$ and $\Psi^d_{j,m,n}(x$ defined as follows

$$\Psi^d_{j,m,n}(x,y) = \psi_{j,m}(x)\psi_{j,n} \quad (4)$$

$$\Psi^v_{j,m,n}(x,y) = \psi_{j,m}(x)\varphi_{j,n} \quad (5)$$

$$\Psi^h_{j,m,n}(x,y) = \varphi_{j,m}(x)\psi_{j,n} \quad (6)$$

with the scaling function:

$$\Phi_{j,m,n}(x,y) = \varphi_{j,m}(x)\varphi_{j,n} \tag{7}$$

where $\Psi^{d}_{j,m,n}(x$ are the diagonal coefficients, $\Psi^{v}_{j,m,n}(x$ are the vertical coefficients and $\Psi^{h}_{j,m,n}(x$ are the horizontal coefficients. With the wavelet base defined, the next step is using it to represent a signal. The sum over all time of the signal multiplied by scaled, shifted versions of the mother wavelet is given by:

$$a_j(n) = \int_{-\infty}^{\infty} f(t)\psi_{j,n}(t \tag{8}$$

where is the signal to be represented. However, for compression this transform is not suitable because it expands the signal into more coefficients than the samples of the signal itself. A better transform, suited for compression and many other applications, is called the discrete wavelet transform (DWT). In [7], the DWT is calculated through a simple algorithm that applies two filters, a low pass filter and a high pass filter. This algorithm is known as the fast wavelet transform [7]. In wavelet analysis of a signal f, we often speak of approximations and details. The approximations are the low-frequency components of the signal, see (9); whereas the details are the high-frequency components of the signal, see (10).

$$a_j[n] = \langle f, \varphi \tag{9}$$

$$d_j[n] = \langle f, \psi \tag{10}$$

B. Discrete Cosine Transform

The discrete cosine transform (DCT) expresses a signal in terms of cosine functions. Such transform is commonly used in the JPEG compression algorithm, the MP3 audio format and the VP8 video format [8]. The discrete cosine transform for a signal $f($ of length is defined as follows [9]

$$C(u) = \alpha(u)\sum_{x=0}^{N-1} f(x)\cos\left[\frac{\pi(2x+1}{2N}\right. \tag{11}$$

where $\alpha($ is defined by:

$$\alpha(n) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for} \quad u = \\ \sqrt{\frac{2}{N}} & \text{for} \quad u = \end{cases} \tag{12}$$

In particular, $u($ is known as the direct current coefficient (DC) and the remaining coefficients are called the alternating current coefficients [10]. Most of the energy of the signal is packed in the DC coefficient.

C. Wavelet Compression

The objective of data compression is to represent a set of data with less information than the original. There are two types of compression: lossy and lossless compression [10]. In lossy, some of the original data is discarded in order to achieve its goal. When the data is reconstructed it will be slightly different from the original. Using lossy compression can help to achieve a better compression ratio than lossless compression if the losses are acceptable over the result. This is a usual practice on images, such as jpeg and jpeg2000 formats [7, 10], and video such as mpeg4 format. When a wavelet transform is applied on an image, the resultant coefficients can then be compressed more easily because the information is statistically concentrated in just a few coefficients. Wavelet compression can reach higher compression ratio than other transforms such as the discrete cosine transform suggested for foveated compression in [3]. In Wavelet lossy compression, the coefficients that contain the most amount of energy are preserved and the rest are discarded. Selecting such coefficients can be done using the wavelet energy profile and choosing a cutoff frequency.

D. Foveation

Foveated images are images which have a non-uniform resolution [6]. Results reported in [11] have demonstrated that the human eye experiments a form of aliasing from the fixation point to the edges of the image. Such aliasing increases in a logarithmic rate on all directions. This can be seen as concentric cutoff frequencies from the fixation point. When it is used in a wavelet, this can be expressed as a function [2]:

$$I_0(x) = \int I(t) C^{-1}(x) s\left(\frac{t-x}{w(x)}\right) \tag{13}$$

where I is a given image, $I_0($ is the foveated image, $w($ is the weight function. The function is called the weighted translation of by . The function is defined as:

$$C(x) = \left\| s\left(\frac{-x}{w(x}\right) \right. \tag{14}$$

The results on a foveated image from [6] are shown in Figure 1. There are several weighted translation functions such as the ones defined in [6]. In [3], the suggested weighted functions are the Hamming window (Figure 2a) and the triangular window (Figure 2b). Such windows offer a smooth degradation from the fixation point. However, in order to preserve a ROI intact, such windows are not useful. A ROI needs to be left with all its coefficients without cutoff. For well defined ROIs, it is proposed to use a Tukey window (Figure 2c) Such windows can be used to define a weighted function where the radio of a fixation point is bigger than one, leaving the coefficients from the ROI untouched and right after the ROI ends the energy begins to decay in a smooth ratio.

3. PROPOSED DWT-BASED FOVEATED ALGORITHM

An implementation of DWT foveated compression was made in order to compare the quality and compression rate of a DWT-based foveated compression using a Tukey window over different wavelets. For compression quantization three different schemes were used, namely, an energy frequency profile cutoff scheme, a Tukey window-based scaling and a JPEG Type 1 quantization based on the Tukey window [8]. Figure 3 shows a foveated image with the different schemes with the fovea point in the eye of the bird and a radius of 60 pixels and four levels of wavelet decomposition using a Daubechies 7 wavelet. Other wavelets may also be considered. For the energy frequency profile cutoff, the algorithm is defined as follows: Given a fixation point, the ROI centroid, and a radius, the algorithm computes the foveated image I of dimension M x N with J levels of decomposition showed in Algorithm 1. The two-dimensional Tukey window, T, is calculated through interpolation of the one-dimensional Tukey window and the function $x^2 +$. Note that the interpolation method used was linear interpolation [13]. The set is a decent ordered set of all the coefficients of the wavelet discrete transform. is the total energy of the coefficients. The set is the wavelet energy profile. The set $Cuto$ is the fovea filter and contains all the thresholds for each pixel of the image. $Cutoj$ is the fovea filter weighted for each level of decomposition. The filter is not applied to the approximation DWT coefficients . This creates a coarser effect than in [3] or in [6]. However, with a well defined ROI, this method leads into a better compression ratio with a slightly less quality outside the ROI. For the Tukey window-based scaling, the cutoff is replaced for the next formula using a scaled $T2$ window:

(a) Original gray level image (b) Foveation point at the right eye

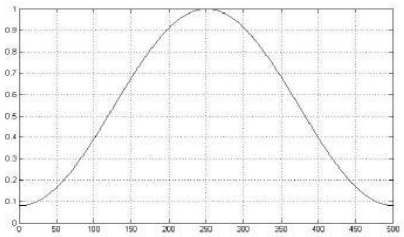
Fig. 1. An image and its wavelet foveated compression

$$d'^{\alpha}_{j}[m,n] := d^{\alpha}_{j}[m,n] * T2D_{j}[m, \quad (15)$$

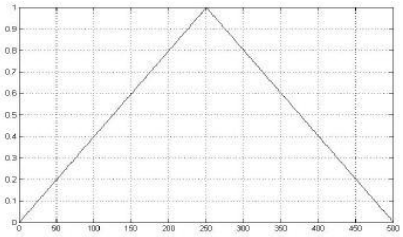
where $\alpha \in \{h, v,$ $T2$ is the T window resized by . In this scheme, the quantization is also applied to the approximation coefficients

$$a'_K[m,n] := a_K[m,n] * T2D_K[m, \quad (16)$$

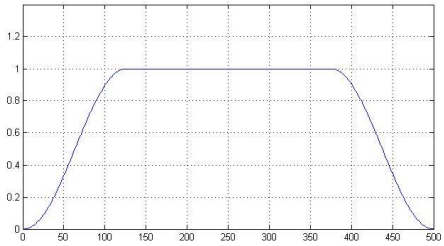
where $K = max($. A gray scale image "lena" of dimensions 512x512 was used for assessing the performance of the proposed algorithm. The original gray scale image is shown in Figure 3a.



(a) Hamming window



(b) Triangular window



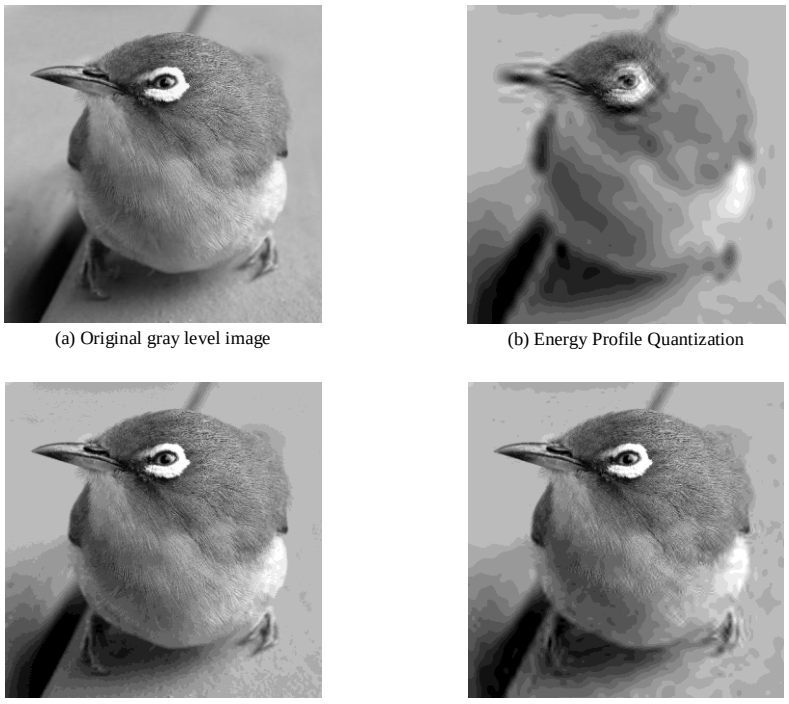
(c) Tukey window

Fig. 2. Different foveating windows

4. SIMULATION RESULTS

Comparisons were carried out with the wavelets Daubechies 7 (db7) and Daubechies 9 (db9) suggested by the JPEG2000 standard [12] with two arbitrary levels of decomposition, J = and J =. To assess the performance of the wavelet-based foveated algorithm the mean squared error (MSE) metric is used. Also, an arbitrary fovea radius used was of 60 pixels. The results are shown in Table 1. The DCT-based foveated compression was also realized and it is shown in Table 1 as dct. The compression was realized using the Type 3 JPEG variable quantization compression [8] and the Tukey window as the quantization weight. Table 1 shows how many coefficients becomes zero after the quantification function is applied. Figure 4 shows the DCT-based foveated image and a wavelet-based foveated image using the Daubechies 9 wavelet

with four levels of decomposition with energy profile quantization, Tukey window-based scaling quantization and JPEG Type-1 quantization. It should be noted in Table 1 that the amount of zeros in the coefficients after applying wavelet-based foveated algorithm using energy profile and JPEG Type-1 quantizations schemes is lower than the DCT-based approach. Furthermore, simulation results show that the DCT-based algorithm created artifacts that make the image fuzzier than the wavelet-based algorithm. The performance of the algorithm depends on the quantization method used. Energy quantization gives a good compression ratio at expenses of computational speed, it is expected a complexity of n^2 because of the wavelet coefficient sorting and the calculation of the cutoff window. Scaling quantization is faster with an expected complexity of n but has a low compression ratio.



(a) Original gray level image (b) Energy Profile Quantization

(c) Tukey window scaling Quantization (d) JPEG Type 1Quantization

Fig. 3. A Wavelet-based foveated image with different quantizations.

5. CONCLUSION

Wavelet foveation compression offers a very good compression ratio at expenses of controlled losses. However, applying foveation with wavelets yields into squared artifacts as mentioned in [3]. These artifacts increase as the decomposition levels increases. However, the DCT also showed a similar behavior in the area outside the ROI as the compression rate increases. With a good model for choosing a ROI, this kind of compression can help in reducing the transmission overhead of video frames, increasing the ability of realizing real-time in video processing. It is intended for future work to compare the wavelet-based foveated compression algorithm with other methods such as the JPEG2000 ROI compression called maxshifts [12].

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support from the Mexican National Council for Science and Technology (CONACYT) and the Puebla State Government under the contract no. 109417.

(a) DCT-based foveated compression

(b) Wavelet-based foveated compression with db9 and Energy profile quantization

(c) Wavelet-based foveated compression with db9 and Tukey window quantization

(d) Wavelet-based foveated compression with db9 and JPEG Type 1 quantization

Fig. 4. Foveated image compression with different methods

TABLE I
COMPARISONS OF DIFFERENT WAVELET-BASED FOVEATEI
COMPRESSION AND DCT-BASED FOVEATED COMPRESSION

Wavelet	J	Zero Coef. Percentage	MSE	Quantization
dct	-	0.9383	2.2402	Type3
db7	2	0.3287	0.2534	Tukey Scaling
db7	4	0.3333	0.2503	Tukey Scaling
db9	2	0.3279	0.3279	Tukey Scaling
db9	4	0.3309	0.2428	Tukey Scaling
db7	2	0.8761	3.8959	JPEG Quantization
db7	4	0.9174	2.4795	JPEG Quantization
db9	2	0.875	3.8167	JPEG Quantization
db9	4	0.9161	2.3974	JPEG Quantization
db7	2	0.9163	2.4192	Energy Profile
db7	4	0.9644	6.0816	Energy Profile
db9	2	0.9153	2.3897	Energy Profile
db9	4	0.9624	6.0116	Energy Profile

REFERENCES

1 . N. Kehtarnavaz and M. Gamadia, "Real-Time Image and Video Processing: From Research to Reality", Morgan and Claypool, University of Texas at Dallas, USA, 2006.

2. E. C. Chang and C. K. Yap, "A wavelet approach to foveating images", *In SCG '97: Proceedings of the thirteenth annual symposium on Computational geometry*, pages 397–399, New York, NY, USA, 1997. ACM.

3. S. Lee and A. Bovik, "Fast algorithms for foveated video processing", *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(2):149 – 162, 2003.

4. Chenlei Guo; Liming Zhang, "A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression", *Image Processing, IEEE Transactions on* , vol.19, no.1, pp.185-198, Jan. 2010

5.A. Graps, "An introduction to wavelets", *IEEE Comput. Sci. Eng.*, 2(2):50–61, 1995.

6. E.-C. Chang, S. Mallat, and C. Yap, "Wavelet Foveation", in *Applied and Computational Harmonic Analysis*, vol. 9, issue 3, pp. 312-335, 2000.

7. S. Mallat, “A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way”, Academic Press, 2008.
8. Digital compression and coding of continuous-tone still images – Requirements and guidelines, *Recommendation T.81 (09/92) ITU-T*; 1992.
9. N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete cosine transform,” *IEEE Transactions on Computers*, vol. C-32, pp. 90-93, Jan. 1974.
10. A. C. Bovik, “The Essential Guide to Image Processing”, Academic Press, 2009.
11. B. A. Wandell. *Foundations of Vision*. Sinauer Associates, Inc., 1995.12. JPEG 2000 image coding system: Core coding system, *Recommendation T.800 (08/02) ITU-T*, 2002
13 . Meijering, E.; ,"A chronology of interpolation: from ancient astronomy to modern signal and image processing," Proceedings of the IEEE , vol.90, no.3, pp.319-342, Mar 2002, doi: 10.1109/5.993400

Generación de Mallas para la Visualización de Estructuras Tubulares

Marco Antonio Caballero Guerrero[1], M. Elena Martinez-Perez[2] y Alejandro Aguilar Sierra[3]

[1] Facultad de Ciencias, UNAM
[2] Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM
[3] Centro de Ciencias de la Atmósfera, UNAM
Paper received on 09/08/10, Accepted on 20/09/10.

Resumen: El presente trabajo propone y describe un sencillo método de reconstrucción de estructuras tubulares, utilizando líneas y mallas de triángulos para generar el modelo tridimensional. Debido a que en el cuerpo humano encontramos numerosos ejemplos de éste tipo de estructuras, sobretodo el sistema vascular, la visualización médica exige aplicaciones que permitan detectar cambios anormales en el grosor de los tubos, ángulos de bifurcación, entre otras características geométricas, diagnosticando posiblemente la presencia de alguna enfermedad de importancia. Así, presentamos los primeros pasos en la generación de una malla que reproduzca adecuadamente la naturaleza de los tubos, resolviendo particularmente las zonas de bifurcación.

1 Introducción

Hoy en día, las aplicaciones por computadora encaminadas a la detección y tratamiento de enfermedades requieren imprescindiblemente la reconstrucción tridimensional de modelos del cuerpo humano, o bien, de órganos específicos que faciliten el análisis médico, permitiendo tener diagnósticos más acertados. De esta forma, la disciplina de la visualización encuentra en tales tareas un extenso campo de trabajo, en el que médicos y especialistas demandan de forma creciente sistemas avanzados que interpreten fiel y claramente los datos procesados. Por tanto, resulta de importancia el uso de métodos y esquemas apropiados para su representación, entre otras, la generación de modelos 3D mediante mallas.

Dentro de las estructuras del cuerpo humano más comunes se encuentran las tubulares, cuyos modelos y técnicas de mallado han sido discutidos ampliamente en la literatura. Así, diversos trabajos han propuesto representar los tubos mediante geometrías elípticas y cilíndricas, tal es el caso de La Cruz *et al.* [2004], en donde describen las superficies de este tipo empleando dimensión, orientación y densidad específicos del interior y exterior del tubo; además las secciones elípticas de cruce son modeladas mediante un punto elegido como centro (x_0, y_0), dimensiones del radio (r_x, r_y) y coeficientes de densidad interna y externa.

Por su parte, Barratt *et al.* [2004] muestran la reconstrucción de bifurcaciones de la arteria carótida, donde las superficies son representadas por esquemas geométri-

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 207-212.

cos, definidos a través de elipses en sentido transversal a la dirección vertical del tubo, así como el uso de curvas *splines* paramétricas.

Otros más como Yim *et al.* [2001], además de definir un sistema paramétrico de coordenadas tubulares, llevan a cabo un procedimiento de mezcla de vértices al fusionar estructuras para mantener espacio entre los vértices y evitar autointersecciones en la superficie. No obstante, algunos autores han optado por el uso de otro tipo de superficies en vez de curvas elípticas y cilindros; como referencia, se encuentra la concatenación de conos truncados propuesta por Hanh *et al.* [2001], donde emplean voxels para generar una estructura de tipo árbol que represente el esqueleto del vaso, suprimen aquellas ramas menos relevantes cuya longitud esté por debajo de cierto umbral definido y generan una bifurcación con el mismo ángulo de abertura para cada rama hija si tal bifurcación tiene un comportamiento simétrico. Por otra parte, Li *et al.* [2007] deciden aproximar el esqueleto tubular a través de curvas NURBS, para posteriormente cubrir el volumen correspondiente extendiendo círculos deformables, cuyos radios varían de acuerdo a la posición de los puntos de control a lo largo de la curva.

De esta manera, ya sea mediante uno de los procesos citados, o algún otro semejante, puede notarse que, generar modelos para reconstruir y visualizar adecuadamente los tubos, facilitarían en buena medida el diagnóstico médico. Siguiendo tal idea, el presente trabajo describe los primeros pasos en la generación de una malla para visualizar estructuras tubulares, en particular, de vasos retinales.

2 Método de mallado 3-D

Comúnmente, la geometría vascular observada aparece constituida por un cuerpo central rodeado de ramificaciones periféricas, o bien, todo un esquema arborescente de ramas y bifurcaciones, donde el estudio médico se enfoca en detectar aquellas zonas a lo largo de los tubos que presentan anomalías, distinguiendo, en ocasiones, ciertos cambios peculiares en el grosor y/o bifurcación de los mismos. Con tal observación, la construcción de la malla debe reproducir lo más exacto y real posible la superficie tubular, además de considerar cómo están dispuestos los valores de entrada para adaptarlos correctamente al método elegido.

Para ello, el punto de partida son los datos numéricos extraídos con el sistema RISA (*Retinal multiScale System Analysis*) [Martinez-Perez et al., 2007] y posteriormente reconstruidos en 3D [Martinez-Perez y Espinosa-Romero, 2004], los cuales están dispuestos en una colección de archivos de texto, cada uno de ellos conteniendo un listado de cinco columnas n_{ki}, x_i, y_i, z_i, r_i, donde la terna (x_i, y_i, z_i), $i=1,\ldots,N$ corresponde a un punto P_i en R^3 con radio r_i (valor promedio para un número finito de puntos), cuya etiqueta n_{ki} es el identificador de los puntos P_i del segmento k en el árbol vascular, donde $k=1,\ldots,M$, con M el número de segmentos del árbol (Figura 1).

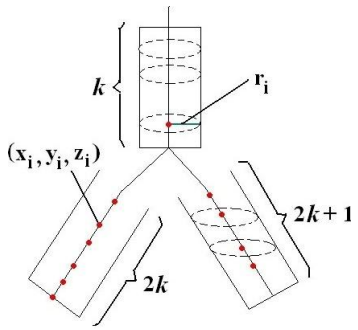


Figura 1. Representación interna de los datos procesados. Los puntos en rojo indican algunos de los puntos P_i de la lista. Las etiquetas (identificadores) de los segmentos se representan bajo la regla [*k, 2k, 2k + 1*], donde *k* indica el identificador para el primer segmento antes de la bifurcación, y *2k*, *2k + 1* son las etiquetas de las ramas hijas de la bifurcación.

2.1 Obtención de puntos de la malla

Con base en la idea de curvas paramétricas y cilindros, proponemos construir la malla proporcionando una colección de vértices y utilizando las primitivas necesarias de la biblioteca gráfica *OpenGL*, sobre el lenguaje de programación *C++* para el trazado de las líneas entre dichos puntos.

Primeramente, se tomó el listado arriba descrito para generar puntos sobre la circunferencia paramétrica C_i determinada por el punto P_i y el valor r_i (centro y radio de la circunferencia, respectivamente), tal que:

$$C_i(t) = \{(x, y) \in \mathbb{R}^2 \mid X^2 + Y^2 = r^2, t \in [0, 2\pi]\},$$

$$\text{donde} \quad \begin{cases} X = x_i + r_i * \cos(t) \\ Y = y_i + r_i * \sin(t) \end{cases}$$

(1)

Como vemos, la coordenada z_i del punto solo es considerada para determinar la posición del plano *XY* sobre la que se coloca la circunferencia, de modo que tales círculos consisten en secciones transversales al eje Z. Así, una vez conseguidos los vértices (los cuales son obtenidos en tiempo de ejecución), se realiza una inspección sobre los datos de **entrada** seleccionados previamente para determinar el primero y el último puntos del segmento *k* en turno (véase Figura 1).

2.2 Trazado de líneas

En esta parte del proceso se unen los puntos mediante GL_LINES_STRIP [4] trazando las circunferencias transversales al eje Z (Figura 2(a)). Asimismo, para generar el cuerpo cilíndrico se trazan líneas o mallas de triángulos (GL_LINES y GL_TRIANGLE_STRIP, respectivamente, según determine el usuario) entre dos círculos siempre y cuando los identificadores nki de las entradas correspondientes pertenezcan al mismo segmento k (Figura 2(b)).

Hasta aquí hemos generado los cilindros de cada segmento sin considerar la zona de bifurcación que hay entre ellos. Para resolver la malla en esta región, de manera preliminar, unimos los vértices finales de la rama padre con los vértices iniciales de cada una de las ramas hijas, sin considerar, por ahora, las posibles intersecciones y superposiciones generadas entre líneas de la malla (Figura 2(c)).

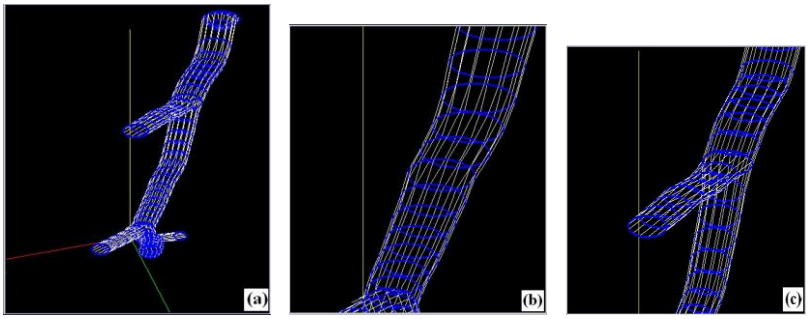


Figura 2. Visualización tridimensional de estructura tubular. (a) Representación 3-D del vaso (el eje Z en color amarillo), (b) región del cuerpo tubular (las circunferencias en color azul, la malla en color blanco y (c) bifurcación, obsérvese la malla que une los segmentos.

3 Resultados

Utilizando el método anterior, se realizó la ejecución del programa con algunos archivos de puntos generados por RISA [Martinez-Perez et al., 2007], [Martinez-Perez y Espinosa-Romero, 2004]. Moviendo la posición de la cámara para observar los detalles más relevantes, vemos en la Figura 3 una representación tubular aceptable, donde la malla parece suave, debido a la relativa simetría de la estructura. La Figura 3(a) muestra la malla usando GL_LINES entre círcunferencias, mientras que

[4] GL LINES STRIP es una de las primitivas básicas de *OpenGL*

en Figura 3(b) empleamos GL_TRIANGLE_STRIP rellenando el volumen de la superficie.

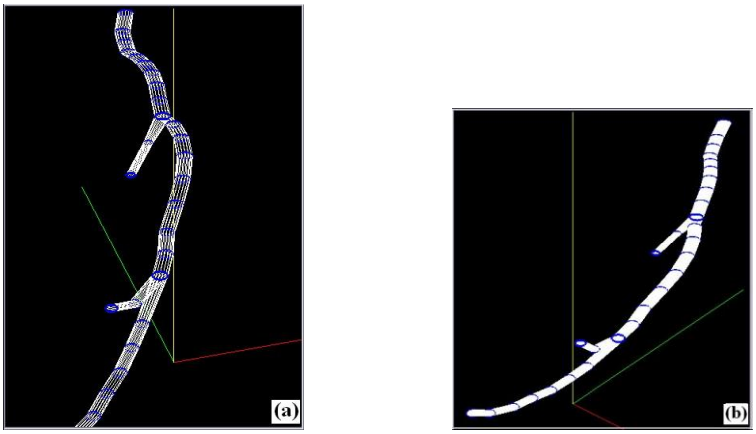


Figura 3. Reconstrucción tubular apropiada, favorecida por la simetría de los puntos centrales del esqueleto. (a) Malla trazada con GL_LINES y (b) malla trazada con GL_TRIANGLE_STRIP vista desde otro ángulo.

Nótese que en las regiones donde el vaso mantiene una dirección uniforme la malla está mejor definida. En cambio, en la Figura 4 en algunas ramas próximas a una bifurcación, las circunferencias se encuentran más separadas tanto horizontal como verticalmente (Figura 4(a)). Esto claramente dificulta el correcto trazado del tubo, pues si realizamos la triangulación entre dos circunferencias contiguas donde se presente este problema, obtenemos una superficie con volumen reducido y casi completamente plana (Figura 4(b), (c)).

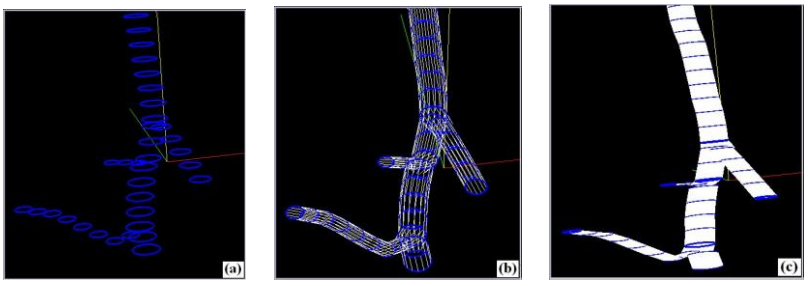


Figura 4. Visualización del trazado de la malla. (a) Circunferencias transversales al eje Z (obsérvese la separación entre circunferencias contiguas), (b) trazado de líneas de la estructura y (c) malla generada con GL{_}LINES{_}STRIP (se distingue una superficie con poco volumen.

4 Conclusión y Trabajo a Futuro

Hemos mostrado un procedimiento que, en términos generales, adapta de manera sencilla los datos que tenemos al modelo tridimensional logrado. Asimismo, la generación de los puntos sobre la malla se consigue fácilmente mediante una parametrización muy simple y poco costosa en cuanto a operaciones aritméticas (Figura 2(a), (b) y Figura 3).

Sin embargo, aunque todavía no se resuelve correctamente las zonas de bifurcación (de modo que evite la superposición de líneas entre segmentos), une los segmentos de forma sencilla y rápida si deseamos tan sólo apreciar la estructura cerrada del tubo (Figura 2(c)). Por otra parte, una desventaja evidente es la disminución del volumen en las regiones donde las circunferencias se encuentran más separadas, lo anterior debido a que se toman los círculos en secciones transversales al eje Z siempre (Figura 4).

Una posible solución a esto último es tomar la línea central del esqueleto formada por los puntos P_i y trazar las circunferencias perpendicularmente a éste. Para solucionar la malla en la zona de bifurcación, podríamos obtener puntos de una esfera central entre los tres segmentos que forman la bifurcación y utilizar técnicas de mallado con NURBS para completar el modelo.

Referencias

1. Barratt, D., Ariff, B., Humphries, K., Thom, S., y Hughes, A. (2004). Reconstruction and quantification of the carotid artery bifurcation from 3-D ultrasound images. *EEE Transactions on Medical Imaging*, **23**(5), 567 –583.
2. Hahn, H., Preim, B., Selle, D., y Peitgen, H.-O. (2001). Visualization and interaction techniques for the exploration of vascular structures. In *Visualization, 2001. VIS'01. Proceedings*, pages 395 –578.
3. La Cruz, A., Straka, M., Kochl, A., Sramek, M., Groller, E., y Fleischmann, D. (2004). Non-linear model fitting to parameterize diseased blood vessels. In *Proceedingsof the conference on Visualization '04*, pages 393–400, Washington, DC, USA. IEEE Computer Society.
4. Li, J., Regli, W. C., y Sun, W. (2007). Mathematical representation of the vascular structure and applications. In *SPM '07: Proceedings of the 2007 ACM symposium on Solid and physical modeling*, pages 373–378, New York, NY, USA. ACM.
5. Martinez-Perez, M. E. y Espinosa-Romero, A. (2004). Retinal blood Wessel 3D Reconstruction from two views. In *4th Indian Conference Vision and Graphics*, págs. 258 – 263.
6. Martinez-Perez, M. E., Hughes, A.D., Thom, S.A., Bharath, A A., y Parker, K. H. (2007). Segmentation of blood vessels from red-free and fluorescein retinal images. *Medical Image Analysis*, 11(1), 47 – 61.
7. Yim, P., Cebral, J., Mullick, R., Marcos, H., y Choyke, P. (2001). Vessel surface reconstruction with a tubular deformable model. *IEEE Transactions on Medical Imaging*, **20**(12), 1411 –1421.

Una Comparación Entre Métodos de Segmentación en Imágenes de Escenas de Interiores de Inmuebles

Salvador Cervantes Álvarez y Raúl Pinto Elías

Departamento de Ciencias Computacionales
Centro Nacional de Investigación y Desarrollo Tecnológico
[scervantes, rpinto]@cenidet.edu.mx
Paper received on 09/08/10, Accepted on 20/09/10

Resumen. En este trabajo se presenta una comparación perceptual entre las segmentaciones producidas por los algoritmos de segmentación Mean Shift, gPb-owt-ucm y la técnica Basada en Grafos de Felzenszwalb aplicados en imágenes de escenas de interiores de inmuebles no controladas. La comparación del desempeño de los algoritmos se realizó considerando los siguientes aspectos: sobre segmentación (partición de una región homogénea en 2 o más regiones), segmentación burda (unión de 2 o más regiones no homogéneas dentro de una sola región) y segmentación consistente. Los algoritmos fueron aplicados sobre las colecciones de imágenes de [17] y PASCAL 2007 [20].

Palabras clave: Segmentación, escenas de interiores.

1 Introducción

El objetivo de las técnicas de segmentación es obtener una partición de la imagen en regiones con apariencia homogénea dada una cierta medida de similitud aplicada sobre un conjunto de descriptores. Varias técnicas de segmentación actuales ([2], [4], [5], [9]) intentan segmentar la imagen en regiones homogéneas que contengan un significado perceptual, donde las regiones aíslen a un objeto o partes de objetos presentes en una imagen (por ejemplo, el respaldo o el descansa brazos de una silla de escritorio).

La segmentación es una etapa indispensable en varios modelos que utilizan información de alto nivel ([1], [3], [10], [11], [18], [19]), sin embargo, la tarea de obtener una segmentación fiable sigue siendo un problema desafiante, y el buscar un algoritmo de segmentación con el mejor desempeño ha sido el objetivo de varios trabajos que realizan estudios comparativos entre los algoritmos de segmentación más ampliamente usados en el análisis de imágenes de escenas ([7], [8], [13], [14]); estos estudios son realizados sobre la colección de imágenes Berkeley [14], la cual, está compuesta principalmente por imágenes de escenas de exteriores. En la literatura no se han encontrado comparaciones de desempeño, realizadas con imágenes de escenas de interiores, motivo por el cual se desarrolló la actual investigación.

El resto de este documento está estructurado de la siguiente forma: en la Sección 2 se explica el funcionamiento de los tres algoritmos de segmentación analizados; en la Sección 3 se describen las características de las imágenes de escenas de interiores de las colecciones de imágenes de [17] y [20] utilizadas en el experimento; en la Sección 4 se presentan los resultados obtenidos con los métodos de segmentación bajo estudio y en la Sección 5 se presentan las con-

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 213-222

clusiones así como las sugerencias para obtener una comparación más precisa de los resultados.

2 Algoritmos de segmentación de escenas

Los algoritmos de segmentación comparados, fueron seleccionados debido a que son los más ampliamente usados en aplicaciones recientes, proporcionan un desempeño razonable ([2], [4], [9]) y sus implementaciones están disponibles públicamente. A continuación se explica en forma general el funcionamiento de cada uno de los métodos de segmentación analizados.

2.1 Mean Shift

En el método de [4], los píxeles de una imagen son descritos en un espacio de características multidimensional; la implementación utiliza el espacio de color CIE L*u*v*. El algoritmo busca las características que corresponden a las regiones más densas en el espacio de características, es decir a los cluster que corresponden a regiones en la imagen. El objetivo del análisis del espacio de características multidimensional, es delinear estos clusters.

El método Mean Shift se ejecuta en forma recursiva y converge hacia el punto con mayor densidad de acuerdo a una *función de densidad fundamental* detectando las *modas* de las densidades, para lo cual, se utiliza la técnica de ventana de Parzen [6] como kernel. Se comienza analizando una región inicial (ventana) y la media local es cambiada hacia la subregión en la que reside la mayoría de los puntos, de esta forma el vector Mean Shift puede definir el camino que lleva a puntos estacionarios de *densidad estimada*. El procedimiento Mean Shift ejecutado en forma sucesiva, calcula el vector Mean Shift y la translación del kernel (ventana) y garantiza la convergencia hacia puntos cercanos estacionarios (ver figura 1).

En [4], se realiza una teselación con ventanas iniciales en el espacio de características y se realiza una ejecución en forma paralela del método Mean Shift para cada ventana. Todos los puntos en el espacio de características evaluados por ventanas que convergen hacia un mismo punto estacionario, pertenecen a una misma región.

A la representación de la imagen en el espacio de características CIE L*u*v*, se incorporan las coordenadas de un píxel, generando una representación de *dominio mixto*. El *dominio mixto* está compuesto por el *dominio espacial* (coordenadas x y y) y el *dominio de rango* (espacio CIE L*u*v*), para los cuales se utilizan kernels por separado y se define el ancho de la ventana de Parzen h_s y h_r, para el domino espacial y de rango respectivamente. La selección del ancho del kernel no es algo trivial ya que un valor muy grande de h puede llevar a obtener una segmentación densa y un valor pequeño de h puede generar una sobre segmentación.

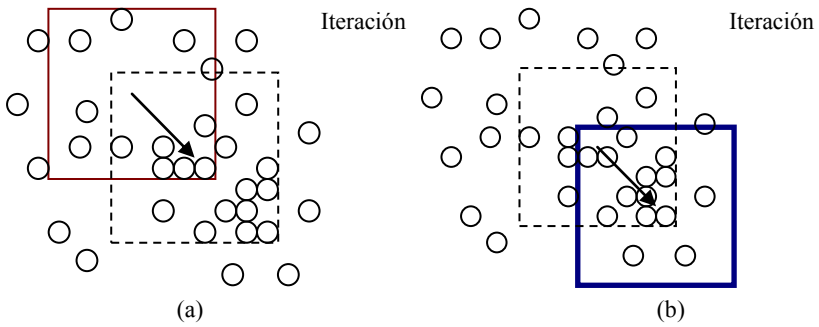


Figura 1. En la figura (a) se muestra el resultado del desplazamiento de la ventana de Parzen en la iteración 1 a partir de una posición inicial hacia una posición intermedia donde se encontró la mayor densidad en el espacio analizado. En la figura (b) se muestra el resultado de la iteración 2 donde a partir de una posición intermedia se alcanza el punto estacionario local.

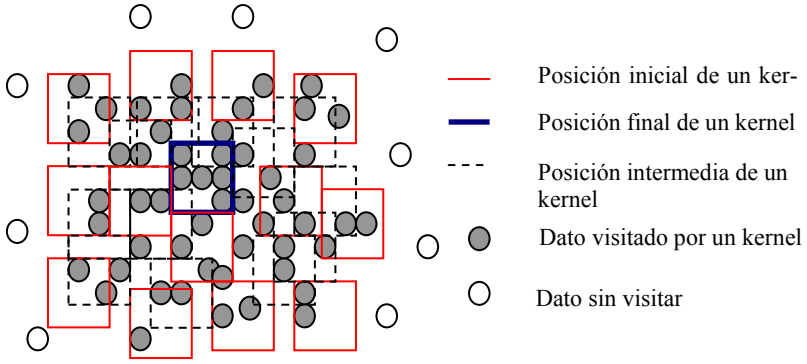


Figura 2. Se presenta una teselación del espacio de características. Todos los puntos que son visitados por ventanas de Parzen que convergen hacia el mismo punto pertenecen a una misma región.

2.2 Técnica basada en grafos de Felzenszwalb

El método de [9] considera a una imagen como un grafo dirigido y cada uno de sus píxeles son vértices conectados con sus píxeles adyacentes (considerando una vecindad cuatro u ocho píxeles) a través de aristas que miden la disimilaridad que existe entre los píxeles. Una descripción formal del método es: Sea $G = (V, E)$ un grafo no dirigido con vértices $u_i \in V$ (el conjunto de elementos a segmentar) y las aristas $(u_i, u_j) \in E$ correspondan a pares de vértices vecinos. Cada arista $(u_i, u_j) \in E$ tiene su correspondiente peso $w((u_i, u_j))$, el cual, es una medida no negativa de la disimilaridad entre los elementos vecinos u_i y u_j. Los elementos de V son píxeles de la imagen y los pesos w de cada arista es alguna medida de disimilaridad entre dos píxeles conectados por dicha arista.

En el enfoque basado en grafos, una segmentación S es una partición de V dentro de varios componentes, tal que cada componente (o región) $C \in S$ corresponde a componentes conectados en un grafo $G' = (V, E')$, donde $E' \subseteq E$. Cualquier segmentación es inducida por un subconjunto de aristas en E. En la segmentación, las aristas de dos vértices en el mismo componente tienen pesos relativamente bajos, y las aristas entre vértices en diferentes componentes tienen pesos altos.

Para realizar la segmentación, se tiene un predicado D, que compara las diferencias entre componentes (regiones) con las diferencias dentro de cada componente y es por lo tanto adaptativo a las características locales de los datos. La *diferencia interna de un componente* $C \in V$, es el peso más grande en el *árbol de expansión mínimo* (MST por sus siglas en inglés) de el componente, MST(C, E). Esto es:

$$Int(C) = \max_{e \in MST(C,E)} w(e) \tag{1}$$

La *diferencia entre dos componentes* C_1, $C_2 \in V$ es la arista de menor peso que conecta los dos componentes. Esto es:

$$Dif(C_1, C_2) = \min_{u_i \in C_1, u_j \in C_2, (u_i, u_j) \in E} w((u_i, u_j)) \tag{2}$$

Si no hay arista conectado a C_1 y C_2 se estable $Dif(C_1,C_2) = \infty$. El predicado de comparación de regiones D evalúa si existe evidencia para un límite entre un par de componentes revisando si la diferencia entre los componentes, $Dif(C_1,C_2)$, es relativamente grande con respecto a la diferencia interna dentro de al menos uno de los componentes, $Int(C_1)$ y $Int(C_2)$. Una función de umbral es usada para controlar el grado al cual la diferencia entre componentes debe ser más grande que la diferencia mínima interna. El predicado de comparación es definido como:

$$D(C_1, C_2) = \begin{cases} verdad & Si \quad Dif(C_1, C_2) > MInt(C_1, C_2) \\ falso & deotraforma \end{cases} \tag{3}$$

Donde la diferencia mínima interna, *MInt*, es definida como:

$$MInt(C_1, C_2) = \min(Int(C_1) + \tau(C_1), Int(C_2) + \tau(C_2)) \tag{4}$$

Donde la función de umbral τ controla el grado a el cual la diferencia entre dos componentes debe ser más grande que las diferencias internas con el objetivo de que exista evidencia de la existencia de un límite entre ellos (que D sea verdadero). Sin embargo, para pequeños componentes, $Int(C)$ no es una buena estimación de las características locales de los datos. En el caso extremo, cuando $|C| = 1$, $Int(C) = 0$. Por lo tanto, se usa una función de umbral basada en el tamaño del componente:

$$\tau(C) = k / | C | \tag{5}$$

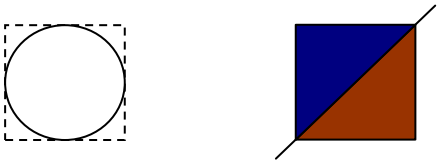
Donde $|C|$ denota el tamaño de C, y k es un parámetro constante. Para pequeños componentes se requiere de una fuerte evidencia para establecer la existencia de un límite. En la práctica k establece una escala de observación, en la

que valores grandes de *k* causan una preferencia por componentes grandes. Sin embargo, *k* no es un tamaño de componente mínimo.

2.3 gPb-owt-ucm

En [2] se genera una segmentación jerárquica de una imagen partiendo de los contornos de la imagen, para lo cual se utiliza el método para obtener de contornos de *probabilidad global de límites* (*gPb*), aunque puede utilizarse cualquier otro método para obtener los contornos. A partir de los contornos se aplica la *transformada watershed orientada* (*owt*) para producir un conjunto de regiones iniciales, con las cuales se construye un *mapa de contornos ultra métrico* (*ucm*). La secuencia de operaciones *owt-ucm* produce un árbol de regiones jerárquico.

El método *gPB* [12] está basado en el detector de contornos *Pb* [13], en el cual se calcula una señal de gradiente orientada $G(x, y, \theta)$ a partir de una imagen de intensidades *I*. El cálculo se realiza mediante colocar un disco en la posición (x, y) y dividirlo en dos mitades en el ángulo θ, aplicado cada uno de los canales del espacio de color CIE Lab y utilizando textones para la textura.



(a) (b)

Figura 4. (a) El disco es reemplazado por una ventana que lo contiene. (b) División de la ventana en un ángulo θ.

En cada mitad del disco se calcula un histograma de intensidad de los píxeles de *I*. La magnitud del gradiente *G* en la posición (x, y) está definida por la distancia X^2 entre los histogramas de las dos mitades del disco. Se realizan tres convoluciones en los canales de brillo (canal *L*), color (canales *a* y *b*) y textura (mapa de textones) con tamaños de disco iguales a $[\sigma/2, \sigma, 2\sigma]$, donde σ es la escala del detector *Pb*.

Los resultados con los diferentes tamaños de disco y en los diferentes canales son combinados en forma lineal con (6).

$$mPb(x, y, \theta) = \sum_{s} \sum_{i} \alpha_{i,s} G_{i,\sigma(i,s)} \quad (x, y, \theta) \tag{6}$$

donde *s* es el índice de las escalas, *i* el índice de los canales de características y $G_{i,\sigma(i,s)}(x, y, \theta)$ mide la diferencia de histogramas en el canal *i* entre dos mitades de un disco de radio $\sigma(i, s)$ centrado en (x, y) y divididos en el ángulo θ. Tomando la repuesta máxima sobre las orientaciones, se obtiene una medida de la fuerza del límite en cada píxel.

$$mPb(x, y) = \max_{\theta} \{mPb(x, y, \theta)\} \tag{7}$$

Para obtener *gPb* se necesita una etapa de agrupamiento espectral, donde se construye una matriz de afinidad *W* donde el máximo valor de *mPb* a lo largo de una línea conecta dos puntos.

$$W_{ij} = \exp\left(- \max_{p \in ij} \{mPb(p)\} / \sigma \right) \tag{8}$$

Donde *ij* es un segmento de línea que conecta a *i* y *j*, y σ es una constante. Se define $D_{ii} = \Sigma_j W_{ij}$ y se resuelve para los eigenvectores generalizados $\{v_0, v_1, \ldots, v_k\}$ del sistema $(D - W)v = \lambda Dv$. Tratando cada eigenvector $\mathbf{v_k}$ como una imagen, se realiza una convolución con *filtros derivados direccionales Gaussianos* en múltiples orientaciones θ, obteniendo una señal $sPb_v(x, y, \theta)$. La información de los diferentes eigenvectores es combinada para obtener el componente espectral del detector de contornos *gPb*.

$$sPb(x, y, \theta) = \sum_{k=1}^{n} \frac{1}{\sqrt{\lambda_k}} . sPb_v(x, y, \theta) \tag{9}$$

La probabilidad global del límite es entonces escrita como una suma ponderada de las señales locales y espectrales:

$$gPb(x, y, \theta) = \sum_s \sum_i \beta_{i,s} G_{i,\sigma(i,s)} (x, y, \theta) + \gamma . sPb(x, y, \theta) \tag{10}$$

El método *gPb* produce contornos cerrados, lo cual es necesario para evitar que el método *owt* fusione regiones que están separadas por un límite. Los valores de $\beta_{i,s}$ y γ son aprendidos utilizando como entrenamiento las imágenes de la base de imágenes Berkeley.

(a) (b) (c)

Figura 5. Imagen original de [17] de 256x256 píxeles. (b) Contornos de la imagen. (c) Segmentación de la imagen.

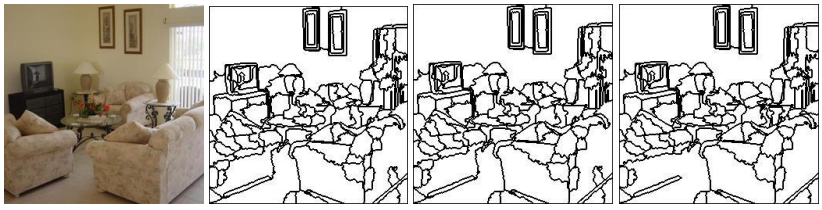
En la figura 5b se muestran los contornos cerrados obtenidos del procesamiento la imagen de la figura 5. Los contornos con mayor gradiente tienen una mayor intensidad y son representados en forma de árbol mediante el proceso *owt-ucm* en donde en el nivel más bajo del árbol (las hojas) se genera una sobre segmentación de la imagen equivalente a las regiones separadas por todos los contornos y en los niveles más altos del árbol se producen segmentaciones burdas (ver figura 5c) en donde los contornos con mayor gradiente segmentan la imagen.

3 Colecciones de imágenes de escenas de interiores

Las colecciones utilizadas para probar el desempeño de los algoritmos de segmentación en escenas de interiores son: imágenes de las categorías de sala, comedor y oficina de la colección generada en [17] e imágenes de interiores de la colección de entrenamiento PASCAL 2007 [20]. Las imágenes de las colecciones son no controladas y las dimensiones de las imágenes varían.

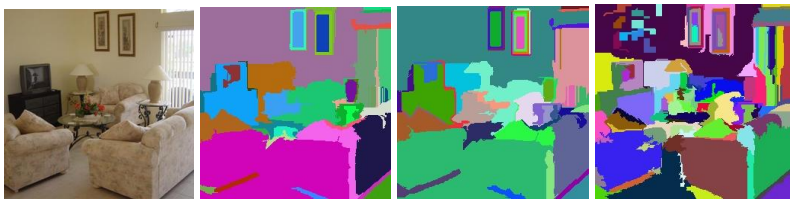
4 Resultados

Para mostrar el comportamiento de los métodos de Mean Shift y de Felzenszwalb, en las figuras 6 y 7 se presentan respectivamente resultados obtenidos con diferentes valores para sus parámetros, intentando evitar segmentaciones burdas. En ambos métodos el parámetro *mínimo* permite establecer la mínima cantidad de píxeles que debe tener una región, en caso de que la región sea menor es fusionada con la región adyacente con mayor similitud. En todos los casos se utilizó una máquina con 1.9 GHz y 3GB en RAM.



(a) (b) (c) (d)

Figura 6. Imagen original de [17] de 256x256 píxeles. (b) hs=1, hr=1.4, mínimo=160 (c) hs=8, hr=1.4, mínimo=160 (d) hs= 8, hr=1.5, mínimo=160.

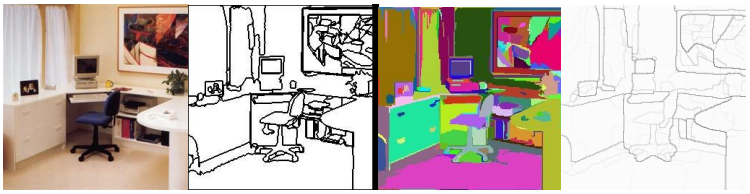


(a) (b) (c) (d)

Figura 7. Imagen original de [17] de 256x26 píxeles. (b) $\sigma = 0.8$, $k = 300$, mínimo=100. (c) $\sigma = 0.5$, $k = 300$, mínimo=100. (d) $\sigma = 0.5$, $k = 100$, mínimo=100.

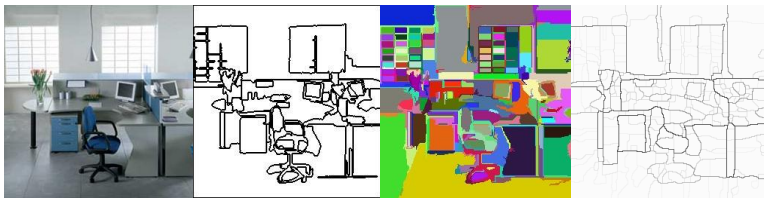
En el método Mean Shift por lo general se obtuvieron sobre segmentaciones de la imagen y en la figuras 6b y 6c se puede observar que se pueden obtener segmentaciones idénticas de la imagen con valores distintos para los parámetros *hs* y *hr*. A diferencia de Mean Shift el método de Felzenszwalb presentó una tendencia a generar segmentaciones burdas. En las figuras 8, 9, 10 y 11 se muestran los resultados de los tres métodos de segmentación; la imagen de la izquierda es la imagen original y las siguientes imágenes son los resultados de los métodos Mean Shift, Felzenszwalb y *gPb-owt-ucm* respectivamente. Se realizó

una selección manual de los parámetros de los métodos Mean Shift y Felzenszwalb intentando que produjeran segmentaciones consistentes; en el caso del método *gPb-owt-ucm* se utilizaron los valores de [2] para el tamaño de disco (σ=5 para el canal L y σ=10 para los demás canales).



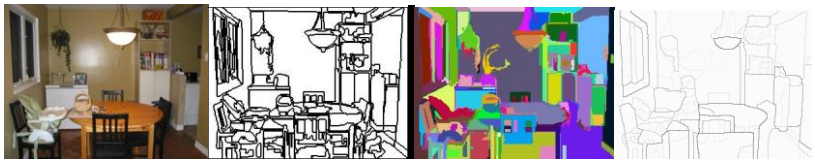
(a (b (c (d

Figura 8. (a) Imagen original de [17] de 256x256 píxeles. (b) hs=10, hr=9.2, mínimo=20. (c) σ =0.5, *k* =200, mínimo=20. (d) Contornos cerrados generados por el método *gPb-owt-ucm*.



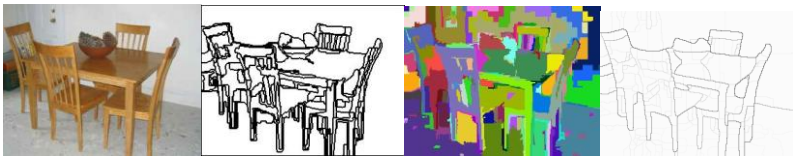
(a) (b) (c) (d

Figura 9. (a) Imagen original de [17] de 256x256 píxeles. (b) hs=4, hr=6.2, mínimo=100. (c) σ =0.5, *k* =200, mínimo=20. (d) Contornos cerrados generados por el método *gPb-owt-ucm*.



(a) (b) (c) (d)

Figura 10. (a) Imagen original de PASCAL 2007 [20] de 310x233 píxeles. (b) hs=1, hr=3.8, mínimo=120. (c) σ =0.5, *k* =200, mínimo=20. (d) Contornos cerrados generados por el método *gPb-owt-ucm*.



(a) (b (c) (d)

Figura 11. (a) Imagen original de [17] de 300x225 píxeles. (b) hs=10, hr=4.8, mínimo=80. (c) σ =0.5, *k* =200, mínimo=20. (d) Contornos cerrados generados por el método *gPb-owt-ucm*.

..El método gpb-owt-ucm requirió 1 hora 38 minutos para procesar una imagen de 321x481 píxeles siendo un método computacionalmente caro, mientras que los métodos Mean Shift y Felzenszwalb requirieron unos cuantos segundos, siendo este último el más rápido. Para lograr una segmentación consistente, los parámetros de los métodos Mean Shift y Felzenszwalb cambiaron en cada imagen, sin embargo, se mantuvo un comportamiento de sobre segmentación en Mean Shift y de segmentación burda en Felzenszwalb, mientras que los resultados de gPb-owt-ucm son dependientes de la elección del nivel del árbol (mediante la definición de un umbral), por tal razón solamente se muestra la imagen de los contornos en las figuras 8d, 9d, 10d y 11d.

5 Conclusiones

Siguiendo la idea de [2] es mejor obtener una sobre segmentación de la imagen y después fusionar las regiones en forma jerárquica basándose en otros descriptores y métricas. Tomando en cuenta lo anterior, la técnica de Felzenszwalb es la de peor desempeño, aunque es de los algoritmos más rápidos según [16].

El método *gPb-owt-ucm* es computacionalmente caro y la calidad de la segmentación depende de la elección del nivel del árbol generado por el proceso *owt-ucm*, sin embargo, tiene la ventaja de poder establecer relaciones de pertenencia entre regiones debido a su estructura de árbol, lo cual, puede ser útil para la representación de objetos como en [1]. Los métodos Mean Shift y Felzenszwalb pueden ser mejorados agregando descriptores de textura en el proceso de segmentación y en el caso del método Mean Shift se puede seguir un proceso de mezcla de regiones en la sobre segmentación, buscando generar una segmentación jerárquica como en [2].

Los resultados de la comparación pueden ser útiles para futuros trabajos de investigación en escenas de interiores, que requieran del empleo de un algoritmo de segmentación tomando en cuenta el tiempo de procesamiento requerido y el desempeño. Sin embargo, para poder obtener una evaluación más precisa del desempeño de los algoritmos de segmentación en imágenes de escenas de interiores, se debe realizar una comparación cuantitativa de los algoritmos de segmentación, para lo cual, se requerirá que varias personas realicen segmentaciones manuales de las imágenes como en la base de imágenes de Berkeley [14] y aplicar una medida cuantitativa a las segmentaciones resultantes como en [7], [8], [13], [14].

Referencias

1. Ahuja, N., Todorovic, S.: Connected Segmentation Tree – A Joint Representation of Region Layout and Hierarchy. Computer Vision and Pattern Recognition, (2008).
2. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: From Contours to Regions: An Empirical Evaluation. Computer Vision and Pattern Recognition, (2009).
3. Christoudias, C., Georgescu, B., Meer, P.: Synergism in Low Level Vision. International Conference on Pattern Recognition, vol. 4, (2002).
4. Comaniciu, D., Meer, P.: Mean Shift: A Robust Approach Toward Feature Space Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 603-619, (2002).
5. Cour, T., Benezit, F., Shi, J.: Spectral Segmentation with Multiscale Graph Decomposition. Computer Vision and Pattern Recognition, (2005).

6. Duda, R. O., Hart, P. E., Store, D. G.: Pattern Classification, Second Edition. Wiley, (2000).
7. Estrada, F, Jepson, A.: Quantitative Evaluation of a Novel Image Segmentation Algorithm. Computer Vision and Pattern Recognition, vol. 2, pp. 1132-1139, (2005).
8. Estrada, F, Jepson, A.: Benchmarking Image Segmentation Algorithms. International Journal of Computer Vision, vol. 85, pp. 167-181, (2009).
9. Felzenszwalb, P. F., Huttenlocher, D.: Efficient Graph-Based Image Segmentation. International Journal of Computer Vision, (2004).
10. Gu, C., Lim, J., Arbelaez, P., Malik, J.: Recognition using Regions. Computer Vision and Pattern Recognition, (2009).
11. Li, L., Socher, R., Fei-Fei, L.: Towards Total Scene Understanding: Classification, Annotation and Segmentation in an Unsupervised Framework, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2009).
12. Maire, M., Arbelaez, P., Fowlkes, C., Malik, J.: Using Contours to Detect and Localize Junctions in Natural Images, Computer Vision and Pattern Recognition, (2008).
13. Martin, D., Fowlkes C., Malik, J.: Learning to Detect Natural Image Boundaries Using Local Brightness, Color and Texture Cues. IEEE Transactions on Pattern Analysis and Machine Learning, vol. 26, pp. 530-549, (2004).
14. Martin, D., Fowlkes C., Tal, D., Malik, J.: A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. On Proceedings of the International Conference on Computer Vision, vol. 2, (2001).
15. Meer, P. Georgescu, B.: Edge detection with embedded confidence. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 1351-1365, (2001).
16. Paris, S., Durand, F.: A Topological Approach to Hierarchical Segmentation Using Mean Shift. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, (2007).
17. Quattoni A., Torralba, A.: Recognizing Indoor Scenes. IEEE Conference on Computer Vision and Pattern Recognition, (2009).
18. Rabinovich, A., Vedaldi, A., Galleguillos, C., Wiewiora, E., Belongie, S.: Objects in Context. International Conference on Computer Vision, (2007).
19. Russell, B. C., Efros, A. A., Sivic, J., Freeman, W.T., Zisserman, A.: Using Multiple Segmentations to Discover Objects and their Extent in Image Collections. Computer Vision and Pattern Recognition, vol. 2, pp. 1605-1614, (2006).
20. Visual Object Classes Challenge, http://pascallin.ecs.soton.ac.uk/challenges/VOC/.

Elicitación y Comparación de Conocimiento Semántico para Ontologías Naturales

L. Cota-Gomez y J. Figueroa-Nazuno

Centro de Investigación en Computación
Instituto Politécnico Nacional
Unidad Profesional "Adolfo López Mateos"
{lcota,jfn}@cic.ipn.mx
Paper received on 28/07/10, Accepted on 27/09/10.

Resumen. Las ontologías son un componente técnico fundamental para la interoperabilidad semántica, objetivo fundamental en la Semantic Web. Se han desarrollado distintos métodos y herramientas para su construcción, que varían desde la anotación directa efectuada por ingenieros de conocimiento hasta aproximaciones de extracción semiautomática en documentos electrónicos. En una u otra forma las ontologías son la descripción por medio de grafos, de diferentes relaciones entre conceptos que a su vez sirven para definir otro concepto, esas relaciones pueden ser por ejemplo: asociaciones, sinonimias, conceptos frecuentemente relacionados, etc. El presente trabajo describe una metodología automatizada para elicitación y comparación cuantitativa de conocimiento semántico, capaz de generar estructuras computacionales para ontologías que reflejan directamente el significado utilizado en una comunidad suscrita a un dominio.

Palabras Clave: Ingeniería del Conocimiento, Elicitación de Conocimiento, Construcción de Ontologías, Comparación de Ontologías, Modelo Semántico, Redes Semánticas Naturales.

1 Introducción

Uno de los principales tópicos constantes en la Inteligencia Artificial es la necesidad de manejar contenido semántico. La búsqueda de interoperabilidad semántica se enfatizó con el surgimiento de la idea de Semantic Web [1] debido a la vasta cantidad de información contenida en medios electrónicos originalmente creada para ser únicamente leída por humanos (human readable information) y su poco potencial de explotación automática (machine-processable information).

Los componentes técnicos considerados con gran importancia para permitir el manejo de contenido semántico son las ontologías. Algunas definiciones sobre lo que es una ontología han sido dadas por Gruber: «una especificación explicita de una conceptualización»; Borst: «una especificación formal de una conceptualización compartida»; Studer: «Una ontología es una especificación formal, explícita, de una conceptualización compartida». Todas estas definiciones asumen una noción informal de «conceptualización»[19]. El propósito primordial de éstas, en su forma más general, consiste en almacenar una colección de conocimiento humano estructurado

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 225-233.

para otorgar significado, por medio de relaciones entre conceptos, habilitando a las computadoras y a las personas a trabajar en cooperación.

Existen muchos tipos de ontologías, y diferentes formas de clasificarlas, que pueden utilizarse para múltiples propósitos, también se han desarrollado distintos métodos y herramientas para su construcción y administración [19].

El problema general de ontologías en el sentido computacional tiene muchas aristas que van desde metodologías totalmente formales para su construcción hasta trabajos que son recolección de datos directos en campo [6,9,13]. Uno de los aspectos más importantes, es la extracción de esas estructuras generalmente en forma de grafos que definen el significado, en dónde los nodos son conceptos con un valor numérico, sin embargo el resultado final y más representativo son las relaciones o red de conceptos.

Algunos métodos para la anotación de ontologías obtienen su conocimiento directamente de humanos a través de técnicas ”activas” como mesas redondas, entrevistas, diálogos y otras semejantes; o técnicas ”pasivas” que incluyen observación, lecturas y protocolo de ”pensar en voz alta”. En todas, resulta sumamente complicado obtener la representación de un consenso, además de que el proceso de anotación se ve afectado por la transcripción.[12,14]

Por otra parte, muchas metodologías para la construcción de ontologías implican el uso de otra auxiliar que proporcione significados representados a través de relaciones entre conceptos, ya sea de cierta área particular o de sentido común no especializado. Un ejemplo de ontología muy usada para éste propósito es WordNet, considerada como el estándar de oro (”gold standard”)[7], y catalogada como base de datos léxica o como ”upper ontology”, que irónicamente su objetivo en un principio no era representar significado.[10]

Este trabajo presenta una metodología para construcción de Ontologías Naturales implementada en software, procedente de la técnica clásica de Redes Semánticas Naturales (RSN) [11] cuya fundamentación teórica y empírica como estructuras que representan el significado en humanos, está sólidamente demostrada por las ciencias cognitivas. [3,2,5,15,21]

Otros trabajos recientes en el área, también se han basado en RSN con diferente procedimiento.[20,4] Las particularidades y sutilezas de la técnica RSN para elicitación del significado, permite generar de forma directa estructuras que contienen palabras y relaciones, proporcionando métricas de distancia y similitud intrínsecas que representan el significado de conceptos importantes con consenso de una determinada comunidad suscrita en cierto dominio de conocimiento. Lo más relevante es que estas estructuras son económicamente manejables computacionalmente, esto es demostrado mediante el ejemplo desarrollado para la construcción de una ontología formada con conceptos del dominio ”Computación”.

2. Construcción de una Ontología Natural del dominio Computación

Se presenta una aproximación para la representación del significado en humanos por medio de la técnica de RSN [11,23], que directamente puede obtener Ontologías

Naturales compartidas por una comunidad, proporcionando estructura y métricas de comparación cuantitativas de forma directa y simple, automatizable y económicamente computacional.

A continuación se describe la metodología para la elicitación y comparación de Ontologías Naturales, mediante la construcción de una ontología en el dominio ”Computación”.

2.1 Metodología para Elicitación y Obtención de la Ontología Natural

Elección de conceptos base dentro del dominio de conocimiento. Para elegir los conceptos que fungirán como base para la construcción de la ontología, se pidió a varios expertos en el área, que determinarán cuáles son los conceptos más importantes en el dominio. Después, se eligieron los 5 conceptos con mayor frecuencia de mención. Este paso puede tener variaciones, inclusive resulta útil hacer uso de la misma técnica que se realiza para obtener las definidoras.

En este caso los conceptos elegidos fueron: Algoritmo, Información, Lenguaje, Red y Sistema Operativo.

Elicitación del significado de los conceptos base. Se eligen 30 sujetos que representen una comunidad inmersa en un dominio de conocimiento, que mediante un sistema de captura, se les solicita lo siguiente:

1. Proporcionar el significado de cada uno de los conceptos base, por medio de un rango de 5 a 10 palabras sueltas, sin usar partículas gramaticales.
2. Jerarquizar las palabras proporcionadas.

Es muy importante tomar en cuenta la sutileza de solicitar el significado de conceptos; ya que esta demostrado que se pueden obtener resultados muy diferentes si se pide dar ”palabras relacionadas”[11].

Para este caso, se eligieron 3 comunidades inmersas en el dominio de Computación con diferentes enfoques para formar las RSN de cada grupo:

RSN1 Grupo de expertos con enfoque en Ciencias.
RSN2 Grupo con enfoque en Sistemas Computacionales.
RSN3 Grupo con enfoque en Telemática.

Cálculo de Pesos Semánticos. Para cada concepto definido, y a su vez, para cada una de las palabras definidoras obtenidas (cabe mencionar que para determinar las definidoras en ocasiones es necesario efectuar un procesamiento previo, en este caso manual, en el que se pueden fusionar palabras sinónimos y eliminar errores ortográficos o léxicos), se calcula la frecuencia de aparición organizada de acuerdo a la jerarquía proporcionada. Las jerarquías del 1 al 10 se convierten mediante una relación de equivalencia en valores semánticos. Es decir, la jerarquía 1, que representa la más importante en cercanía al concepto, toma el valor 10, que es el máximo valor semántico con base en nuestro análisis. La jerarquía 2 toma el valor 9, de esta forma continua hasta la jerarquía 10 que toma el valor 1. Una vez dadas las equivalencias,

se calcula el valor M [23] de cada palabra definidora, mediante, la sumatoria de la frecuencia de la jerarquización por el valor semántico correspondiente. O bien:

$$M = \sum_{i=1}^{10} (Frecuencia_i \times ValorSemantico_i) \quad (1)$$

El valor M representa el peso semántico de la definidora en relación al concepto que define. Ver Tabla 1. Este valor es uno de los más importantes ya que en base a éste, la técnica de RSN tiene muchas formas de obtener distancia semántica.

Tabla 1. Valores M del grupo RSN3 para el concepto ”Algoritmo”.

Definidora	Valor
PROGRAMACION	68
SOLUCIÓN	58
INSTRUCCIONES	53
SECUENCIA	52
COMPLEJIDAD	47
LOGICA	33

Generación del Canónico General para la Ontología Natural. Este es un paso optativo, útil cuando se tienen ontologías de diferentes grupos. Una vez obtenidas las ontologías para los grupos definidos previamente, se construye el canónico general de éstas sumando el valor M de cada relación concepto-definidora encontrado.

Ontologías Obtenidas En la Figura 1, se muestra la Ontología Natural obtenida mediante el software de elicitación que implementa RSN y otro de graficación. En el grafo que se muestra cada nodo representa una palabra “definidora” y cada arista la relación entre éstas, obtenidas a partir de la información dada por los grupos de sujetos.La Tabla 2 muestra algunas métricas de las ontologías obtenidas. La columna con el valor J indica la cantidad de palabras [23], la columna ”Links” indica la cantidad de relaciones encontradas entre las palabras.

Todas las palabras y relaciones obtenidas con la metodología descrita, se pueden aplicar directamente para la anotación de ontologías mediante algún lenguaje como OWL o RDF.

2.2 Metodología para Comparación de Ontologías Naturales

Se puede efectuar la comparación de Ontologías Naturales de diferentes formas, la metodología usada para este trabajo se describe a continuación:

Reconstrucción Automática de Matriz Los datos obtenidos después del cálculo de peso semántico, nos permiten representar las relaciones concepto-definidora de la Ontología Natural en una matriz cuadrada.

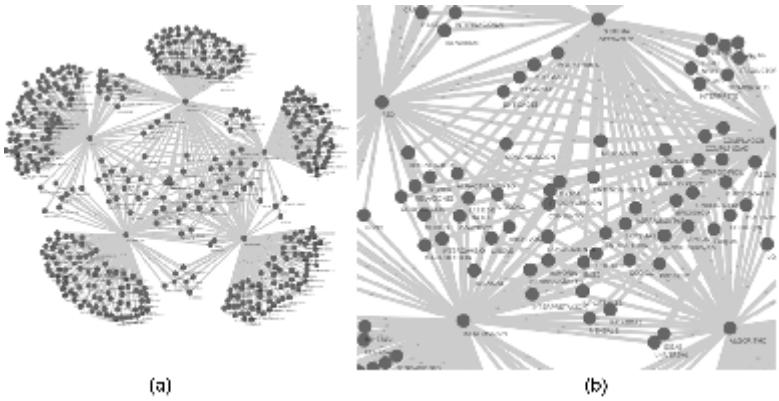


Fig. 1. (a) Visualización gráfica para la Ontología Natural obtenida del canónico formado por los grupos RSN1, RSN2, RSN3. (b) Acercamiento

Tabla 2. Número de palabras y relaciones obtenidas para cada ontología natural

Ontología	ValorJ	Links
RSN1	68	73
RSN2	374	471
RSN3	172	212
Canónico	471	617

Cálculo de Eigenvalores Los Eigenvalores son conocidos como una forma poderosa de representar matrices y con ellos se puede hacer comparación de tipo elástica, ya que el vector de *Eigenvalores* obtenido es una representación en R1 mucho más fácil de manejar que una matriz completa. Se calcularon los 6 Eigenvalores con mayor magnitud, como se muestra en la Tabla 3. [16]

Table 3. *Eigenvalores* para las matrices de cada ontología.

Canónico	RSN1	RSN2	RSN3
38.53132229	-0.000957142	21.6079831	5.916079783
-24.28726989	-0.000957142	-13.93788685	-5.916079783
-24.28726989	0.000957142	-7.67009625	-2.19342E-14
10.04321749	0.000957142	-9.04058E-05	-2.19342E-14
3.74373E-14	1.55378E-16	4.52029E-05	-5.15254E-16
3.74373E-14	1.55378E-16	4.52029E-05	-5.15254E-16

Es necesario utilizar todo el lexicón usado en las Ontologías para construir las matrices que nos permitirán compararlas dónde cada [i][j]ésimo elemento representa el valor M de la relación concepto-definidora. De tal forma que se obtuvieron 4 matrices esparcidas de 471x471.

Cálculo de Distancias Se obtuvo el cálculo de distancia entre los Eigenvalores correspondientes a cada Ontología Natural. Se pueden aplicar diferentes medidas, para este trabajo las distancias calculadas automáticamente fueron Distancia Chebyshev (2), Distancia Euclideana al Cuadrado (3), Distancia Minkowski (4):

$$dis(U,V) = \max\left(|u_i - v_i|_{i=1}^{n}\right) \quad (2)$$

$$dis(U,V) = \sum_{i=1}^{n}(u_i - v_i)^2 \quad (3)$$

$$dis(U,V) = \sqrt[\lambda]{\sum_{i=1}^{n}|u_i - v_i|^{\lambda}} \quad (4)$$

Donde:

$U =$ Muestra 1
$V =$ Muestra 2
$u =$ Valor de la variable i en la muestra U
$v =$ Valor de la variable i en la muestra V
$n =$ Número total de variables.
$\lambda =$ Órden.

Las Tablas 4, 5 y 6 muestran los resultados del cálculo de distancias para comparar las matrices de las Ontologías Naturales. Cada métrica de distancia nos proporciona diferente información, no obstante todas coinciden en que los valores de distancia para las matrices más parecidas son menores, y viceversa.

Tabla 4. Distancia Chebyshev

Chebyshev	**Canónico**	**RSN1**	**RSN2**	**RSN3**
Canónico	0	38.532279	16.923339	32.615243
RSN1	38.532279	0	21.60894	5.917037
RSN2	16.923339	21.60894	0	15.691903
RSN3	32.615243	5.917037	15.691903	0

Tabla 5. Distancia Euclideana al Cuadrado

Euclideana^2	**Canónico**	**RSN1**	**RSN2**	**RSN3**
Canónico	0	2765.3265	770.50763	2091.9924
RSN1	2765.3265	0	720.02937	70.000004
RSN2	770.50763	720.02937	0	369.4156
RSN3	2091.9924	70.000004	369.4156	0

Tabla 6. Distancia Minkowski

Minkowski	**Canónico**	**RSN1**	**RSN2**	**RSN3**
Canónico	0	44.289372	22.60896	38.3119
RSN1	44.289372	0	23.662343	7.453794
RSN2	22.60896	23.662343	0	16.905289
RSN3	38.3119	7.453794	16.905289	0

Como se puede observar en el paso 1, se obtiene una matriz global de todos los conceptos generados. Estas matrices esparcidas son la base para encontrar, dados los grupos de referencia que se utilizaron, por un lado, cuales son los conceptos con más altas relaciones y, por otro lado, permiten comparar las matrices de resultados de los grupos de expertos con los no expertos.

En la matriz de cada grupo están todas las relaciones entre conceptos dadas por los sujetos. Y dado que se aceptan los diferentes grados de conocimiento que tienen estos sujetos por su origen, entonces se puede ver que la técnica permite encontrar que tan diferentes son entre sí. Lo cual es muy importante para encontrar acuerdo entre expertos y conocer cómo la estructura de red es diferente en la de no expertos.

Una forma más específica de ver la diferencia entre los grupos es el uso de tres medidas diferentes de distancia, lo que nos ayuda a demostrar la diferencia entre grupos tanto en el espacio euclidiano como en el de Minkowski y también nos indica la posible operación de la red en un espacio n-dimensional.

3 Discusión

La metodología presentada hace un énfasis en obtener la información directamente de especialistas en el tema y estos datos se obtienen en forma sistemática y con diferentes restricciones lo cual elimina la posibilidad de textos ambiguos o de tipo narrativo y hace énfasis en la restricción de conceptos que son las "definidoras" que dan los sujetos. Lo más importante de esta técnica y que se ha demostrado en muchos trabajos es la repetibilidad sistemática de definidoras.

El caso más extremo de demostración del uso de conceptos para definir fenómenos de alta complejidad es el trabajo de Dravnieks sobre olores presentado en la revista Science[8]. Así mismo existen muchos trabajos en dónde las respuestas generadas por los sujetos son altamente consistentes y repetibles en diferentes situaciones [24,22,17].

El aspecto central de la organización de los conceptos dados por los sujetos es el hecho de la alta consistencia que tienen los pesos semánticos generados por éstos, donde es muy fácil ver que hay un grupo de definidoras que es constante y repetitivo. Estos conceptos altamente usados como definidoras son la base para encontrar las redes de relación.

Usar la técnica de RSN como método de elicitación para la anotación de conocimiento en la construcción de ontologías, asegura que se está representando el significado mediante los conceptos y relaciones entre éstos, basándose en las teorías modernas del significado además de estar sólidamente demostrada por las ciencias cognitivas.

Es un método que obtiene el significado directamente de humanos, logrando encontrar consenso y proporcionando estructura y valores semánticos. Estos valores semánticos pueden procesarse para proporcionar mayor información, como distancia entre conceptos, y a su vez distancia entre el conocimiento de grupos.

La representación en grafo, aunque da una idea visual de las relaciones, es incompleta debido a que las longitudes de los links en el esquema sólo son una aproximación respecto a su peso semántico. No obstante, sus datos son fácilmente representados en una matriz esparcida cuadrada, en la que se pueden identificar las relaciones y sus pesos semánticos. A su vez esta matriz puede ser reducida a sus eigenvalores, y de esta forma aplicar diferentes técnicas para calcular distancia, y de esta forma comparar conocimiento humano.

La ontología obtenida fue desarrollada mediante un software que implementa esta metodología. Es un modelo de elicitación y representación plausible computacionalmente, que puede ser de gran utilidad en el objetivo de interoperabilidad semántica planteado por la Semantic Web.

4 Conclusiones

La metodología de elicitación y comparación de Ontologías Naturales propuesta en este trabajo, es una aproximación que puede ser de utilidad para las tecnologías destinadas a la interoperabilidad semántica; ya que además de generar una representación del significado manejado por humanos basada en una sólida teoría cognitiva (RSN), es fácilmente automatizable, tal como se demostró mediante la utilización del software para obtener la Ontología ejemplificada.

La técnica RSN permite identificar los conceptos, sus relaciones y pesos semánticos que usa una comunidad. Su estructura se forma directamente de los datos proporcionados por grupos de sujetos, permitiendo normalizar y comparar el conocimiento. Pueden ser la base para la construcción y/o anotación de ontologías.

Referencias

1. Tim Berners-Lee, Jim Hendler, and Ora Lassila. The semantic web. Scientific American, 284(5):34–43, 2001.
2. R. J. Brachman. What is-a is and isn't: An analysis of taxonomic links in semantic networks. IEEE Computer, 16(10):30–36, 1983.
3. R.J. Brachman. On the epistemological status of semantic networks. In Associative Networks: Representations and Use of Knowledge by Computers. Findler, N. V., 1979.
4. C. Bustillo-Hernandez y J. G. Figueroa. Procedimiento para la extracción y normalización de conocimiento en humanos: Una aproximación para la anotación de ontologías en la semantic web. Memorias ROC&C 2009, IEEE Sección México., 2009.
5. Allan M. Collins and M. Ross Quillian. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8:240–247, 1969.
6. M. Cristani and R. Cuel. A survey on ontology creation methodologies. International Journal on Semantic Web and Information Systems, 2005.
7. Yang Dongqiang. Lexical Semantic Similarity: Word Similarity in Semantic Networks and Distributional Structures. Verlag Dr. Müller, 2009.

8. Andrew Dravnieks. Odor quality: semantically generated multidimensional profiles are stable. Science, 218:799–801, 1982.
9. Marc Ehrig. Ontology Alignment: Bridging the Semantic Gap. Springer, 2007.
10. Christiane Fellbaum, editor. WordNet: An Electronic Lexical Database. MIT Press, 1998.
11. Jesús G. Figueroa, Esther G. González, and Victor M. Solís. An approach to the problem of meaning: Semantic networks. Journal of Psycolinguistic Research, 5(2):107–115, 1976.
12. A. Hammed, D. Sleeman, and A. Preece. Detecting mismatches among experts ontologies acquired through knowledge elicitation. Elsevier Knowledge Based Systems, 15:265–273, 2002.
13. K. Jung-Min, Byoung-II C., S. Hyo-Phil, and K. Hyoung-Joo. A methodology for constructing of philosophy ontology based on philosophical texts. Elsevier Computer Standards and Interfaces, 29:302–315, 2007.
14. Y.V. Kapitonova. General principles of construction of knowledge computer systems. Cybernetics and Systems Analysis, 42:531–546, 2006.
15. David E. Meyer and Roger W. Schvaneveldt. Meaning, memory structure and mental processes. Science, 192:27–33, Abril 1976.
16. E.V. Ortega-Gonzalez. Una técnica para el análisis de similitud entre imágenes. Tesis de Maestría. Centro de Investigación en Computación. Instituto Politécnico Nacional., 2007.
17. V. A. Ramírez, A. Zermeño, y A. C. Arellano. Redes semánticas naturales: Técnica para representar los significados que los jóvenes tienen sobre televisión, internet y expectativas de vida. Estudios sobre las culturas contemporáneas, 22:305–334, 2005.
18. John Sowa, editor. Principles of Semantic Networks. The Morgan Kaufman Series in Representation and Reasoning. Morgan Kaufmann Publishers, Inc., 1991.
19. S. Staab and R. Studer, editors. Handbook on Ontologies. International Handbooks on Information Systems. Springer, 2 edition, 2009.
20. J. Gamboa Suasnavart, J. Osorio Reséndiz, R. Lom Romero, O. Tapia Leyva, E. Vargas Medina, y J. Figueroa Nazuno. Las redes semánticas naturales como procedimiento para extracción de conocimiento, para su uso en interoperabilidad semántica. Memorias ROC&C 2008, IEEE Sección México., pag. 42, 2008.
21. Endel Tulving and Daniel L. Schacter. Priming and human memory systems. Science, 247:301–306, 1990.
22. J. L. Valdez y Reyes L. Las categorías semánticas y el autoconcepto. Revista Psicología Social en México, IV:193–199, 1992.
23. J. L. Valdez-Medina. Las redes semánticas naturales, uso y aplicaciones en psicología social. Universidad Autónoma del Estado de México, 2004.
24. E. Vargas-Medina y C. Calzada-Ugalde. Evolución de la representación conceptual de la física en estudiantes universitarios y preuniversitarios. Revista del Centro de Investigación, Universidad La Salle, 1(2):41–61, 1994.

Which Algorithm and Approach for Arabic Part Of Speech Tagging

Mourad MARS[1,2], Georges Antoniadis[1], Mounir Zrigui[2]

[1] Lidilem Laboratory, University of Grenoble,
BP 25, 38040 Grenoble cedex 9, France
[2] UTIC laboratory, University of Monastir, Tunisia
Mourad.mars@e.u-grenoble3.fr
Georges.antoniadis@u-grenoble3.fr
Mounir.zrigui@fsm.rnu.tn
Paper received on 06/08/10, Accepted on 05/10/10

Abstract. This paper presents a recent study aiming to develop a Part Of Speech (POS) tagger for Arabic language. Part-of-speech (also known as POS, Word classes, morphological classes, or lexical tags) of Arabic is an active topic of research in recent years, it is the process of assigning to each sequence of words the right POS tags. The paper is organized as follows. In the first section, an overview of recent work and a briefly introduce to different configurations and tested approaches in POS tagging. In the following section, we describe the different kinds of languages models with various smoothing techniques, as well as several resources and tested approaches to justify the methodology and the appropriate setting used to build our POS tagger. We also tested and evaluate the robustness of the tagger. Furthermore, conclusions, open areas, and future directions are provided at the end.

Keywords: Part-Of-Speech tagging, POS tagger, Arabic, disambiguation, HMM, Algorithm, Language Model.

1 Introduction

Arabic is one of the Semitic languages; known for its rich and systematic but very complex morphological structure and also for its several problems in natural language processing (agglutinative form, run-on word, free concatenation and orthographic variation) [21]. Part-of-speech (also known as POS, Word classes, morphological classes, or lexical tags) of Arabic language is an active topic of research in recent years, it is the process of assigning to each sequence of words the right POS tags.

Many computational methods and algorithms for assigning POS to words have been used including rule-based tagging (hand-written rules), statistical methods (HMM tagging and maximum entropy tagging) as well as other methods like transformation-based, memory-based tagging and also hybrid method for tagging.

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 235 -245

Before turning to present our approach in POS tagging and our tests, let's begin with an overview of recently published work.

Table **1.** Sample of Arabic tagged.

Arabic sentence	كتب جميلة .
Transliteration	ktb jmylp .
POS tag	NN Adj PUNCT
Meaning	Nice books.

2 Recent Work on POS Tagging: A Brief Overview

In this section, a briefly overview of different POS taggers for Arabic language recently achieved is presented. Khoja [12] combines both statistical and rule-based techniques and uses a tagset of 131 derived from the BNC tagset. Maamouri and Cieri system [14] achieved an accuracy of 96%, it consist of the automatic annotation output produced by AraMorph, which is the morphological analyzer of Tim Buckwalter [4]. Yahya O. Mohamed Elhadj [21] presented an Arabic POS tagger that uses a HMM model, the evaluation shown an accuracy of about 96%. Freeman's tagger [7] uses a machine learning (ML) approach as in Brill's POS tagger [3], a tagset of 146 tags extracted from the Brown corpus for English, is used. Mona Diab et al. [6] use Support Vector Machine (SVM) method to build their tagger for Arabic and the LDC's POS tagset with the number of 24 tags. Fatma al Shamsi and Guessoum [2] HMM POS tagger has been tested and has achieved a state-of-the-art performance of 97%. Tlili Guiassa [22] uses a hybrid which consists of combining based-rules and memory-based learning methods. This system reports a performance of 85%.

3 Part Of Speech Tagging

What is the best sequence of tags which corresponds to a given sequence of words? The use of a POS tagger can answer this question; it attempts to assign the correct tag or lexical category to all words in a text. This is an example of a tagged text (POS tagger output):

ktb/VB rsAlp/NN ./PUNCT (1)

ktb/NN jmylp/ADJ ./PUNCT (2)

In both sentences, automatically assigning one tag per word is not easy. For example, ktb is ambiguous. That is, it has more than one possible tag. It can be a

verb (write) as in (1) "Write a letter" or a noun (books) as in (2) "Nice Books". The objective of POS tagging is to resolve these ambiguities, choosing the proper tag for the word.

This table reports the problem of ambiguity, tag number per word for ambiguous corpus [5], [10].

Table 2. The amount of tag ambiguity for word types in a ≈ 135000 words corpus.

Tag per word	% in Arabic corpus
Unambiguous (1 tag per word)	24.15%
Ambiguous (more than 1 tag per word)	24.15%

Most tagging algorithms fall into one of two classes : Rule-Based taggers [3], [1] which consists of coding the necessary knowledge in a set of hand written disambiguation rules database (3) or Stochastic Taggers which resolve problem by the use of a training corpus to calculate probability of a given words and tags.

VB VB - Rule
Given tags: Vb or NN

if(-1 is Vb); /*if previous tag is Verb*/ (3)

then eliminate Vb tags

Most experimental results, in Arabic and other languages, demonstrate that statistical theories are the commenly used methods in order to obtain the best accuracy and precision in POS tagging.

Use of a Hidden Markov Model (HMM) to Part Of Speech tagging is a special case of Bayesian inference. POS tagger tries to choose the tag sequence t_1^n which is most probable given the observation sequence of n words w_1^n [24]

$$t_1^n = \arg\max P\ \ t_1^n / w_1^n\ \ . \tag{4}$$

We apply Bayes' rule, (4) will be transformed to (5), a simpler formula easier to compute.

$$t_1^n = \arg\max P\ \ w_1^n / t_1^n\ \ P\ \ t_1^n\ \ . \tag{5}$$

HMM taggers make other simplifying assumption is that the probability of a tag appearing is dependent on the n-1 previous tags. To summarize, the most probable tag sequence t_1^n given some word w_1^n can be computed by the following equation:

$$t_1^n = \arg\max P\ \ t_1^n / w_1^n\ \ \approx \arg\max \prod_{i=1}^{n} P(w_i / t_i) P(t_i / t_{i-1}, t_{i-2}, ..., t_{i-n+1}). \quad (6)$$

3.1 Part of Speech Data

The achieved POS tagger gets its input from our morphological analyzer [15], [16], [17] module which segments lexical units and gives, for each surface form, all informations consisting of lemma, lexical category and morphological inflection information.

To calculate all statistical parameters for our POS tagger, we make use of a hand-tagged corpus of Arabic sentences collection. We start by dividing the data into a training set which is approximately about 120,000 words (TrSet≈2.2Mo) and a test set (TestSet≈38ko). The training consists in estimating two types of parameters: lexical P(W/T) saved in Map file and contextual P(Ti/Ti-1,…,T-n+1) saved in language model (LM) file.

3. Lexical Model

Due to data sparseness and in order to overcome this problem, it is necessary to introduce smoothing techniques, while using with HMM [11]. All experiments illustrate that overall accuracy of the tagger improves significantly with the introduction of smoothing techniques. Thus we find that the HMM together with different smoothing methods shows improved results than ordinary Hidden Markov models.

Smoothing methods have been proposed and described in the literature, these methods aimed at re-evaluating and harmonizing the probabilities of words which are quite unlikely to occur and then assign them appropriate non-zero probabilities [11], [23]. These methods adjust downward high probabilities. Two smoothing methods was tested, Good-Turing and Chen and Goodman's modified Kneser-Ney discounting [8] [9], [13], [18], [19].

Using the same training set (TrSet) and under the identical test environment, we tried to compare different smoothing methods for different size of training corpus in order to select the best method and get which techniques improve with more data, and which get worse. The comparison results of different smoothing techniques are shown that all technique improves with more data and the Chen and Goodman's modified Kneser-Ney smoothing method gives best result.

3.3 Contextual Parameters (Language Model)

For our generative modeling approach, the first step consists of creating an N-gram language model from the tagged corpus. The estimation of the n-gram probabilities was carried out using the SRI Language Modeling toolkit1[1] [20] for language models (LM), we compare Chen and Goodman's modified Kneser-Ney discounting, as implemented by the ngram-count tool, and Good-Turing smoothing method. The LM is given in ARPA format. We measure the size of language models in total number of n-grams which is the sum of all n-gram orders [19] (n from 1 to 3).

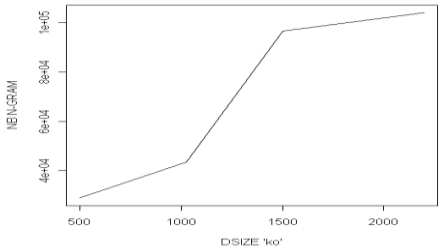


Fig. 1. Number of n-grams for varying training data size.

Now, we first describe and justify our use of perplexity as an evaluation technique. We then describe the experimental techniques used, in POS tagging, in the following sections.

The performance evaluation of a language model is usually based on the probability the language model assigns for an independent test text. The most commonly used method for measuring language model performance is perplexity which is equal to the geometric average of the inverse probability of the words measured on test data [8]:

$$\sqrt[n]{\prod_{i=1}^{n} \frac{1}{P(w_i / w_1 \ldots w_{i-1})}}. \quad (7)$$

[1] SRILM*:* SRI Language Modeling Toolkit is a toolkit for building and applying statistical language models (LMs)

Perplexity has many properties that make it attractive as a measure of language model performance; among others, the "true" model for any data source will have the lowest possible perplexity for that source. Thus, the lower the perplexity of our model, the closer it is, in some sense, to the true model.

We measured perplexity for different tested smoothing techniques using various training data size and our results show that perplexity is lower with Chen and Goodman's modified Kneser-Ney, Namely when the data size is high.

In general, smoothing is an important factor in language modeling especially when the data size is too small and we haven't enough corpora; it is the case of Arabic language [14].

3.4 Viterbi Algorithm

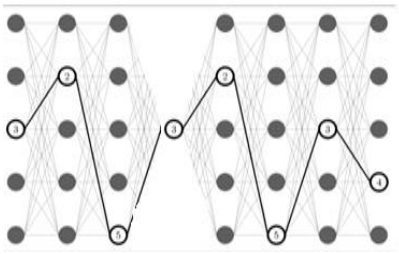
Having computed the transition and emission probabilities and assigning all possible tag sequences to all the input words, now we are in the situation to construct a lattice showing association of different tags to the input words, we need an algorithm that can search the tagging sequence and maximize the product of transition and emission probabilities. For this purpose we use Viterbi algorithm, a dynamic programming process, which compute the maximized as well as optimal tag sequence with best score.

3.5 Disambiguation Approach

In this work, a combination of statistical and linguistic approaches is employed, and our approach is to use variable windows. In other way, the text sequence $w_1 \ldots w_n$ being disambiguated is always boundered by the beginning and the end of a sentence. We disambiguate each sentence separately, because as in other languages, the sentence structure is independent from the previous and the next sentence, so there is no grammatical relation between the last word of a sentence and the first one of the upcoming sentence. Computing probabilities for these sequences have no meaning and can reduce POS tagger performance. Other point, our POS Tagger divide sentences to sub sentences delimited by unambiguous words having only one tags (see figure 2). The best sequence of tags for each sub sentence is finded by a modified classical use of viterbi algorithm. For each sequence, we try to maximize this probability (n=3):

$$\arg\max \prod_{i=1}^{n} P(w_i / t_i) P(t_i / t_{i-1}, t_{i-2}, ..., t_{i-n+1}). \quad (8)$$

This is a sample of a sentence having one unambiguous word.



Sub sentence1 Sub sentence2

Fig. 2. Sentence decomposition.

3.6 Architecture of POS Tagger

The diagram hereafter gives the architecture of our POS tagger for Arabic language. morphological system of analysis of Arab texts. The system comprises three modules: the morphological analyzer making it possible to assign with each lexeme one or several labels, the theory of probability which make it possible to assign a probability with each potential sequence of labels, and modulate it clarification making it possible to calculate the most probable sequence of labels for each sequence of the text to be analyzed.

In support of this and other works, we have described an architecture for our POS tagging system for Arabic language [16]. This architecture defines two components for disambiguation systems:

Training, in which we compute HMM parameters from an unambiguous tagged corpus (lexical model and Language model). These parameters are respectivly saved in Map file and LM file.

POS Tagger, it is implemented as an analysis module. Figure 2 illustrates the overall architecture; it attempts to choose the correct tag or lexical category, from a list of tags, for each words in a given text. This operation is achieved by the use of Viterbi Algorithm and all needed files like Map file and LM file computed in traning level.

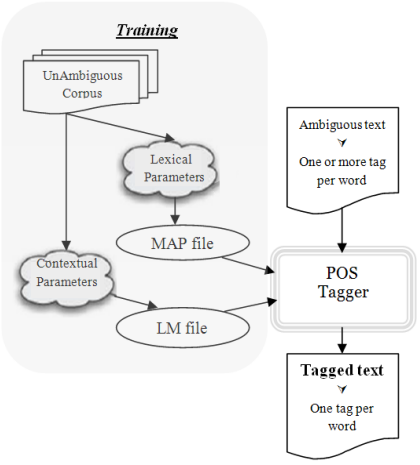


Fig. 3. Our Arabic POS tagger architecture.

3.7 Evaluation

Taggers are often evaluated by comparing them with a hand tagged test set, based on accuracy. We have prepared and tagged a test corpus (TestSet). Experimental results have been evaluated through standardize formulation: Precision, Recall and F-measure [8].

- *Precision= Correct number of token tag pair occurrence / Total number of token tag pair.*
- *Recall= Correct number of token tag pair occurrence / Number of correct token tag pair that is possible.*

$$F = 2 * \frac{precision * recall}{precision + recall}.$$

The proposed POS tagger has been tested and has achieved a state of the art performance of 96% which is very encouraging.

A confusion matrix contains information about correct and predicted tag done by the POS tagger. Performance of such systems can be represented using the following matrix. The following table shows a part of the confusion matrix.

Table3. Part of the confusion matrix

Tag	VB	NN	DT
VB	98.20%	01.40%	00.80%
NN	01.38%	92.30%	04.32%
DT	01.51%	3.39%	94.10%

4 Conclusion and Future Work

To summarize, POS tagging of Arabic language is not an area for the faint of heart or easily depressed. It is sometimes very difficult even for human to identify the correct tag for a given word. A great work was done and our results compare favorably and very promising to other recently reported works results.

As future work, we will try to proceed to test disambiguation in two steps, or in hierarchic manner, to give more information for each word. We disambiguate the first level of tags and then the second. For example, for the word "ktb" it can be a VB, as word class, while it can be also a NN. The first level of disambiguation will disambiguate the word in its context and return us "ktb" as VB. The second steps decide if this VB is in the Past or in its Passive Form. This choice is due to the use of our morphological analyzer in platform of language learning for Arabic which needs largest informations about words in text to allow the automatic generation of pedagogical activities.

Other future work is to combine taggers in series (rule-based-tagger and stochastic tagger) by using the rule-based approach to remove some of the impossible tag possibilities for each word, and then the HMM tagger to choose the best sequence from the remaining tags. Other possibility is to use the HMM tagger to choose the best tag for each word in corpus, or multiple tags in certain cases. Carrying some ambiguity from one processing step into the next, in order not to prune good solutions, will be performed by the use of a confidence interval, and then we use rules to keep only one tag per word.

The proposed approach can also be applied to other NLP processing. We estimate that this additional work we planned will improve our system's performance even more.

Acknowledgments. This work is part of three large research projects, the first one titled "Oreillodule" which is a tool for automatic speech recognition, translation, and synthesis for Arabic language. The others are "MIRTO" & "@rabLearn": two Intelligent Computer Assisted Language Learning (ICALL) platforms. The integration of our POS tagger in both ICALL systems will help teachers to generate, automatically, pedagogical activities and scenarios for learner and to migrate towards new generation of intelligent feedback.

Glossary

Corpus, pl. corpora. A collection of text, typically representing either a random sample from numerous genres, or a large archive of text from a single source (such as the archives of a single newspaper).

Hidden Markov Model (HMM). A stochastic finite-state model whose output is known, but for which the state sequence that produced is unknown.

Informations theory. A branch of mathematics that quantifies information's content of communications, with applications that include compression, error correction, and encryption.

Language model. A statistical model defining a probability distribution over the word sequences of a language.

Smoothing. Statistical estimation technique that are used when the training data is insufficient for estimating all parameters of a large model.

Supervised (unsupervised) learning. In supervised learning, the learning algorithm is provided with training data that has been manually labeled with the correct answers. In unsupervised learning, only unlabeled data is provided.

Viterbi algorithm. An efficient algorithm which, given the output sequence produced by an HMM, computes the most-likely sequence of states that the HMM passed through.

References

1. Ahmad, A., Salah, A.: A Rule-Based Approach for Tagging Non-Vocalized Arabic Words: In The International Arab Journal of Information Technology, Vol. 6, No. 3. (2009)
2. Fatma, A., Ahmed, G.: A Hidden Markov Model –Based POS Tagger for Arabic: In JADT&06. (2006)
3. Eric, B.: Transformation-based error-driven learning and natural language processing: a case study in part of speech tagging: In Computational Linguistics 21, pages 543-565. (1995)
4. Tim, B.,: Arabic Morphological Analyzer Version 1.0. LDC: University of Pennsylvania. (2002)
5. Debili, F., Emna, S: Etiquetage grammatical de l'arabe voyellé ou non: In Proceedings of the Workshop on Computational Approaches to Semitic Languages, Montreal, Quebec, Canada. (1998)
6. Mona, D., Kadri, H., Daniel, J.: Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks: In HLT-NAACL, pages 149-152. (2004)
7. Freeman, A.: Brill's POS tagger and a morphology parser for Arabic: In ACL'01 Workshop on Arabic language processing. (2001)
8. Goodman, J.: A bit of progress in language modeling: Extended Version, Microsoft Research Technical Report MSR-TR-72. (2001)
9. Jelinek, F.: Aspects of the Statistical Approach to Speech Recognition: In IEEE International Symposium on Information Theory, Washington D.C. (2001)
10. Karin, C.R.: A Reference Grammar of Modern Standard Arabic: Cambridge University Press, ISBN: 0521777712, 736 pages. (2005)

11. Katz S.M.: Estimation of probabilities from sparse data for the language model component of a speech recognizer: In IEEE Transactions on Acoustics, Speech and Signal Processing, volume ASSP-35, pages 400-401. (1987)
12. Khoja, S.: APT: Arabic part-of-speech tagger: In proceeding of the Student Workshop at the 2nd Meeting of the NAACL, (NAACL'01), Carnegie Mellon University, Pennsylvania. (2001)
13. Kneser R., Ney H.: Improved backing-off for n-gram language modeling: In Conference on Acoustics, Speech and Signal Processing, volume 1, pages 181-184. (1995)
14. Maamouri, M,. Ann, B.: Developing an Arabic treebank: Methods, guidelines, procedures, and tools: In Proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages (COLING), Geneva. (2004)
15. Mourad, M., Mohamed, B.: Development Of A Morphological Analyzer For Arabic Language, Tool For Creation Of Educational Activities for Training Of Arabic: In International Conference in Technology Enhanced Learning (TEL) in Working Context, Grenoble France. (2006)
16. Mourad, M., Mounir, Z., Mohamed, B., Anis, Z., Georges, A.: A Semantic Analyzer for the Comprehension of the Spontaneous Arabic Speech: In 9th International Conference on Computing CORE08, Journal Research in Computing Science (Journal RCS), ISSN: 1870-4069, Vol 34, pp 129-140, CORE0, Mexico. (2008)
17. Mourad, M., Georges, A., Mounir, Z.: Nouvelles ressources et nouvelles pratiques pédagogiques avec les outils TAL: In TICEMED 08, Journal of Information Sciences for Decision Making (ISDM Journal), ISDM32, N°571. (2008)
18. Ney, H., Essen, Kneser, R.: On structuring probabilistic dependencies in stochastic language modeling: In Computer Speech and Language, volume 8(1), pp. 1-28. (1994)
19. Rosenfeld, R.: Two decades of Statistical Language Modeling: Where Do We Go From Here? In Proceedings of the IEEE, 88(8) (2000)
20. Stolcke, A.: SRILM - An Extensible Language Modeling Toolkit: in Proceeding of International Conference in Spoken Language Processing, Denver, Colorado. (2002)
21. Yahya, O., Mohamed, E.: Statistical Part-of-Speech Tagger for Traditional Arabic Texts: Journal of Computer Science 5 (11): 794-800,. (2009)
22. Yamina, T.G.: Hybrid Method for Tagging Arabic Text: Journal of Computer Science 2 (3): 245-248, ISSN 1549-3636. (2006)
23. Witten, I.H., Bell, TC: The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression: IEEE Transactions Information Theory, vol. 34, number 4, pp. 1085-1094. (1991)
24. Waqas, A., Xuan, W., LuLi, X.: Hidden Markov Model Based Part of Speech Tagger For Urdu: Information Technology Journal Vol: 6, 1190-1198. (2007)

A vertical handover mechanism with security services for mobile applications

Juan A. Vargas Enríquez [1] and Arturo Díaz Pérez[2]
[1] Instituto Tecnológico de Cd. Victoria,
Blvd. Emilio Portes Gil # 1301 Pte. C.P. 87010
Cd. Victoria, Tamaulipas
Jvargd@itcv.edu.mx
[2] CINVESTAV, Unidad Tamaulipas
Laboratorio de Tecnologías de Información
Parque Científico y Tecnológico TECNOTAM
Km. 5.5 carretera Cd. Victoria-Soto La Marina
Cd. Victoria Tamaulipas, México
Paper received on 26/08/10, Accepted on 15/09/10.

Abstract. A mechanism, called vertical handover, for changing networks on mobile devices is presented. Particularly, we explain how to do such changes in cell phones owning Wi-Fi controllers. The proposed mechanism enables applications accessing the Internet on a cell phone to automatically switch between WLAN and GPRS networks. Additionally, in order to assess the performance of the handover mechanism, it is presented an application that transfers data through a secure channel via Internet from a cell phone to a desktop computer that acts as an application server. The secure channel provides confidentiality, integrity and authentication services. The tests conducted showed that the handover mechanism makes network changes in the three possible scenarios: GPRS to WLAN, GPRS to WLAN and WLAN to WLAN.

1 Introduction

Currently the use of the cell phone is not just limited to making phone calls. According to a survey conducted in the United States by America On Line in 2006 [1], 8% of the users use their cell phone to access the web, and 7% use it to send email. The integration of Wi-Fi to the cell phone has become decisive when it comes to accessing the Internet due to its low cost and higher data rates compared to GPRS (General Packet Radio Service) of GSM (Global System for Mobile communications). The number of cell phones with a Wi-Fi interface is growing. According to the report "Wi-Fi Component forecast and Vendor Share" of the 2005 [2], in 2009 most of the Wi-Fi chips had been installed in cell phones. On the other hand, wireless networks are widely available in offices and public places such as hotels, airports, malls and schools [3]. The integration of these technologies is needed in order

©E.Cuatecontzi Cuahutle, A.Cortés Fernández,
J. F. Ramírez Cruz, J. H. Sossa Azuela. (Eds.).
Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 249-259

to allow mobile devices to roam seamlessly between WLAN and GPRS networks, a process called vertical handover [11].

In this paper we propose a vertical handover mechanism between WLAN and GPRS networks for mobile devices. In order to evaluate the performance of the handover mechanism it was developed an application in the health care area for patient monitoring. The application for services in the area of health consists of a cell phone with Wi-Fi interface and a sensing device, the system monitors the patient's vital signs and sends this information to a medical facility through the GPRS network or through a wireless network if available. Handover events occur when the patient enters or leaves the coverage areas of wireless networks. We also propose a security mechanism to provide authentication, data integrity and confidentiality services in order to establish a secure communication channel [12] between the cell phone and the application server.

The remainder of this paper is organized as follows. In section 2 the system architecture is presented and a description of the modules that comprises is given. Section 3 discusses the implementation of the architecture. Section 4 discusses the security services implemented. Sections 5 and 6 discuss the performance of the system and the analysis of results. Section 7 presents the conclusions drawn.

2 System architecture

The proposed architecture for patient monitoring system is shown in Fig. 1. The architecture follows the client server model, where the client is the application that runs on the cell phone and the server is the application server for Internet communications.

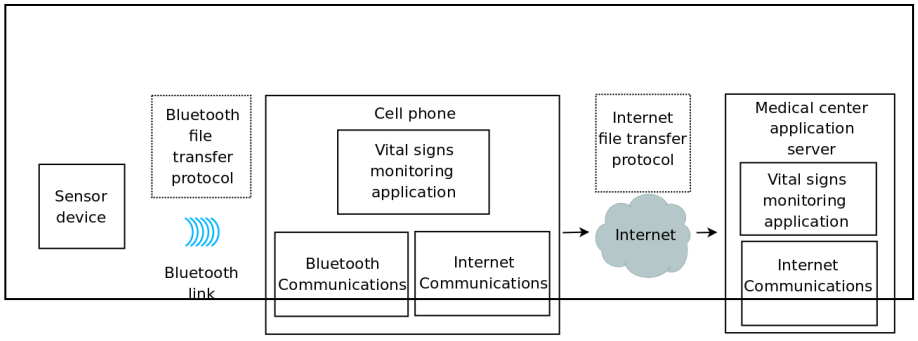


Fig. 1. Vital signs monitoring system architecture

The client architecture consists of three modules which are described below.

- Bluetooth file transfer module: The function of this module is to transfer data stored in the internal memory of the monitor device to the cell phone.
- Data display module: The function of this module is to present in the cell phone data received from the monitor device.
- Internet file transfer module: The function of this module is to transfer the data file from a cell phone to the application server providing the security services of authentication, confidentiality and integrity. This module also provides the handover mechanism to make the automatic change between GPRS and WLAN networks.

The server architecture consists of four modules which are described below.

- Internet file transfer module: This module receives the data file sent from the cell phone using the developed file transfer protocol. This module also performs two other functions, the first one is to respond to authentication requests from clients and the second one is to provide session keys that require both client and server to provide integrity and confidentiality services.
- Data display module: The function of this module is to present to staff in the health care facility the data received from the monitor device.

3 Communications

The system for monitoring vital signs performs a transfer of information from the monitor device to the application server; this transfer is done through two different types of networks, Bluetooth and Internet. To make this transfer of data reliable two communication protocols are used one for Bluetooth and one for Internet.

3.1 Bluetooth file transfer

Transferring data via Bluetooth interface requires the implementation of a file transfer protocol for two reasons. Firstly, the vital signs monitor device has limited computing capabilities and has only a basic implementation of RFCOMM transport protocol. Secondly, the communications via Bluetooth present frequent errors [13] and need an error recovery mechanism additional to that provided by RFCOMM. When messages are lost or when communication errors occur, the error recovery mechanism of the protocol stops completely the file transfer to enable the cell phone

and the monitor device to synchronize again. Once the devices are synchronized, the file transfer is resumed from the beginning.

3.2 Internet file transfer

To make the file transfer between cell phone and the application server through the Internet a protocol was developed over the TCP/IP protocol. The Internet transfer protocol keeps the state of the transfer, in this way a mechanism for more efficient error recovery can be added. If the file transfer is interrupted, it resumes at the point where it was stopped. This error recovery scheme is necessary because of interruptions in the data transfer can occur more frequently in this protocol. The commands in this protocol are shown in Table 1.

Table 1. Commands of the Internet file transfer protocol

Commands issued by cell phone	Answer from server
SEND_FILE	COMMAND_RECEIVED
RESEND_FILE	Last data block received
EOF	EOF_RECEIVED
Data block	BLOCK_RECEIVED

The command SEND_FILE tells the application server a client (cell phone) wants to start a transfer session to send a file to the server.

The command RESEND_FILE is issued by the cell phone and tells the server that the file transfer failed, it also tells the server that the file transfer is going to restart at the point it was interrupted.

3.3 Vertical handover GPRS/WLAN

The vertical handover process has been extensively addressed in the formal literature and several solutions have been proposed.

In [4] Salkintzis discusses the motivations for integrating WLANs with 3G cellular networks and presents an outline for a tightly coupled architecture between WLAN and GPRS networks.

In [5] it is proposed a hybrid architecture to support vertical handover between an IEEE 802.11 network and an UMTS network (Universal Mobile Telecommunications System), which incorporates SIP (Session Initiation Protocol) and mobile IP protocol.

In [6] it is proposed a scheme of vertical handover between WLAN and GPRS networks based on routing. In addition, it is presented a model for the handover decision, which reduces the latency time of handover procedure.

In an article by Song et al [7] it is proposed a hybrid coupling scheme to support interworking between UMTS and WLAN networks, which differentiates the data paths based on the type of the traffic. For real-time traffic a tightly coupled network architecture is chosen. For non real-time traffic the loosely coupled network architecture is chosen.

In this paper, it is presented a mechanism called *vertical handover*, which allows changing networks on mobile devices, particularly cell phones that have Wi-Fi interface. The proposed mechanism enables applications accessing the Internet on a cell phone to automatically switch between WLAN and GPRS networks without restarting the data transmission from scratch. The algorithm used in the process of searching the best Internet access network is shown in Fig. 2. If there are any available wireless networks, the one with the greatest signal strength is selected, otherwise GPRS is chosen as Internet access point. The developed handover mechanism considers 3 scenarios in which there may be a network change as shown in Fig. 3:

Network change GPRS to WLAN: The network change occurs when the signal strength of the WLAN network reaches a minimum value of -75 dBm, this value was determined by tests that were performed.

Network change WLAN to GPRS: As explained before, the minimum signal strength to establish a connection is -75 dBm, when the signal strength falls below this value is considered that the cell phone is out of range of the wireless network and the handover mechanism proceeds to close the connection and makes the switch to the GPRS network.

Network change WLAN to WLAN: When the cell phone moves away from the WLAN to which is connected, the WLAN signal will be lost eventually, when this happens the handover mechanism will change the access network to the WLAN that has the strongest signal.

3.4 Handover during file transfer.

The use of a handover mechanism for file transfer is necessary so the transfer service can properly handle the loss of the network signal in use, or availability of another network with better features. The purpose of the vertical handover module is to provide the best access network to the Internet during the file transfer. The handover module is independent of the transfer protocol, is placed in a lower layer and is responsible for providing the best available communication network. The error recovery mechanism used by the file transfer protocol is initiated in the cell phone and can resume an interrupted transfer from the point of interruption.

During a file transfer several network changes may occur, these changes however occur only under two circumstances when the signal is lost or when a better network is available, in either case the handover mechanism is invoked in the same way. When the lost of the network signal that is being used in the cellular phone is detected, the transfer protocol responds with the error recovery mechanism, this in turn invokes the handover module to provide the new best communication network. When the cell phone sets the transfer session with the server, it checks if there is any

transfer in progress reviewing the value of the variable that stores the number of last block sent, if this is different from zero means the file transfer was not completed and a RESEND_FILE command is issued. When the server receives the RESEND_FILE command, sends in response to the cellular phone the number of the last block received. When the cellular phone receives the number of last block, the file transfer is initiated from the next block. The protocol performs all these operations through a series of messages exchanged between the cell phone and the application server as shown in Fig. 4.

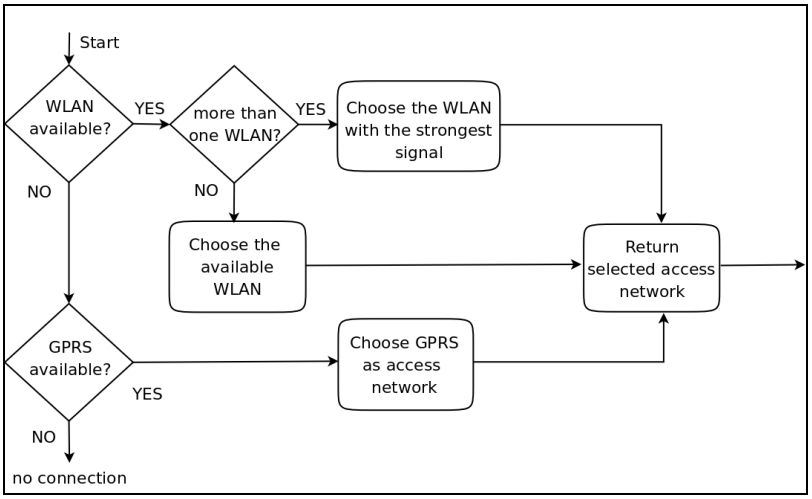


Fig. 2. Flow chart for selection of Internet access network during handover

4 Security Services

Beside the handover module, the security services module provides authentication, confidentiality and integrity services [8] to applications placed in the upper layer.

The mechanism of distribution of session keys is responsible for distributing between the client and server the session keys that are required for data encryption. In addition, the protocol for the management of session keys performs the authentication function. Under this scheme the application server acts as a distribution center of session keys and simultaneously authenticates the mobile phone at the start of a transfer session. The implemented authentication is performed only in one way, the application server authenticates the mobile phone but the cell phone does not authenticate the application server. The process of distribution of session keys assumes that both client and server share a secret key previously distributed in some way. In these implementations the life time of a session key was set in one day.

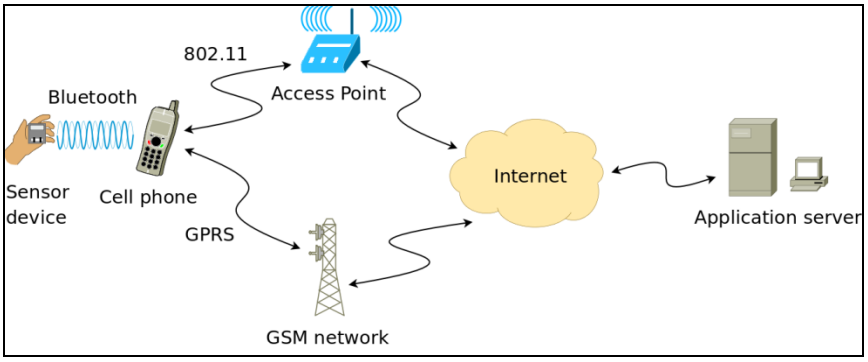


Fig. 3. Patient monitoring system scenario

Confidentiality is implemented using symmetric key cryptography. The encryption algorithm used is AES with a key of 128 bits [9]. The confidentiality service is deployed on both ends of the communication channel. However, in this application, the data only flows from the client to the server but not vice versa. Therefore the cell phone encrypts the data file before sending it to the application server using a 128 bits session key that was previously distributed during the set up of the transfer session. After the transfer of the data file is completed in the application server, the data file is decrypted using the session key to restore it to its original condition.

The integrity service is provided using the SHA-1 hash function [10]. Once the data is encrypted and transferred from the cell phone to the application server, the integrity service is implemented by calculating the SHA-1 hash function of the original data file, the digest that generates SHA-1 function is then sent to application server. In the application server, the integrity service is implemented after the data file and its corresponding digest are received from the cell phone, the data file is first decrypted and then SHA-1 function is calculated, then both digest are compared, the one that was received and the one that was calculated in the application server, if both digests are equal that means that the file was received fine, and an acknowledgment message is sent to the cell phone, if they do not match that means the file was corrupted during transfer, in this case the application server does not return any acknowledgment message and the cell phone sends the data file again.

5 Functionality testing

The client application was installed on a Nokia N91 cell phone, this cell phone has Bluetooth, GPRS and 802.11 (Wi-Fi) communication interfaces, the server application was installed on a desktop computer with Linux operating system. There were also two wireless networks available, both were registered as Internet access

points in the configuration of the cell phone. Authentication to access the wireless networks was done using MAC address filtering. Three tests were conducted to test the handover mechanism during the change of communication network, from WLAN to GPRS, from GPRS to WLAN and from WLAN to WLAN. In each test an image file in jpeg format was sent from the mobile phone to the application server.

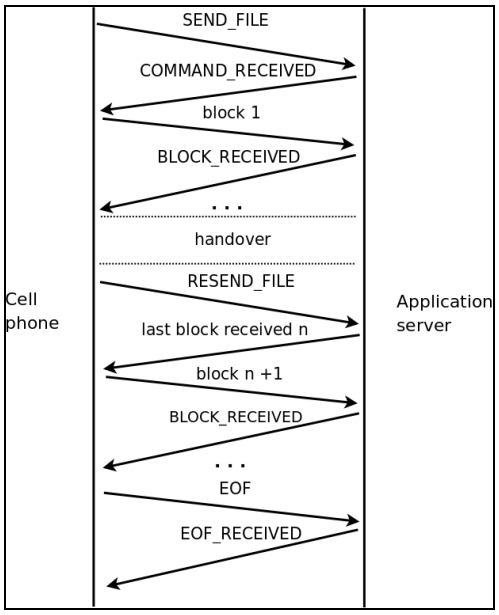


Fig. 4. Message exchange during handover

6 Performance testing

To assess the performance of vertical handover mechanism 15 tests were carried out; the performance was measured in categories such as time of handover, transfer time and file encryption and decryption time, and power consumption.

The handover time is the time it takes for the application on the mobile phone to make a change from one type of network to another. It is measured from the moment the current connection with the application server is closed until the connection is opened with the new selected network. The average times for the three possible scenarios of network change are shown in Table 2.

Table 2. Handover time in seconds

Handover type	Average time	Standard deviation
WLAN/GPRS	18.69 s	9.10 s
GPRS/WLAN	3.76 s	0.97 s
WLAN/WLAN	1.25 s	0.14 s

In the encryption tests conducted it was found that the average rate of encryption is 47.89 KB/s with a standard deviation of 1.4 KB/s, showing that the rate of encryption is kept more or less constant regardless of the size of the file being encrypted. Similar results were obtained for decryption, the decryption rate averaged 50.00 KB/s with a standard deviation of 2.29 KB/s, it was observed that the rate of decryption is slightly higher for files whose size is greater than 2 MB.

To determine the time of a file transfer and the data transfer rate when using a WLAN, six files whose sizes ranged from 128 KB to 5 MB were transferred from cell phone to the application server.

To determine the overhead imposed by security services to the file transfer it was performed the same procedure as in the previous test but the time measurements were taken considering the time involved in making the operations of authentication, file encryption and hash function calculation. The results of this test are shown in Table 3.

Table 3. File transfer times and data transfer rates using a WLAN

	128 KB	256 KB	512 KB	1 MB	3 MB	5 MB
Average transfer time in seconds	1.08	2.00	2.93	4.48	11.97	19.94
Average transfer time with security services	3.83	7.30	13.62	25.70	74.66	125.61
Overhead % due to security services	354.60	365.00	464.80	573.60	623.70	629.90

To determine the time of file transfer and the data transfer rate when using a GPRS network, six files whose sizes ranged from 4 KB to 256 KB were transferred from cell phone to the application server. The sizes of these files were smaller than those of the same test for WLAN because the nominal transfer rate of GPRS is only 150 KB/s, transferring files over 1 MB would be slow as well as costly. The average transfer times and transfer rates with and without considering the time added by the security services is shown in Table 4.

Table 4. File transfer times and data transfer rates using GPRS

	4 KB	8 KB	32 KB	64 KB	128 KB	256 KB
Average transfer time in seconds	20.74	44.41	159.58	287.8	519.42	905.65
Average transfer time with security services	21.39	44.70	163.26	288.38	519.61	921.74
Overhead % due to security services	1.03	1.01	1.02	1.00	1.00	1.01

To determine the battery life for data encryption, the battery was fully charged and then a 1 MB text file was encrypted repeatedly until the battery ran out. At the same time the percentage of battery use was recorded periodically. It was found that an average 776 MB of data can be encrypted with a fully charged battery. The process for determining the power consumption is similar for decryption. The tests showed that energy consumption is slightly higher for decryption. It was found that on average of 749 MB of data can be decrypted with a fully charged battery.

The test of power consumption of data transfer was only carried out through a WLAN because doing this test through GPRS was impossible due to economic constraints. As in previous tests the battery was fully charged and a text file of 1 MB was sent repeatedly from the cell phone to the application server until the battery ran out. It was found that an average of 580 MB of data can be transferred with a fully charged battery.

7 Conclusions

The vertical handover platform works as a separate module that can be used by any application that accesses the Internet from a cell phone. The integration of the handover platform to different applications is feasible as demonstrated by the development and implementation of the system for monitoring vital signs. However, the handover mechanism developed is suitable mainly for applications tolerant to transmission delays.

Tests showed that the handover mechanism successfully performs network changes in the three possible scenarios. This system, as demonstrated by tests, is capable of transferring information from the cell phone to the application server on scenarios of network change, ensuring the integrity and confidentiality of the information.

References

1. How americans use their cell phones, http://www.pewinternet.org/pdfs/PIP_Cell_phone_study.pdf
2. Wi-Fi component forecast and vendor share 2005, http://www.strategyanalytics.net/default.aspx?mod=ReportAbstractViewer&a0=2480
3. Smith, R.: Wi-Fi Home Networking, pp. 16--19. McGraw-Hill, (2003)

4. Salkintzis, A.K.: Interworking between WLANs and third-generation cellular data networks. Vehicular Technology Conference, 2003, 3:1802--1806, (2003).
5. Good, R., Ventura, N.: A multilayered hybrid architecture to support vertical handover between IEEE802.11 and UMTS. IWCMC 2006, pp. 257--262, (2006)
6. Rong-Hong Jan, Wen-Yueh Chiu.: An approach for seamless handoff among mobile WLAN/GPRS integrated networks. Computer Communications, 29:32--41, (2005)
7. Jee-Young Song, Hye Jeong Lee, Sun-Ho Lee, Dong-Ho Cho.: Hybrid coupling scheme for UMTS and wireless lan interworking. International Journal of Electronics Communications, 61:329--336, (2007)
8. Rodríguez-Henríquez, F., Saquib N.A., Díaz-Pérez, A., Kaya Koc, C.: Cryptographic Algorithms on Reconfigurable Hardware, pp. 8--9.wblock Springer, (2006)
9. NIST. Announcing the ADVANCED ENCRYPTION STANDARD (AES), November 2001.
10. NIST. Announcing the SECURE HASH STANDARD, August 2002
11. Apostolis K. Salkintzis. Interworking between WLANs and third-generation cellular data networks. Vehicular Technology Conference, 2003, 3:1802{1806, 2006.
12. Stallings, William.: Cryptographic and Network Security Principles and Practices, pp. 11--14,353. Prentice Hall (2005)
13. Houda Labiod, Hossam A__, and Costantino de Santis. Wi-Fi Bluetooth ZigBee and WiMax, page 104. Springer, 2007

WEBCEL: herramienta para generar contenido web para comercio móvil

Ernesto E. Quiroz M [1], Lirio Ruiz G. [2], Isaura González R. A.
.
[1] Centro de Investigación y Desarrollo de Tecnología Digital del IPN
Av. del Parque No. 1310, Mesa de Otay Tijuana, Baja California, México CP-22510
correo-e: {eequiroz, isaura}@citedi.mx
[2] Universidad de la Sierra del Sur, Calle Guillermo Rojas Mijangos S/N, Esq. Av. Univ., Col. Ciudad Universitaria, Miahuatlán de Porfirio Díaz, Oax. 70800. México
lruiz@unsis.edu.mx
Paper received on 07/08/10, Accepted on 14/09/10.

Abstract. Las comunicaciones celulares han logrado que los teléfonos móviles tengan presencia en todos los escenarios de la actividad humana, sin distingo de edad, género, estado social, etc. Los servicios de voz, texto, correo electrónico, y últimamente el de Internet son los más demandados por los usuarios. El comercio móvil (*m*-commerce) tiene una aceptación que crece con el tiempo. En este orden de ideas, este artículo presenta un paquete de software para elaborar sitios web corporativos desplegables en terminales celulares. Los usuarios potenciales son microempresas que no tienen un especialista o departamento de informática, pero cuentan con una persona con conocimiento promedio en el manejo computadoras personales. El programa genera páginas XHTML-MP que se pueden descargar desde cualquier celular habilitado para Internet. Pequeñas compañías pueden promover sus productos y servicios a clientes que tengan un celular, descartando el requerimiento de laptops costosas o computadoras de escritorio atadas cables.

1 Introduction

La forma de efectuar compras ha sido transformada por el comercio electrónico (*e*-commerce) a través de Internet, el cual se ha convertido en un medio común para la búsqueda de un producto, evaluación de alternativas, compra y entrega del mismo. Se estima que el volumen de transacciones al final del 2010 será de 301 mil millones de dólares [1] solo en EUA. Una nueva avenida del e-commerce se abrió a partir del 2002, con la aparición de la versión 2 del Protocolo de Aplicación Inalámbrica (WAP: Wireless Access Protocol), el cual hizo posible la navegación web desde celulares de una manera práctica. Por sus características sui generis, que la diferencían del e-commerce tradicional, se le ha dado la denominación particular de comercio móvil (*m*-commerce). Coincidentemente entre 2001 y 2003 se introdujeron las primeras redes 3G que ofrecen una experiencia más rica en el manejo de multimedia en el web, derivadas de enlaces con tasas de 1.5 a 2 Mbps. La expresión comercial del Internet por celular es una veta de actividad económica que ha crecido

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 261.-271

rápidamente, de tal forma que se estima que en 2011 será un mercado de alrededor de 10 mil millones de dólares [2].

El futuro del *m*-commerce se hace patente por el número de celulares en operación actualmente, 4,700 millones al final del 2009 [3], más de la mitad de la población global (6.8 mil millones) [4]; y lo confirma la tendencia reportada por "The Mobile Internet Report [5]" la cual afirma que mas usuarios se conectarán a Internet por medio de notebooks y teléfonos inteligentes que PCs fijas dentro de 5 años.
En este contexto, este artículo describe una Herramienta de Diseño de Contenido para Publicidad Móvil (WebCel) para crear páginas web accesibles desde los micronavegadores de celulares 2.5G, 3G, 4G. El programa es intuitivo y amistoso, pensado para un usuario con conocimientos generales en el uso de una computadora. El propósito principal del paquete es poner al alcance de negocios pequeños la posibilidad de dar a conocer sus productos y servicios a clientes potenciales que poseen un celular (77 % de los habitantes en México [6]), independientemente del proveedor celular y la tecnología (CDMA Uno, GSM, UMTS/3G, LTE/ 4G).

Hay una gran cantidad de editores HTML (Dreamweaver, Bluefish, Frontpage, etc.), no obstante las opciones se reducen al requerir la capacidad de desarrollo con XHTML-MP. Además, estos editores está dirigido normalmente a desarrolladores de software, y muy pocos a usuarios no especialistas. Una de las mejores herramientas en este último género es el Mobisite Galore, el cual es un sitio web dedicado al desarrollo de aplicaciones para móviles, entre cuyas facilidades están: Modificación con asistente, acceso al código fuente, validación del código, hojas de estilo, publicación remota via FTP, y otras mas. Apuntando al nicho de micro y mini-empresas, WebCel llena el objetivo de proveer un ambiente para generar páginas web corporativas para móviles, por usuarios que no conocen el lenguaje ni las herramientas de HTML.

2 Herramienta de publicidad móvil autocontenida

Considerando el ambiente de trabajo de una microempresa, se proyectó que WebCel contuviera todo lo necesario para generar un sitio web corporativo y consultable, con los requerimientos mínimos de una computadora personal y una conexión de Internet.

Las principales características de WebCel se describen a continuación:

- Interfaz amigable al usuario basada en asistentes y plantillas de diseño.
- La información de las páginas contiene texto e imágenes (fijas y/o móviles).
- Interfaz de usuario con barra de menús contextuales y barra de herramientas.
- Área de trabajo para mostrar los avances y ediciones del diseño.
- Servicios web de Publicación, Borrado, Consulta, y Actualización.
- Repositorio de datos para que el sistema pueda guardar información del usuario.
- Documentación en línea.
- Compresor de imágenes.

- Instructivo paso a paso para convertir la computadora en un servidor (si tienen el Internet Information Server de Windows).
- Publicación local (en la misma computadora) o remota (a un servidor de hospedaje).

En el resto del artículo se presenta el desarrollo de las partes de interfaz de usuario y del generador de páginas web-móvil de WebCel.

3 Desarrollo de una herramienta de publicidad para sistemas celulares

3.1 Contenido Web para Sistemas Celulares

Las primeras versiones de acceso web a portales inalámbricos aparecieron en 1996 [7] en versiones propietarias. La conjunción de varios factores ha cambiado diametralmente la aceptación de la navegación web: (a) la evolución de la tecnología celular de GSM (Gobal System for Mobile Communications) a EDGE (Enhanced Data rates for GSM Evolution) (374 Kbps) y UMTS (Universal Mobile Telephone Systems) (2 Mbps). (b) terminales celulares con pantallas a color, mayor resolución, y procesadores más rápidos. (c) mejoramiento del protocolo WAP (versión 2.0) por la OMA (Open Mobile Alliance) [8]. (d) Establecimiento en 2008 del Grupo de Trabajo de Publicidad Móvil por la OMA, con el propósito de estandarizar procedimientos para operadores y publicadores para crear publicidad basada en el perfil y demandas del usuario.

WAP 2.0 define el uso de XHTML-Perfil Móvil (XHTML-MP) como el lenguaje de marcado para crear páginas web para equipos móviles [9, 10]. El Servicio de Validación de Marcado [11] permite la validación de código de XHTML-MP vía web.

3.2 Análisis

El objetivo de WebCel es generar un sitio web a partir de la información proporcionada por el usuario.

La información capturada se estructura en secciones. Se consideran las siguientes secciones: la identidad de la empresa, información de contacto, productos, servicios, clientes, trabajos realizados, noticias y una opción final que el usuario puede personalizar.

Por cada sección se genera una página dentro del sitio. La información de la empresa se guarda en un archivo que se puede modificar posteriormente. Una vez generado el sitio, este puede publicarse en un servidor local o remoto para que pueda ser consultado desde cualquier dispositivo móvil que cuente con un navegador web. (Fig. 1)

Los requerimientos generales de WebCel son entonces:

1. Capturar información.
2. Crear un sitio web a partir de la información.

3. Guardar Información.
4. Leer Información para su modificación.
5. Publicar el sitio Web generado.

A partir de estos requerimientos se identificaron los casos de uso (Ver Fig. 2)

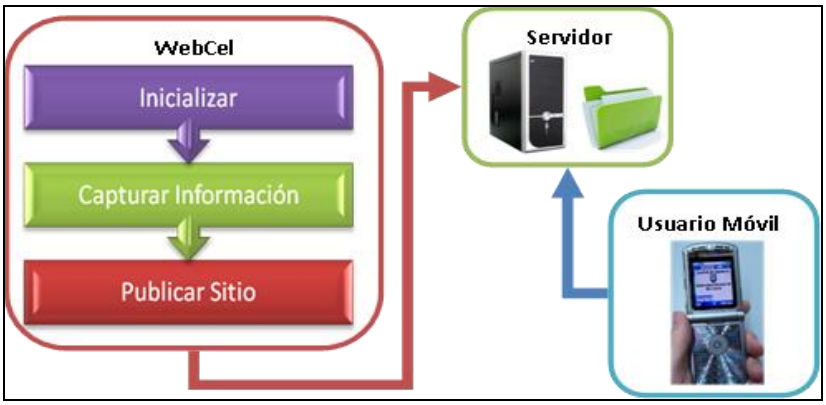


Fig. 1 Esquema general de funcionamiento.

3.3 Diseño

Una vez identificados los casos de uso y utilizando el patrón de diseño de ingeniería de software MVC (Modelo-Vista-Controlador) [14], [15] se diseñaron las clases (Fig. 3). La *Vista* representa las clases utilizadas para la interfaz de usuario, donde son capturados y presentados los datos. El *Modelo*, representa los datos de la empresa. Y el *Controlador* está compuesto por las clases que permiten guardar la información, cargarla y generar el sitio web.

Dependiendo de las secciones elegidas por el usuario en la *Vista*, el *Controlador* actualiza la información del *Modelo*.

Las clases que componen el *Modelo* son las que guardan información de la empresa se separaron en: Identidad, Descripción, Contacto, Noticia, Cliente, Producto, Trabajo y DatoUsuario, esta última es personalizable para el usuario de tal manera que puede guardar en ella otra información no prevista. Se cuenta con la clase central *Empresa* que se relaciona con las demás clases por una relación de agregación (Ver Fig. 4)

La captura de datos se realiza mediante el componente *Vista*, cada una de las clases que lo conforman es una ventana desde la que se capturan y donde se presenta la información de la empresa (Ver Fig. 5). Así por ejemplo se tiene la clase *FormProducto* que permite al usuario capturar la información del producto. Se tiene la clase *FormAsistenteInicio* que es la primera ventana del asistente. Es en ella donde el usuario selecciona las secciones que contendrá su sitio, es por eso que en el diagrama se muestra como clase central ya que según lo que el usuario elija en ella, se

crearan las demás clases. El usuario puede volver a esta ventana inicial para modificar su selección en cualquier momento.

El componente *Controlador* (Fig. 6) contiene clases que permiten realizar el proceso de guardar los datos para su posterior lectura y la generación del sitio. La clase *Asistente* define los métodos para actualizar los datos, por eso se relaciona directamente con la clase *Empresa.* Dichos datos se guardan en un objeto serializable con la ayuda de la clase *Directorio.* Se tienen también la clase *Sitio* que se *compone* de la clase *DocHtml* y esta a su vez de la clase *Nodo*. Asemejando la realidad donde un sitio web está compuesto por páginas web que no son más que documentos HTML, y los documentos están compuestos por nodos.

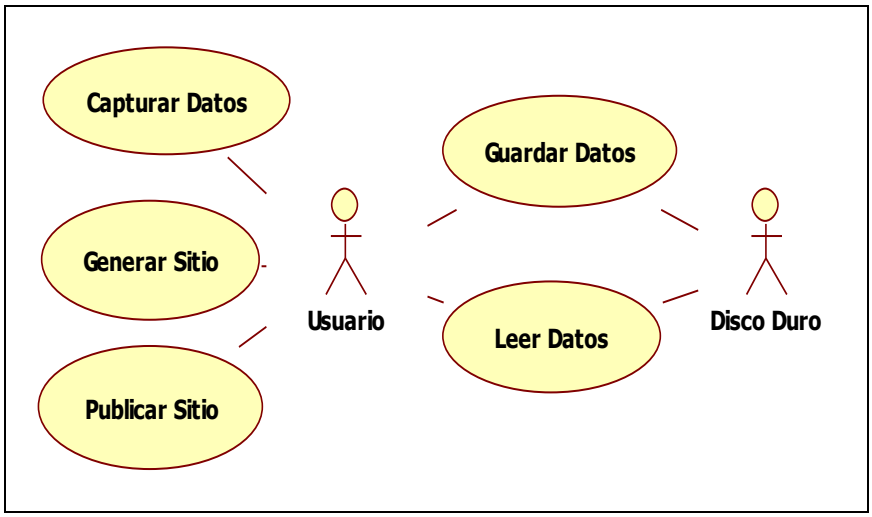


Fig. 2 Diagrama general de clases de uso

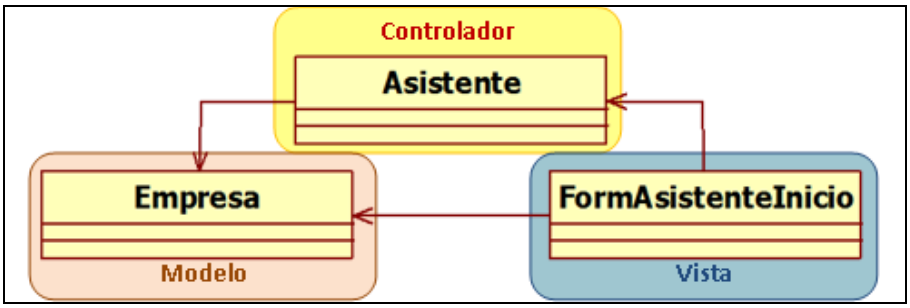


Fig. 3 Diagrama general de clases basado en MVC

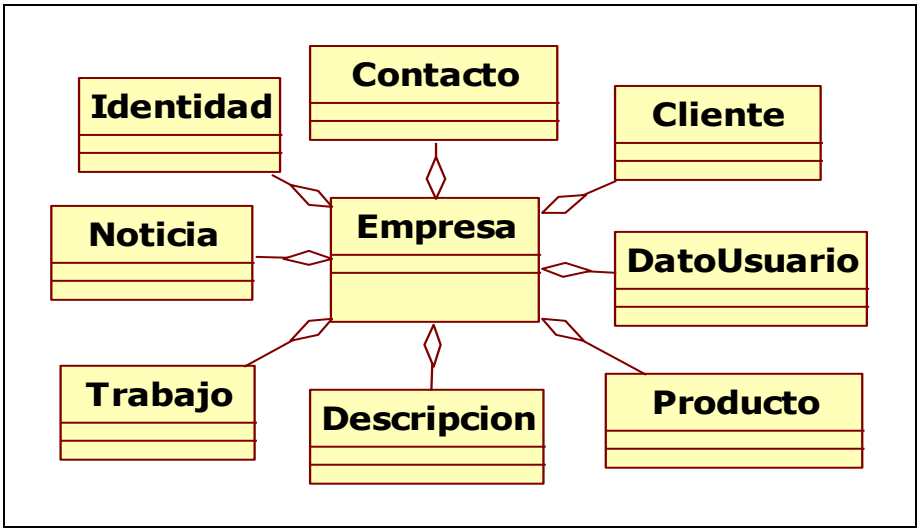


Fig. 4 Diagrama de clases del componente *Modelo*

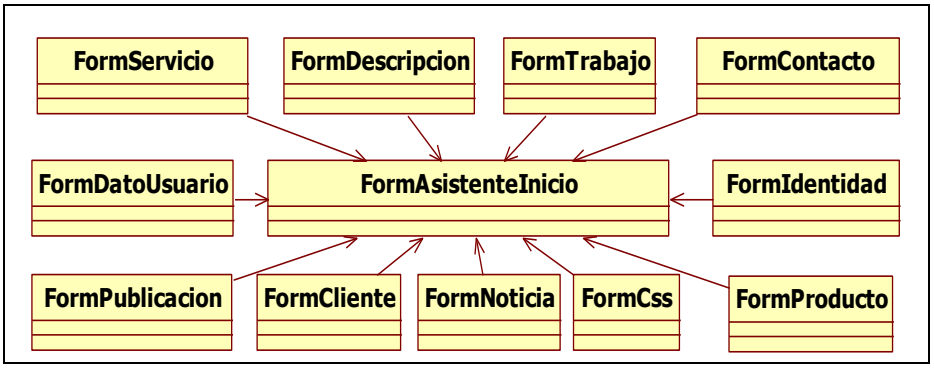


Fig. 5 Diagrama de clases del componente *Vista*

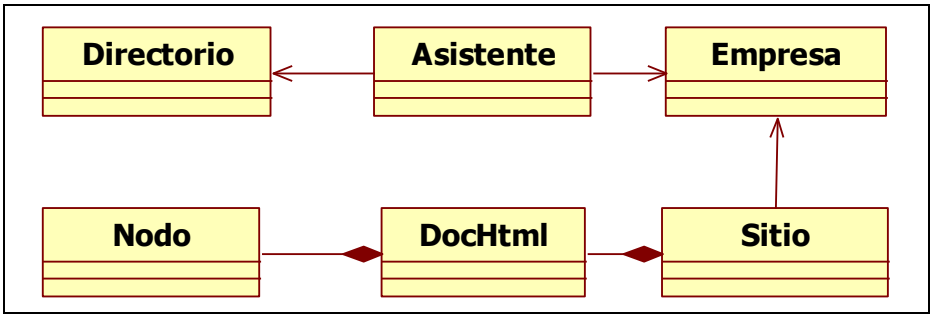


Fig. 6 Diagrama de clases del componente *Controlador*

4 Operación

La pantalla de apertura de WebCel despliega tres opciones (a) Diseñar Nuevo Sitio (b) Actualizar Sitio (b) Publicar Sitio (Fig. 7). Pulsando en el primer ícono, se abre una pantalla con nueve aspectos de la empresa, que al ser seleccionados permitirán más adelante introducir información acerca de la compañía (Fig. 8). Sólo los temas marcados abrirán plantillas en pasos subsecuentes, mientras los omitidos pueden activarse en sesiones posteriores. Esta plantilla (Fig. 8) corresponde a la clase *FormAsistenteInicio* del componente *Vista* (Fig. 5).

El botón "Siguiente" en la parte baja de la página abre una nueva pantalla con tres opciones de combinación de colores de fondo y fuentes de letras, que permite seleccionar, e inclusive estar cambiando la imagen corporativa (Fig. 9).

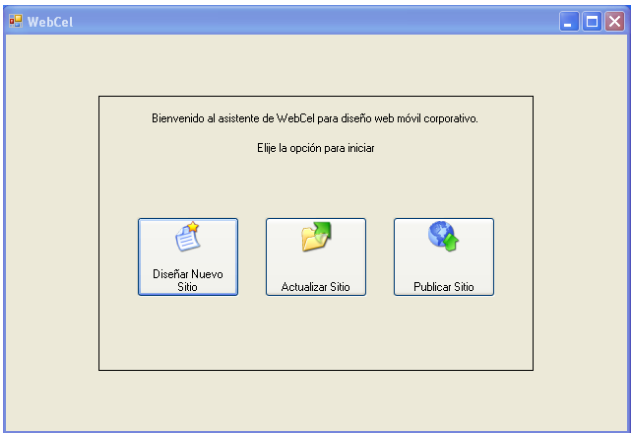


Fig. 7 Pantalla de entrada

Haciendo clic en "Siguiente" se inicia la captura de información de la empresa mediante la inserción de texto e imágenes, En cada plantilla sucesiva el usuario encontrará campos de manera similar a la de *Identidad* (Fig. 10). En cualquier etapa del proceso, el usuario puede pulsar el botón "Vista Previa" que despliega en modo de presentación final, la información introducida hasta ese momento.

Una vez que el usuario ha terminado el llenado de todas las plantillas, la información se guardará en una carpeta que contiene las carpetas secundarias de hojas de estilo (css) y las imágenes insertadas (img).

Al terminar la sesión puede publicar lo hecho con la opción “Publicar Sitio”, o reanudar la sesión posteriormente. La opción de publicar convierte la información al lenguaje XHTML-MP, y lo guarda en el servidor local. WebCel también proporciona la opción de publicar en un servidor remoto. La información completa de operación de WebCel se encuentra en [16].

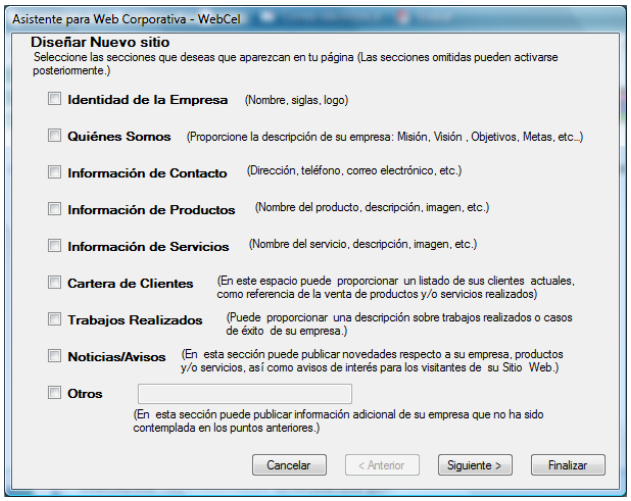


Fig. 8 Plantilla con información selecionable

Fig. 9 Plantilla con hojas de estilo

5 Sitios web desarrollados con WebCel

Se han elaborado y publicado sitios-muestra de compañías virtuales utilizando WebCel (http://webcel.citedi.mx/). Se llevaron a cabo pruebas para acceder y navegar los sitios con diversos teléfonos móviles, la tableta Nook de Barnes&Noble y dos proveedores de servicios celulares. Debido al compresor de imágenes integrado a WebCel, los contenidos son ligeros y las descargas rápidas. Cuando el tamaño del contenido es mayor que el área de la pantalla, los navegadores proveen una o dos ba-

rras deslizables (scroll bar), que permiten al usuario desplazar la página. Otros navegadores incorporan opciones para ver página completa y ajuste de tamaño (zoom) (Fig. 9). La transferencia tecnológica del programa se efectuó a un organismo que aglutina empresas de software entre otras.

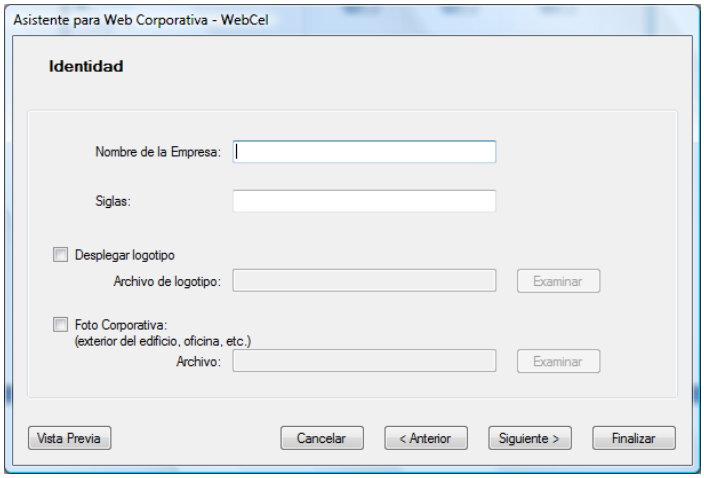


Fig. 10 Plantilla con información de identidad.

6 Conclusiones

M-commerce es una industria naciente, que ofrece un enorme espectro de posibilidades de investigación y desarrollo tecnológico, del cual WebCel es una muestra terminada, transferida y operando.

WebCel es una Herramienta de Diseño de Contenido para Publicidad Móvil, que permite generar sitios web corporativos para ambiente telefónico celular, consultables vía Internet móvil. WebCel se diseñó como una opción de bajo costo para pequeñas empresas con pocas o mínimas posibilidades de sufragar los honorarios de un especialista en elaborar publicidad web, por lo cual es de uso amigable para un usuario promedio de computadora. Una aplicación alterna del paquete es la explotación por compañías de publicidad en línea que ofrecen diseño y hospedaje de páginas web.

Webcel se basa en plantillas y ayudas en línea (wizards) que guían al usuario, requiriendo información en forma de texto e imágenes. Proporciona documentos de ayuda, compresor de la imagen, y visualización previa de páginas a cada paso del proceso. Al finalizar genera las páginas web en lenguaje XHML-MP, valida el contenido en línea, y lo publica en el servidor designado.

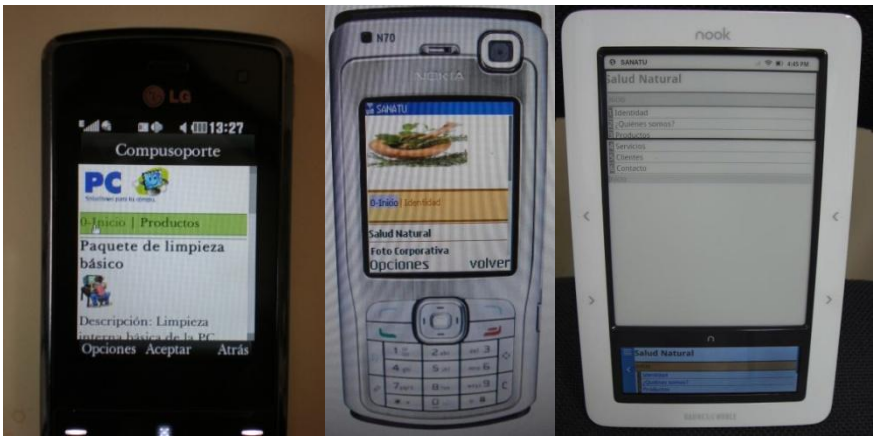


Fig. 11 Fotos de sitios web-móvil creadas con WebCel.

Reconocimientos.

Este trabajo fue financiado por el Gobierno del Estado de Baja California y el CONACYT mediante proyecto Fomix 2005-2/22634, con apoyos complementarios del IPN (SIP 2009-0995) y beca de COFAA Exclusividad.

References

1. Sucharita Mulpuru, et al, Forrester Research Inc., "U.S. E-Commerce Forecast: 2008 to 2012, January 18, 2008.
2. Open Mobile Alliance, Quarterly 2010, Volume 1.
3. The Top 25 LTE Operator Commitments: Deployment Scenarios and Growth Opportunities, Maravedis Inc. 2010.
4. US Census Bureau http://www.census.gov/, Last query: August 1, 2010.
5. The Mobile Internet Report, Morgan Stanley Research, Dec., 15, 2009.
6. Dirección de Información Estadística de Mercados COFETEL, http://www.cofetel.gob.mx/wb/ Cofetel_2008/Cofe_telefonia_movil.
7. NetHopper 2.0 First true web browser for Newton, PenComputing Magazine. Issue No. 11, 2006.
8. Open Mobile Alliance, 2008 Annual Report.
9. Developer's Home. XHTML Mobile Profile / XHTML MP. Disponible en: Tutorial http://www.developershome.com/wap/xhtmlmp/ Ultima Consulta: Marzo 2008.
10. Open Wave Developer Network. XHTML-MP Style Guide Disponible en: http://developer.openwave.com/dvl/support/documentation/guides_and_references/xhtml-mp_style_guide/index.htm Ultima Consulta: Marzo 2008.
11. Markup Validation Service. Disponible en: http://validator.w3.org/ Ultima Consulta: Marzo 2008.
12. Micah Dubinko, Leigh L. Klotz, Jr., Roland Merrick, T. V. Raman, XForms 1.0, World Wide Web Consortium, Document TR/2002/CR-xforms-20021112, November 2002.
13. Model-View-Controller Available at: http://java.sun.com/blueprints/patterns/MVC-detailed.html Last query: March 2009

14. Stephener R. Schach.: Análisis y diseño orientado a objetos con uml y el proceso unificado. McGraw Hill. México(2005)
15. Ian Sommerville.: Ingeniería de Software. Adison Wesley. Madrid(2005)
16. Ernesto E. Quiroz M. y Lirio Ruiz G., Reporte Técnico del proyecto Fomix 2005-2/22634, Centro de Desarrollo de Tecnología Digital del Instituto Politécnico Nacional, México, Abril 2010.

Implementación de un metódo híbrido usando secuencias de ADN, para la obtención de probabilidades de enfemedades (diabetes tipo2, hipertensión e insuficiencia renal)

Vianney Morales Zamora, José Crispín Hernández Hernández, José Federico Ramírez Cruz

Avenida Instituto Tecnológico S/N C.P. 90300,
Apizaco, Tlaxcala, México
vimoza@hotmail.com,josechh@itapizaco.edu.mx,
Federico_ramirez@itapizaco.edu.mx

Abstrac. En este articulo se presenta un algoritmo hibrido que permite mediante la alineación de secuencias de ADN (algoritmo genético y programación dinámica), obtener una probabilidad (cadenas de Markov) de tener cierta enfermedad, como lo es diabetes tipo2, Hipertensión e insuficiencia renal. Debido a que en los últimos 3 años el número de personas con estas enfermedades han aumentado de manera considerable hasta en un 100% en nuestro país, es importante saber que probabilidad existe de tener las enfermedades anteriores, donde se utilizan secuencias de ADN, que debido a la gran cantidad de información de estas secuencias es necesario utilizar métodos híbridos que nos permitan mejorar la eficiencia de los resultados.

K**eywords:** Alineación de secuencias, algoritmo genético, programación dinámica, cadenas de Markov.

Introducción

En los últimos diez años las computadoras han tenido un papel muy importante en las áreas de la biología, y la medicina. La computación se ha enfocado al análisis de secuencias biológicas, cinco años atrás hubo muchos logros en la identificación de secuencias de genomas, como el de la mosca, y los primeros intentos en el proyecto de identificación de la secuencia del genoma humano [4][8]. Estas nuevas tecnologías y otras de alto desempeño como los arreglos de ADN (Ácido Desoxirribonucleico) y espectrografía de masas han tenido un progreso considerable [5]; y debido a que la información es demasiado grande, es necesario el uso de métodos híbridos que procesen de una manera efectiva y eficiente dicha información, como es el caso de los algoritmos genéticos y la programación dinámica para buscar la mejor alineación local óptima. Cuando se visualiza el ADN, ARN o las secuencias de proteínas como una cadena o lenguaje formal sobre alfabetos de cuatro nucleóti-

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, pp. 275-288

dos o 20 amino ácidos, o como una representación gramatical y métodos de inferencia gramatical que pueden ser aplicados a varios problemas para análisis de secuencias biológicas[2].

Para obtener las probabilidades de enfermedades, primero se alinean las secuencias de ADN usando un algoritmo genético como se muestra en sección 2.1 y después el resultado será la entrada a la programación dinámica presentado en la sección 2.2, y teniendo las secuencias alineadas por pares, a continuación se obtienen las probabilidades de las enfermedades utilizando cadenas de Markov como se expresa en la sección 2.3.

Método propuesto

Algoritmo Genético para la alineación de secuencias por pares

La idea de este método es intentar generar alineaciones mediante reacomodaciones que simulan la inserción de gaps (huecos) y acciones de recombinación durante la replicación para obtener puntajes más altos para el alineamiento [10][15][18][21].

Se tomaron como entradas al algoritmo genético pares de secuencias, para los cual solo analizaremos los primeros 5 pares de secuencias. Cada par se ingreso al algoritmo genético, obteniendo como salida la mejor alineación.

1. Se toma el primer par de secuencias, y se pide el número de individuos (NI) a generar, y el número de generaciones a realizar (NG).
2. Se busca la cadena mas larga (x) del par de secuencias, se calcula la longitud de esta (L=lenght (x)), y se obtiene un factor (F=Lx0.25). Este factor da los espacios (gaps) extras que se van a usar para mover las cadenas, explícitamente se genera un número aleatorio entre 0 y F que será e l numero de gaps a insertar al inicio, y posteriormente se insertan gaps al final para hacer que las cadenas tengan la misma longitud. Así se obtienen los nuevos posibles alineamientos. Esto se realiza (Ni) veces de acuerdo con el número de individuos.
3. A continuación se saca el puntaje de cada alineamiento, teniendo en cuenta las calificaciones establecidas de acuerdo en la matriz de blosum62 que se muestra en la figura 1, cuya ecuación es la siguiente:

$$a_{ij} = \left(\frac{1}{\lambda}\right) \log\left(\frac{p_{ij}}{q_i * q_j}\right) \qquad (1)$$

Donde p_{ij} es la probabilidad de que dos aminoácidos *i* y *j* reemplacen uno al otro en una secuencia homóloga, mientras que q_i y q_j son las probabilidades últimas de encontrar los aminoácidos *i* y *j* en cualquier secuencia de proteína de forma aleatoria. El factor λ es un factor de escala para asegurar que, tras su aplicación y la de un

necesario redondeo al entero más cercano, la matriz contenga valores enteros dispersos y fácilmente tratables.

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Fig. 1**.** Matriz Blosum62

Utilizando este puntaje se selecciona la mitad de alineamientos para ser utilizados posteriormente en el proceso de selección y reproducción.

4. A continuación la otra mitad de los alineamientos pasan por un proceso de mutación, se obtiene el tamaño total de las secuencias(TT) tomando en cuenta los espacios anteriormente insertados, al igual se obtiene la diferencia del número de variables máximo (Nmax) y el menor numero de variables (Nmin) del par de alineamientos, y se generan números aleatorios entre 0 y ese rango (A=random(Nmax-Nmin)) que indican el numero de gaps a insertar, en seguida de acuerdo al tamaño total de las secuencias, se genera por cada gap a insertar un numero aleatorio entre 1 y (TT) que indican la posición donde será insertado cada espacios (A) realizándose este proceso con los pares de secuencias.

Y finalmente obtenemos la puntuación de las nuevas alineaciones conforme a la matriz de blosum62 expresada en la figura 1 y se compara el puntaje obtenido con los alineamientos anteriores; y si el puntaje del nuevo alineamiento es mayor, se seleccionan las nuevas alineaciones, de lo contrario se repite el proceso hasta que exista alguna mejora.

Con el proceso anterior se busca darle una mayor probabilidad a los alineamientos “fuertes” de ser elegidos.

5. Después tanto la primera mitad de los mejores alineamientos como la segunda mitad de los alineamientos mutados se juntan, y luego son calificados y reordenados. Que serán utilizados en el proceso de reproducción.
6. Ahora se realiza el proceso de reproducción basado en el método de torneo en el cual se selecciona aleatoriamente parejas de alineamientos; aquí el que tenga un mejor puntaje va a ser elegido como padre, de cada par de secuencias. Obteniendo los padres para el proceso de reproducción.
7. A partir de estos alineamientos se genera un número aleatorio dentro del rango de uno hasta la longitud de la cadena más corta (Nmin), ese será el punto de corte para cada par de secuencias. Este numero aleatorio (AC), contara solo cada carácter sin contar los espacios.
8. Después se intercambia la primera parte de la matriz superior con la segunda de la inferior y viceversa, y se obtiene la puntuación de cada hijo.
9. **S**e elije el hijo que tengan la mejor calificación y es agregado a la nueva generación.

Y de acuerdo al número de la población (NI), se realiza el proceso 7, 8 y 9.

Se hace mención que en el proceso de reproducción se eliminan las columnas llenas de gaps, para evitar que la secuencia crezca exponencialmente.

Una vez que se ha generado el número de hijos suficientes para cumplir con el total de la población (NI), se retorna al paso tres para seguir con el procedimiento hasta completar el número solicitado de generaciones (NG).

A continuación se muestra un ejemplo resultado de alinear con el algoritmo genético, la figura 2 muestra las alineaciones de entrada y la figura 3 muestra la salida obtenida con el algoritmo genético:

```
TCACCTGGGTGTGTGGGTGCCGTTCCAGGCTGTCAGAGCTCGCGTGGGGGTGTGGGTGCTGCTCCAGGCT
TTCGGAGCTCACCTGGGGGTGCAGGGTGCTGTTCCAGGCTGTCAGATGCTCACCTGGGGGTGTGGTTGCT

GCTCCAGGCTGTCAGATGCTCACTTGGGGGTGCAGCGTGCTGTTCCAGGCTGTCAGATGCTAACCTGGGG
TTGTGAGAGCTGTTGCAGGCTGTCAGATGCTCACTGGGGGTGTGCGTGCTGCTCCAGCCTGTCAGATGCT

CACCTGGGGGTGTGGGTGCTGTTCCAGGATATCAGATGCTCACCTGGGGGTGTGGGTGCTGCTCCAGGCT
GTCGGATGCTCACCTGGGGGTGTCGGTGCTGTTCCAGGCTGTTGGATGCTCACCTGGGGGTGTGGTTGCT

GCTCCAGGCTGTCAGGTGCTCACTTGGGGGTGCAGCGTGCTGTTCCGGGCTGTCAGATGCTCACCTGGGG
TTGTGAGAGCTGTTCCAGGCTGTCAGATGCTCACCTGGGGGTGTGGGTGCTGTTCCAGGCTGTCAGATGC

TCACCTGTTGGTGTGGGTGCTGCTTCAGGCTGTCAGATGCTCACCTGGAGTGTGGGTGCTGCTCCAGGCT
GTCTGATCCTCACTCGCGGGTGCAGGGTTCTGTTCCAGGCTGTCAGATGCTTGCCTGGGGGTGTGGTTGC
```

Fig. 2. Ejemplo de la secuencia de entrada

Solución obtenida:

```
puntuacion1 =  91.88
Alineacion1 =
---------------TCACCTGGGTGTG-TGGGTGCCGTTCCAGGCTGTCAGA-GCTCGCGTGGGGGTGTGGGTGCTGCTCCAGGCT
           |||||||||||:||||||||||||||||||||||||||:|||||||||| ||| |  |||
TTCGGAGCTCACCTGGGGGTGCAGGGTGCTGTTCCAGGCTGTCAGATGCTCACCTGGGGGTGT--G-G-T--T----GCT

puntuacion2 =  87
Alineacion2 =
---------------GC---T-C-CAGGCTGTCAGATGCTCACTTGGGGGTGCAGCGTGCTGTTCCAGGCTGTCAGATGCTAACCTGGGG
             ||  |   |||||||||||||||||||||||||||||||||   ||| ||||||||||||||| ||
TTGTGAGAGCTGTTGCAGGCTGTCAGATGCTCAC-TGGGGGTG-TGCGTGCTGCTCCAGCCTGTCAGATG-----CT----

puntuacion3 =  88.777
Alineacion3 =
-----------------CACCTGGGGGTGTGGGTGCTGTTCCAGGATATCAGATGCTCACCTGGGGGTGTGGGTGCTGCTCCAGGCT
              |||||||||||||||||||||||||||:| :|||||||||||||||||||| ||| |  |||
GTCGGATGCTCACCTGGGGGTGTCGGTGCTGTTCCAGGCTGTTGGATGCTCACCTGGGGGTGT--G-G-T--T----GCT

puntuacion4 =  88.787
Alineacion4 =
---------------GC-----TCCAGGCTGTCAGGTGCTCACTTGGGGGTGCAGCGTGCTGTTCCGGGCTGTCAGATGCTCACCTGGGG
             ||     |||||||||||||:|||||||||||:| |||||||||||:|||||||||||  ||||||||     |
TTGTGAGAGCTGTTCCAGGCTGTCAGATGCTCACCTGGGGGTGTGG-GTGCTGTTCCAGGCTGTCAGATG-----C-----

puntuacion5 =  77.99
Alineacion5 =
--TC---A----CCT-GTTGGTGTGGGTGCTGCTTCAGGCTGTCAGATGCTCACCTGGAGTGTGGGTGCTGCTCCAGGCT
  ||   |  |||  | ||||              ||| ||||||||||||||||||:|||||||||||| || : ||
GTCTGATCCTCACTCGCGGGTGCAGGGTTCTGTTCCAGGCTGTCAGATGCTTGCCTGG-G-G-GTGTG--G-T--T-GC-
```

Fig .3. Secuencia alineada y sus puntajes de alineación, obtenidas con el algoritmo genético. La alineación obtenida se introduce como entrada en la programación dinámica.

Programación dinámica para la alineación de secuencias

La Programación Dinámica (PD) es una alternativa de descomposición en que se resuelven subproblemas más pequeños y luego se juntan asumiendo que en cada etapa futura se tomaran decisiones correctas [13].

Los pasos del diagrama se definen a continuación:

1.-Se introducen en pares las secuencias para nuevamente alinearlas.
2.- Se elabora su matriz de encajes, como se muestra en la figura 4 y si existe coincidencia de letras se asigna a la coincidencia un 1 y si no hay se coloca un cero.

		1	2	3	4	5	6	7	8	9	10	
		A	T	C	C	G	C	T	T	A	C	A
1	C	0	0	1	1	0	1	0	0	0	1	
2	T	0	1	0	0	0	0	1	1	0	0	
3	C	0	0	1	1	0	1	0	0	0	1	
4	C	0	0	1	1	0	1	0	0	0	1	
5	T	0	1	0	0	0	0	1	1	0	0	
6	A	1	0	0	0	0	0	0	0	1	0	
B	7	G	0	0	0	0	1	0	0	0	0	0

Fig. 4. Matriz de encajes

El camino que se eligió para este problema fue el diagonal como se muestra la figura 5, en el que se puede ir seleccionando un paso a la derecha y así mismo

hacia abajo, formando un camino diagonal a partir de la parte superior izquierda, como se muestra a continuación.

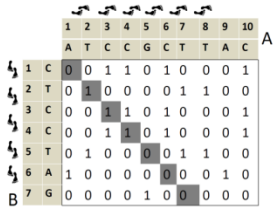


Fig. 5. Movimiento en diagonal

Buscar el mejor alineamiento entre 2 secuencias es buscar el camino que obtenga la mejor puntuación. La estrategia de la programación dinámica es dividir el problema global en subproblemas. Así se busca primero el mejor camino que empieza en cada una de las casillas. Empezamos por la última fila. En esta fila la puntuación del mejor camino para cada casilla coincide con la puntuación de encaje. En las casillas de la penúltima fila el mejor camino es de un sólo paso. Para computar las casillas del resto de las filas también basta con dar un sólo paso porque cada casilla a la que nos podemos mover lleva ya acumulada la mejor puntuación que se puede obtener pasando por ella. Así el problema de elegir el mejor camino que empieza en una casilla se reduce a elegir la siguiente casilla con mejor puntuación.

Se elije la máxima puntuación de las casillas marcadas en gris y se le suma a la puntuación de encaje tal como se muestra en la figura 6, y el resultado de este proceso se muestra en la figura 7.

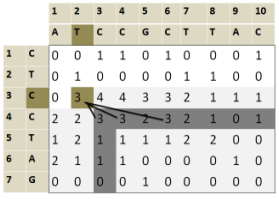


Fig.6. Obtención de puntajes máximos

Así se van a ir generando los caminos de alineación haciendo un recorrido en diagonal, y una vez obtenidos las posibles soluciones, se elije la mayor puntuación como se presenta en la figura 8, y esto se realiza con cada par de secuencias. Aplicando el ejemplo anterior de secuencias de ADN, las siguientes alineaciones obtenidas se ilustran en la figura 9.

		1	2	3	4	5	6	7	8	9	10
		A	T	C	C	G	C	T	T	A	C
1	C	5	4	5	4	3	4	2	1	1	1
2	T	4	5	4	3	3	2	3	2	1	0
3	C	3	3	4	4	3	3	2	1	1	1
4	C	2	2	3	3	2	3	2	1	0	1
5	T	1	2	1	1	1	2	2	0	0	
6	A	2	1	1	1	0	0	0	0	1	0
7	G	0	0	0	0	1	0	0	0	0	0

Fig. 7. Resultados de puntajes máximos

		1	2	3	4	5	6	7	8	9	10
		A	T	C	C	G	C	T	T	A	C
1	C	5	4	5	4	3	4	2	1	1	1
2	T	4	5	4	3	3	2	3	2	1	0
3	C	3	3	4	4	3	3	2	1	1	1
4	C	2	2	3	3	2	3	2	1	0	1
5	T	1	2	1	1	1	2	2	0	0	
6	A	2	1	1	1	0	0	0	0	1	0
7	G	0	0	0	0	1	0	0	0	0	0

Fig. 8. Alineamientos que alcanzan el máximo puntaje

En la figura 10 se presentan los resultados obtenidos de introducir las secuencias en un algoritmo genético y después en la programación dinámica.

Con lo anterior podemos comparar nuestros resultados, resaltando la mejora en la alineación usando un algoritmo genético y programación dinámica, como se muestra en la Tabla 1

Por lo tanto el método propuesto de usar un algoritmo genético y después programación dinámica, nos permite tener una mejor alineación de secuencias, como se mostro en la tabla anterior. Y ya teniendo las secuencias alineadas, ahora el siguiente paso es buscar patrones que nos permitan identificar enfermedades y poder tener una probabilidad de contraer alguna enfermedad y así poder prevenir problemas futuros. Para la solución de dicho problema se hace uso de los modelos de Markov.

Obtención de probabilidades utilizando cadenas de Markov

Uno de los aspectos fundamentales en la Bioinformática, es el diseño de modelos matemáticos que interpreten los sesgos de las secuencias biológicas e identifiquen patrones de las mismas. Uno de los modelos más utilizados para este propósito, lo son las cadenas de Markov.

Una cadena de Markov es una secuencia de variables aleatorias X (p) donde la distribución de probabilidad de cada X (i) depende de una cantidad de k valores precedentes X (i-1), X (i-2),.., X(i-k) .

Cada uno de los nucleótidos de una secuencia puede diferenciarse por su base correspondiente; por tal motivo es usual hablar indistintamente de nucleótido o de base.

```
Alineacion1 =

--------TCACCTGGGTGTG-TGGGTGCCGTTCCAGGCTGTCAGA-GCTCGCGTGGGGGTGTGGGTGCTGCTCCAGGCT
        ||||||||| ||| :|||||| |||||||||||||||| ||||:| |||||||||  | | |  |    |||
TTCGGAGCTCACCTGGGGGTGCAGGGTGCTGTTCCAGGCTGTCAGATGCTCACCTGGGGGTGT--G-G-T--T----GCT


puntuacion1 =

   92.3333


Alineacion2 =

--------GC--TC-CAGGCTGTCAGATGCTCACTTGGGGGTGCAGCGTGCTGTTCCAGGCTGTCAGATGCTAACCTGGGG
        ||  |  |||||||||||||||||||| |||||||| :|||||||| ||||| ||||||||||     ||
TTGTGAGAGCTGTTGCAGGCTGTCAGATGCTCAC-TGGGGGTG-TGCGTGCTGCTCCAGCCTGTCAGATG-----CT----


puntuacion2 =

   88.6667

Alineacion3 =

----------CACCTGGGGGTGTGGGTGCTGTTCCAGGATATCAGATGCTCACCTGGGGGTGTGGGTGCTGCTCCAGGCT
          |||||||||||||| |||||||||||||| |:| :|||||||||||||||||||||  | | |  |    |||
GTCGGATGCTCACCTGGGGGTGTCGGTGCTGTTCCAGGCTGTTGGATGCTCACCTGGGGGTGT--G-G-T--T----GCT


puntuacion3 =

    89


Alineacion4 =

--------GC---TCCAGGCTGTCAGGTGCTCACTTGGGGGTGCAGCGTGCTGTTCCGGGCTGTCAGATGCTCACCTGGGG
        ||   |||||||||||||:|||||||| |||||||| :| ||||||||||:||||||||||||     |
TTGTGAGAGCTGTTCCAGGCTGTCAGATGCTCACCTGGGGGTGTGG-GTGCTGTTCCAGGCTGTCAGATG-----C-----


puntuacion4 =

    89
Alineacion5 =

-TC--A----C-CT-GTTGGTG-TGGGTGCTGCTTCAGGCTGTCAGATGCTCACCTGGAGTGTGGGTGCTGCTCCAGGCT
 ||  |    | || |  |||| :|||| ||| | |||||||||||||||| :||||| | | | |||  | |  : ||
GTCTGATCCTCACTCGCGGGTGCAGGGTTCTGTTCCAGGCTGTCAGATGCTTGCCTGG-G-G-GTGTG--G-T--T-GC-


puntuacion5 =

    78
```

Fig.9. Resultados obtenidos únicamente con la PD

Fig. 10. Resultados obtenidos con el algoritmo genético (AG) y programación dinámica (PD).

Tabla1. Resultados en la alineación

ALGORITMO GENÉTICO	PROGRAMACIÓN DINÁMICA	AG-PD
91.88 %	92.3333 %	95 %
87 %	88.6667 %	89.888 %
88.777 %	89 %	90 %
88.787 %	89 %	89.888 %
77.99 %	78 %	78.777 %

Los modelos de cadenas de Markov aplicados a secuencias de DNA permiten calcular la probabilidad de ocurrencia de un nucleótido (es decir, de la base correspondiente a un nucleótido dado) b en la secuencia dependiendo de los k nucleótidos inmediatamente anteriores a b en la secuencia.

Los modelos de Markov se utilizan bastante en el análisis de secuencias de datos biológicos. Los más utilizados son las cadenas de Markov de orden fijo, en las cuales el contexto (es decir, la secuencia de k nucleótidos precedentes) se utiliza en cada posición donde se desea hallar la probabilidad de ocurrencia de un nucleótido dado.

Cualquier modelo de cadenas de Markov de orden fijo predice un nucleótido de la secuencia de ADN usando los nucleótidos anteriores de la secuencia; el modelo de mayor uso es el de cadenas de Markov de quinto orden.

A diferencia de un modelo simple como lo es el IDD (Independent Identically Dsitributed) donde la ocurrencia de cada nucleótido (A, T, G, C) seria independiente de los demás, en las cadenas de Markov el valor tomado de una variable es dependiente del valor tomado en el estado anterior.

En el modelo de Markov, la probabilidad de encontrar un patrón está dada por la probabilidad (P_1) del modelo de distribución de α en la primera posición y la probabilidad condicional de β en la posición i>1 (P_2), es decir,

$$\mathbf{P(x \mid M) = P_1 (X_1)} \quad \prod_{i=2.....n(x)} \quad \mathbf{P_2 (X_1|X_{1-1})} \qquad \mathbf{(2)}$$

P_2 se calcula con la siguiente fórmula:

$$P_2 (\beta|\alpha) = P (\alpha\beta)/ P (\alpha) \qquad \mathbf{(3)}$$

Para esto es necesario conocer las frecuencias de cada uno de los nucleótidos y cada uno de los pares en la secuencia dada.

Por ejemplo: Dada la siguiente secuencia de 25 nucleótidos,

AACGTCTCTATCATGCCAGGATCTG

¿Cuál sería la probabilidad de encontrar el patrón **CAAT**?

Se obtienen las frecuencias de cada uno de los nucleótidos y cada uno de los pares en la secuencia dada.

A = 6/25, C = 7/25, T = 7/25, G = 5/25

Donde el numerador en la fracción es el número de veces que existe la letra en la cadena de secuencias, y el denominador es el tamaño de la cadena, y en el caso de los pares será el tamaño total de cadena menos 1.

(AA) = 1/24 (CA) = 2/24 (TA) = 1/24
(GA) = 1/24 (CG) = 1/24 (GT) = 1/24

(TC) = 4/24 (GG) = 1/24 (AC) = 1/24
(CT) = 3/24 (AT) = 3/24 (TG) = 2/24
(GC) = 1/24 (CC) = 1/24 (AG) = 1/24

Con lo cual podemos calcular P_2 (β|α), por ejemplo del di-nucleótido AC es:

P_2 (C|A) = $P_{AC/}$ P_A = (1/24)/ (6/25) = 25/144

Por lo tanto la tanto la probabilidad de encontrar CAAT usando cadenas de Markov de primer orden es*:*

P(C) P (A|C) P (A|A) P (T|A)
P(C) P (A|C) P (A|A) P (T|A)
(7/25) (2/23)/ (7/25) (1/24)/ (6/25) (3/22)/ (6/25) =
(7/25) x (50/161) x (25/144) x (75/132) =0.00857
=0.0086

Así la probabilidad de encontrar el patrón **CAAT,** en la secuencia AACGTCTCTATCATGCCAGGATCTG, es de 0.0086.

En este caso se busca la probabilidad de encontrar patrones de enfermedades en las secuencias. Las enfermedades a encontrar serán: Diabetes, Hipertensión e Insuficiencia Renal. Cuyos patrones, serán evaluados en las secuencias para obtener una probabilidad de tener dicha enfermedad, y debido a que las secuencias ya se encuentran alineadas, la identificación de patrones será de una manera más rápida.

Para el siguiente ejemplo se tomo aleatoriamente de las bases de datos de las enfermedades 10 personas. Cuyas entradas son las siguientes:

Tabla 2. Entradas de personas con enfermedad

No. De persona	Enfermedad
1	Diabetes tipo 2
2	Diabetes tipo 2
3	Diabetes tipo 2
4	Insuficiencia renal
5	Insuficiencia renal
6	Hipertensión
7	Hipertensión
8	Hipertensión
9	Diabetes tipo 2
10	Insuficiencia renal

Y los resultados de las probabilidades obtenidas con las cadenas de Markov son los siguientes:

Tabla3. Probabilidades obtenidas

No de persona	Diabetes tipo2	Insuficiencia renal	Hipertension
1	**0.87**	0,20	0,15
2	**0.84**	0,02	0,01
3	**0.80**	0,15	0,10
4	0,19	**0.80**	0,17
5	0,13	**0.85**	0,03
6	0,17	0,01	**0.85**
7	0,15	0,10	**0.90**
8	0,11	0,01	**0.89**
9	**0.91**	0,03	0,11
10	0,02	**0.90**	0,17

Por lo que obtenemos:

Tabla 4. Resultados de las probabilidades respecto a su enfermedad

No de persona	Enfermedad	Probabilidad obtenida
1	*Diabetestipo2*	0,87
2	*Diabetestipo2*	0,84
3	*Diabetestipo2*	0,80
4	*Insuficienciarenal*	0,80
5	*Insuficienciarenal*	0,85
6	*Hipertensin*	0,85
7	*Hipertensin*	0,90
8	*Hipertensin*	0,89
9	*Diabetestipo2*	0,91
10	*Insuficienciarenal*	0,90

Conclusiones

Debido a la gran cantidad de información que se procesa al manejar cadenas de ADN, se recurren a diversos métodos y técnicas computacionales o bien sistemas híbridos que nos permiten la manipulación de dicha información para analizarla y así poder obtener mayor información de las secuencias que forman el ADN del ser humano y poder desarrollar sistemas que permitan facilitar el manejo y acceso de información a especialistas en el área de la biología molecular. Considerando estos sistemas computacionales como una herramienta esencial de ayuda.

Tras haber realizado algunas pruebas con diferentes grupos de secuencias, se evidenciaron varios aspectos que hay que tomar en cuenta para obtener buenos resultados, con este algoritmo. Dado que uno de los problemas de los algoritmos genéticos en general es la existencia de los máximos locales, en algunos casos el algoritmo se ejecuta varias veces sobre el mismo conjunto de secuencias y escoger entre las diferentes corridas el mejor resultado. Por otro lado, es necesario jugar un poco con los parámetros del algoritmo tales como el número y longitud de gaps (huecos), que se pueden insertar en las secuencias mutadas, y el tamaño de la población. Esta variación en los parámetros es muy importante porque no es lo mismo trabajar con

secuencias que son muy similares entre si, a trabajar con secuencias altamente disparejas.

Con las pruebas también se pudo detector que cuando se está realizando la implementación del algoritmo, es de vital importancia evitar que las secuencias aumenten de tamaño de manera exagerada porque si no se toman tales precauciones, puede suceder que todos los alineamientos estén llenos de secuencias que son en un alto porcentaje gaps; y si este error se propaga el resultado del algoritmo puede ser muy pobre. Por lo que la propuesta es aplicarle otro filtro a esas secuencias alineadas obtenidas del algoritmo genético, pasando por la programación dinámica para volver a alinearlas, y así poder optimizar su alineación.

Finalmente al tener una buena alineación de secuencias, se prosigue a identificar patrones de enfermedades en las secuencias mediante el uso de otro modelo estocástico como lo son la cadena de Markov, que nos permitirán obtener una probabilidad de tener una enfermedad y poder prever problemas graves y se puedan atender.

El inconveniente que presentan los modelos de cadenas de Markov es que el aprendizaje de algunos modelos puede producir dificultad cuando no se tienen suficientes datos de un contexto determinado. Sin embargo en este caso si contamos con los patrones a trabajar y nuestra salida requerida es una probabilidad, cuyo problema es fácil resolver con las cadenas de Markov.

Lo que se tiene es un sistema que nos puede mostrar diferentes soluciones a un problema y es ahí donde la capacidad de interpretar estas propuestas puede hacer la diferencia.

Bibliografía

1. D.J. Attwood, T.K y Parry-Smith. Introducción a la Bioinformática. Madrid, 2002.
2. Pierre Balde and Soren Brunak. Bioinformatics: The Machine Learning approach. Cataloging in publication data, 2001.
3. G. Baldi, P. y Pollastri. A machine learning strategy for protein analysis, volume 17. 2002.
4. U Bandyopadhyay, S y Maulik. Gene identification: classical and computational intelligence approaches. Man and cybernetics, Part C: Applications and Reviews, IEEE, 3:40_56, 2005.

5 M. Bertone, P. y Gerstein. Integrative data mining: new direction in bioinformatics engineering in medicine and biology. IEEE Magazine, 20:33_40, 2001.

6. Bethesda. Los riñones y la insuficiencia renal. National Kidney and Urologic Diseases Information Clearinghouse, Agosto 2009.
7. MD Bethesda. Diabetes. National Institute Health, December 2009.
8. Liaison Branch. National human genome research institute. Communications and Public Liasion Branch, December 2009.
9. T.Koike R. Lopez T.J. Gibson D.G. Higgins J.D. Thompson Chenna, H.Sugawara. Multiple sequence alignment with the clustal series of programs. IEEE, 3:37_45, 2003.
10. M; Merler S; Jurman G Furlanello, C; Serafini. Semi supervised learning for molecular profiling. IEEE Computational Biology and Bioinformatics, 2:110_118, 2005.
11. D.B; Knowles J. Handl, J; Kell. Multiobjetive optimization in bioinformatics and computational biology. IEEE ACM Transactions on Computational Biology and Bioinformatics, 4:279_292, 2007.

12. M Hawkins, J; Boden. The applicability of recurrent neural networks for biological sequence analysis computational biology and bioinformatics. IEEE ACM, 2:243_253, 2005.
13. Moulton J. Stoye, V and A.W. And efficient implementation of the divided and conquer approach to simultaneous multiple sequence alignment. Biosci, 6, 2000.
14. D.G. Higgins J.D. Thompson and T.J. Gibson. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighling, positions- specific gap penalties and weight matrix choice. Nucleic Acid Research, 22, 1999.
15. A Keedwell L, Narayanan. Discovering gene networks with a neural-genetic hybrid computational biology and bioinformatics. IEEE ACM, 2:231_242, 2007.
16. Nelson Fausto Kumar, Abul K. Hypertensive vascular disease, volume 8. Saunders (Elsevier), 2009.
17. Sverlin A Manavsky and Giorgio Valle. Cuda compatible CPU as efficient hardware accelerators for smith-waterman sequence aligment. CRIBI, 2008.
18. C. Notredame and D.G Higgins. Saga: Sequence alignment by genetic algorithm. Nucleic Acid Research, 24, 2000.
19. John L. Semmlow. Biosignal and Biomedical Image Processing, Matlab based application. 2004.
20. Albert W; Sat Sharma y Claude Kortas. Hypertension and the kidney. Department of Medicine, University of Western Ontario, Canada, February 2010.
21. H. Carrillo y D.J. Lipman. The multiple sequence alignment problems in biology. SIAM J. Appl., 48:4_7, 2000.
22. Yufeng Wang Yijuan Lu; Qi Tian; Feng Liu; Sanchez M. Interactive semi supervised learning for microarray analysis. IEEE ACM, 4:190203, 2007.

An efficient embedded gene selection method for microarray gene expression data

Edmundo Bonilla Huerta, José C. Hernández Hernández,
Roberto Morales Caporal, José F. Ramírez Cruz and
Luis A. Hernández Montiel

LITI, Instituto Tecnológico de Apizaco,
Av. Instituto Tecnológico S/N. C.P. 90300
{edbonn,josechh,robertomorales,framirez}@itapizaco.edu.mx
Paper received on 17/07/10, Accepted on 23/09/10.

Abstract. In this paper an embedded LDA based gene selection method is proposed. We combine a Genetic Algorithm (GA) with Fisher Linear Discriminant Analysis (LDA) for selecting a small subset of genes of microarray gene expression data. This LDA-based GA algorithm uses a LDA classifier in its fitness function and uses the discriminant coefficients of LDA in its dedicated crossover and mutation operators. This paper studies the effect of these specialized operators on the evolutionary process. The proposed algorithm is assessed on a several well-known datasets from the literature and compared with recent algorithms. The results obtained show that our filter-embedded approach obtains globally high classification accuracies with very small number of genes to those obtained by other methods.

Key words: LDA, genetic algorithm, embedded, filter, gene selection, microarray data.

1. Introduction

Feature selection consists to select a minimal subset of *m* features from the original set of n features ($m << n$)[11]. Recently, those methods have been utilized for gene selection in microarray data. Feature selection methods may be categorized into three main families [9] 1) the filter approach, 2) the wrapper approach and 3) the embedded approach.

The filter approach select gene subsets independently of the learning algorithm that is used for classification. In most cases, the selection relies on an individual evaluation of each gene [2], therefore the interactions between genes are not taken into account. The filter methods separate the gene selection process from the classification process.

In contrast, the embedded approach evaluates each candidate gene subset according to their quality to improve sample classification accuracy. Embedded methods are generally computation intensive since the classifier must be trained for each candidate subset. Several strategies can be considered to explore the space of

Advances in Intelligent and Information Technologies.
Research in Computing Science 50, 2010, 2010, pp.289 -299.

possible subsets. Recently, evolutionary algorithms have been proposed for the analysis of microarray gene expression data [12, 22, 13, 15].

Finally, in embedded methods, an inductive algorithm is used as feature selector and classifier. A representative work of this approach is the method that uses support vector machines with recursive feature elimination (SVM/RFE) [1].

In this paper, we propose an embedded GA-LDA-based gene selection approach for gene subset selection for microarray gene expression data where Fisher's Linear Discriminant Analysis (LDA) is used to provide useful information to a Genetic Algorithm (GA) for an efficient exploration of gene subsets space. LDA has been used for several classification problems and recently for microarray data [7, 27, 28]. Experimental results show that our approach obtains globally high classification accuracies with very small number of genes to those obtained by others similar methods.

The organization of the rest of this paper is as follows. Section 2 describe in detail the Fisher's LDA method and discusses the calculus that must be done in the case of the small sample size problem. Section 3 presents our embedded method for gene selection. Section 4 shows the experimental results on seven microarray datasets and presents a table of comparisons with other well-known gene selection methods. Finally, conclusions are presented in Section 5.

2. LDA and Small Sample Size Problem

2.1. Linear discriminant analysis

LDA is one of the most effective dimension reduction and classification methods, where the data are projected into a low dimension space according to Fisher's criterion, such that the classes are well separated. As we use this method for binary classification problems, we shall restrict the explanations to this case. Given a data set of n samples consisting of two classes C_1 and C_2, with n_1 samples in C_1 and n_2 samples in C_2. Each sample is described by q variables. So the data form a matrix $X = (x_{ij})$, $i = 1, \ldots, n$; $j = 1, \ldots, q$. We denote by μ_k the mean of class C_k and by μ the mean of all the samples:

$$\mu_k = \frac{1}{n_k} \sum_{x_i \in C_k} x_i \text{ and } \mu = \frac{1}{n} \sum_{x_i} x_i = \frac{1}{n} \sum_k n_k \mu_k$$

The data are described by two matrices S_B and S_W, where S_B is the between class scatter matrix and S_W the within-class scatter matrix defined as follows:

$$S_B = \sum_k n_k(\mu_k - \mu)(\mu_k - \mu)^t \quad (1)$$

$$S_W = \sum_k \sum_{x_i \in C_k} (x_i - \mu_k)(x_i - \mu_k)^t \quad (2)$$

If we denote by SV the covariance matrix for all the data, we have $S_V = S_B + S_W$.

LDA seeks a linear combination of the initial variables on which the means of the two classes are well separated, measured relatively to the sum of the variances of the data assigned to each class. For this purpose, LDA determines a vector w such that $w^t S_B w$ is maximized while $w^t S_W w$ is minimized. This double objective is realized by the vector w_{opt} that maximizes the criterion:

$$J(w) = \frac{w^t S_B w}{w^t S_W w} \quad (3)$$

One can prove that the solution w_{opt} is the eigen vector associated to the sole eigen value of $S^{-1}{}_W S_B$, when $S^{-1}{}_W$ exists. Once this axis w_{opt} is determined, LDA provides a classification procedure (classifier), but in our case we are particularly interested in the *discriminant coefficients* of this vector: the absolute value of these coefficients indicates the importance of the q initial variables for the class discrimination.

2.2 Generalized LDA for small sample size problems

When the sample size n is smaller than the dimensionality of samples q, S_W is singular, and it is not possible to compute $S^{-1}{}_W$. To overcome the singularity problem, recent works have proposed different methods like the null space method [28], orthogonal LDA [26], uncorrelated LDA [27, 26] (see also [17] for a comparison of these methods). The two last techniques use the pseudo inverse method to solve the small sample size problem and this is the approach we apply in this work. When S_w is singular, the eigen problem is solved for $S^+{}_w S_b$, where $S^+{}_w$ is the pseudo inverse of S_w. The pseudo-inverse of a matrix can be computed by Singular Value Decomposition. More specifically, for a matrix A of size $m \times p$ such that *rank (A)* = r, if we denote by $A = U\_V\, T$ the singular value decomposition of A, where U of size $m \times r$ and V of size $r \times p$ have orthonormal columns, $\sum$ of size $r \times r$, is diagonal with positive diagonal entries, then the pseudo-inverse of A is defined as $A^+ = V \sum^{-1} UT$.

3. Embedded method

In this section we describe our embedded method (LDA-GA) for gene subset selection. First, we apply a filter Bss/Wss [7] to retain a group *Gp* of *p* top ranking genes (In this work about p = 150). Then, the LDA-based GA is used to conduct a combinatorial search within the space of size 2p. The purpose of this search is to determine from this large search space small sized gene subsets allowing a high predictive accuracy. In what follows, we present the general procedure and then show the components of the LDA-based Genetic Algorithm. In particular, we explain how LDA is combined with the Genetic Algorithm.

3.1. General GA procedure

Our LDA-based Genetic Algorithm follows the conventional schema of a generational GA and uses also an elitism strategy.

- Initial population: The initial population is generated randomly in such a way that each chromosome contains a number of genes ranging from $p \times 0.6$ to $p \times 0.75$. The population size is fixed at 100 in this work.
- Evolution: The chromosomes of the current population P are sorted according to the fitness function (see Section 3.3). The "best" 10% chromosomes of *P* are directly copied to the next population P^1 and removed from *P*. The remaing 90% chromosomes of P^1 are then generated by using crossover and mutation.
- Crossover and mutation: Mating chromosomes are determined from the remaining chromosomes of *P* by considering each pair of adjacent chromosomes. By applying our multi-parent recombination operator (see Section 3.4), one child is created each time. This child undergoes then a mutation operation (see Section 3.5) before joining the next population P^1.
- Stop condition: The evolution process ends when a pre-defined number of generations is reached (fixed at 250 generations in this work).

3.2. Chromosome encoding

Conventionally, a chromosome is used simply to represent a candidate gene subset. In our GA, a chromosome encodes more information and is defined by a couple:

$$I = (\tau; \phi)$$

where τ and ϕ have the following meaning. The first part (τ) is a binary vector and represents effectively a candidate gene subset. Each allele τ_i indicates whether the corresponding gene g_i is selected (τ_i =1) or not selected (_i=0). The second part of the chromosome (ϕ) is a real-valued vector where each ϕ_i corresponds to the

discriminant coefficient of the eigen vector for gene g_i. As explained in Section 2, the discriminant coefficient defines the contribution of gene g_i to the projection axis w_{opt}. A chromosome can be thus represented as follows:

$$I = (\tau_1, \tau_2, \ldots, \tau_p; \phi_1, \phi_2, \ldots, \phi_p)$$

The length of τ and ϕ is defined by p, the number of the pre-selected genes with a filter. Notice that this chromosome encoding is more general and richer than those used in most genetic algorithms for feature selection in the sense that in addition to the candidate gene subset, the chromosome includes other information (LDA discriminant coefficients here) which are useful for designing powerful crossover and mutation operators (see Section 3.4 and 3.5).

3.3 Fitness evaluation

The purpose of the genetic search in our LDA-GA approach is to seek "good" gene subsets having the minimal size and the highest prediction accuracy. To achieve this double objective, we devise a fitness function taking into account these (somewhat conflicting) criteria.

To evaluate a chromosome I= (τ;ϕ), the fitness function considers the classification accuracy of the chromosome (f_1) and the number of selected genes in the chromosome (f_2). More precisely, f_1 is obtained by evaluating the classification accuracy of the gene subset τ using the LDA classifier on the training dataset and is formally defined as follows[1]:

$$f_1(I) = AccuracySVM \tag{4}$$

We apply 10-fold cross-validation error estimation to calculate the accuracy of the classifier SVM for each candidate gene subset. The second part of the fitness function *f2* is calculated by the formula:

$$f_2(I) = \left(1 - \frac{m_\tau}{p}\right) \tag{5}$$

where m_τ is the number of bits having the value "1" in the candidate gene subset τ , *i.e*. the number of selected genes; *p* is the length of the chromosome

[1] For simplicity reason, we use *I* (chromosome) instead of τ (gene subset part of *I*) in the fitness function even if it is the gene subset τ that is effectively evaluated.

corresponding to the number of the pre-selected genes from the filter ranking. Then the fitness function f is defined as the following weighted aggregation:

$$f(I) = \alpha f_1(I) + (1-\alpha) f_2(I)$$
$$\text{subject to } 0 < \alpha < 1$$

where α is a parameter that allows us to allocate a relative importance factor to f_1 or f_2. Assigning to α a value greater than 0.5 will push the genetic search toward solutions of high classification accuracy (probably at the expense of having more selected genes). Inversely, using small values of α helps the search toward small sized gene subsets. So varying α will change the search direction of the genetic algorithm.

3.4 Informative LDA-based Crossover

We use the discriminant coefficients from the LDA classifier to design our specialized operators (crossover and mutation). Here, we explain how our LDA-based specialized genetic operators operates.

The crossover combines two parent chromosomes I_1 and I_2 to generate a new chromosome I_{new} in such a way that 1) top ranking genes in both parents are conserved in the child and 2) the number of selected genes in the child I_{new} is no greater than the number of selected genes in the parents. The first point ensures that ”good” genes are transmitted from one generation to another while the second property is coherent with the optimization objective of small-sized gene subsets.

Before inserting the child into the next population, I^c undergoes a mutation operation based in the LDA-coefficients to remove the gene having the lowest discriminant coefficients.

3.5 Informative LDA-based Mutation

In a conventional GA, the purpose of mutation is to introduce new genetic materials for diversifying the population by making local changes in a given chromosome. For binary coded GAs, this is typically realized by flipping the value of some bits (1→0, or 0→1). In our case, mutation is used for dimension reduction; each application of mutation eliminates a single gene (1→0). To determine which gene is removed, we use discriminant coefficients obtained from LDA classifier. Given a candidate gene subset, we identify the smallest LDA discriminant coefficient and remove the corresponding gene, this is the least informative genes among the current candidate gene subset.

4. Experiments on microarray datasets

4.1 Microarray gene expression datasets

In this section, we use 7 public microarray datasets for our experiments (more details in table 1). In this table is shown the number of genes, the number of samples and the first publication that has presented an analysis of this dataset.

Table 1. Summary of datasets used for experimentation.

Dataset	Genes	Samples	References
Leukemia	7129	72	Golub et al [2]
Colon	2000	62	Alon et al [4]
Lung	12533	181	Gordon et al [8]
Prostate	12600	109	Singh et al [21]
CNS	7129	60	Pomeroy et al [20]
Ovarian	15154	253	Petricoin et al [19]
DLBCL	4026	47	Alizadeh et al [3]

4.2 Experimental results

In order to enable a fair comparison, all the crossover operators were tested under the same conditions on seven microarray datasets (Leukemia, Colon cancer, Lung cancer, Prostate cancer, Ovarian, CNS and DLBCL). The following parameters were used in this experiment: a) population size $|P| = 50$, b) maximal number of generations is fixed at 250, c) individual length (number of pre-selected genes) $p = 150$. We use a classical mutation where each bit of an individual has a mutation probability of 0.01. For the single point and uniform crossover operators, we use a crossover probability of 0.875, whereas the general settings for our LDA based crossover operator are explained in subsection 3.4.

The results of figure 1 shown that α does have a clear influence on the search direction of the genetic algorithm. A smaller value of α allows the search to obtain smaller sized gene subsets at the price of lower classification accuracy and vice-versa.

We note that the embedded approach is able to achieve very good performance even with a small population of 30 individuals. However the population size of 100 provides a slight improvement in the sense that it offers a better compromise between two objectives: good classification accuracy and small number of genes.

Table 2 summarizes the best accuracies (in bold) obtained by other methods and by our filter-embedded approach on the seven datasets presented previously. An entry with the symbol (–) in this table means that the paper does not treat the corresponding dataset. All the methods reported in this table use a process of cross validation. Each cell contains the classification accuracy and the number of genes when this is available. We remark that each cell contains the classification accuracy and the number of genes when this is available.

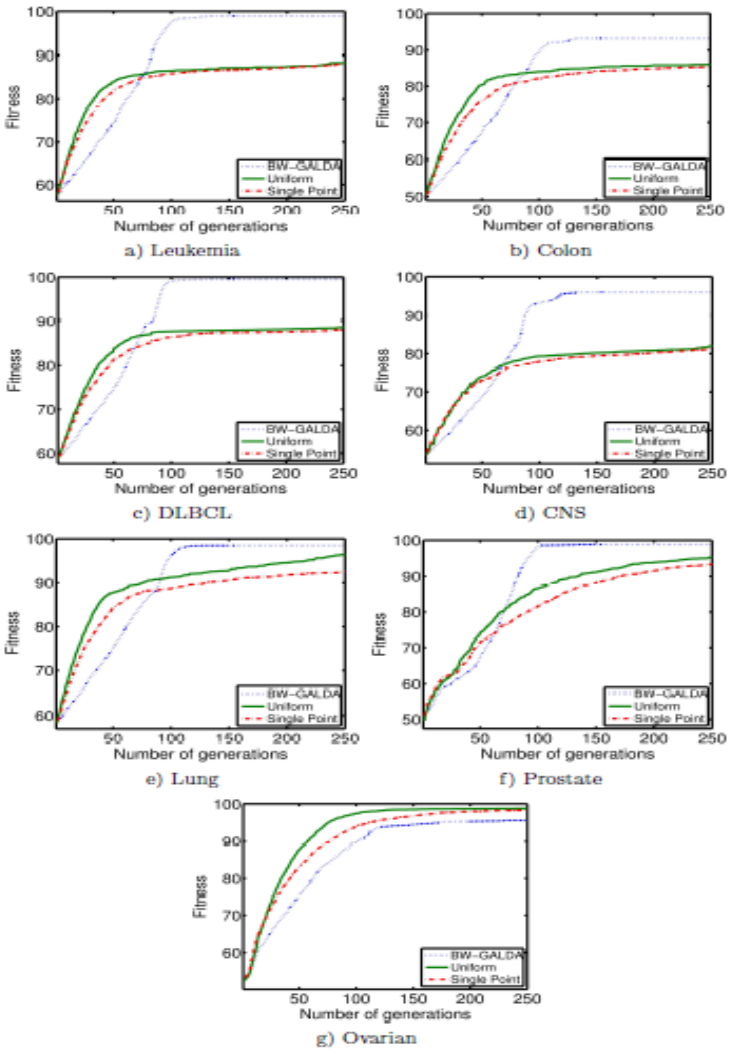


Fig. 1. Evaluation of our method with $\alpha = 0.50$

According to these observations, it seems clear that our embedded approach is very competitive in comparison with some top-performing methods. It is difficult to assess this statement by statistical tests since many methods only deal with two datasets

Table 2. Results of our embedded LDA-based GA (two last lines) compared to the most relevant works on cancer classification.

Author	Leukemia	Colon	Lung	Prostate	CNS	Ovarian	DLBCL
[6]	**100**	93.5	97.2	–	–	–	–
[5]	95.9(25)	87.7(25)	–	–	–	–	93.0(25)
[25]	73.2	84.8	–	86.88	–	–	–
[23]	95.8(20)	**100**(20)	–	–	–	–	95.6(20)
[16]	94.1(35)	83.8(23)	91.2(34)	–	65.0(46)	98.8(26)	–
[14]	97.1(20)	83.5(20)	–	91.7(20)	68.5(20)	**99.9**(20)	93.0(20)
[29]	**100**(30)	90.3(30)	**100**(30)	95.2(30)	80(30)	–	92.2(30)
[28]	83.8(100)	85.4(100)	–	–	–	–	–
[15]	**100**(4)	93.6(15)	–	–	–	–	–
our model	**100**(**4**)	95.1(**3**)	99.3(**2**)	**98.0**(**4**)	**98.8**(**2**)	99.3(**2**)	**100**(**3**)

5. Conclusions and discussion

In this paper we proposed an embedded method with specialized genetic operators for the gene selection and classification of microarray gene expression. The propose approach begins with the B/W filter that pre-selects the first 150 top-ranked genes. To further explore the combinations of these genes, we rely on a hybrid Genetic Algorithm combined with Fisher's Linear Discriminant Analysis. In this LDA-GA, LDA is used not only to assess the fitness of a candidate gene subset, but also to inform the crossover and mutation operators. This GA and LDA hybridization makes the genetic search highly efficient for identifying small and informative gene subsets.

We use a double function fitness that provides a interesting way for the LDAGA to explore the gene subset space either for the minimization of the selected genes or for the maximization of the prediction accuracy.

We have extensively evaluated our embedded approach on seven public datasets using a rigorous 10-fold cross validation process. A large comparison was carried out with 10 state-of-art algorithms that are based on a variety of methods. The results clearly show the competitiveness of our filter-embedded approach. For all the datasets, our approach is able to select small gene subsets while ensuring the best or the second best classification rate. The proposed approach has another practically useful feature for biological analysis. In fact, instead of producing a single solution (gene subset), our approach can easily and naturally provide multiple non-dominated solutions that constitute valuable candidates for further biological investigations.

Acknowledgments: This work is supported by the PROMEP project ITAPIEXB-000.

References

1. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. Machine Learning, 46(1-3):389–422, 2002.
2. 2. T. Golub, D. Slonim, et al. M. Loh, J. Downing, M. Caligiuri, C. Bloomfield, and E. Lander. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science, 286:531–537, 1999.

3. A. Alizadeh, M.B. Eisen, et al. Distinct types of diffuse large (b)–cell lymphoma identified by gene expression profiling. Nature, 403:503–511, February 2000.
4. U. Alon, N. Barkai, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Nat. Acad. Sci. USA., 96:6745–6750, 1999.
5. S.-B. Cho and H.-H. Won. Cancer classification using ensemble of neural networks with multiple significant gene subsets. Applied Intelligence, 26(3):243–250, 2007.
6. C. Ding and H. Peng. Minimum redundancy feature selection from microarray gene expression data. Bioinformatics and Computational Biology, 3(2):185–206, 2005.
7. S. Dudoit, J. Fridlyand, and T. Speed. Comparison of discrimination methods for the classification of tumors using gene expression data. Journal of the American Statistical Association, 97:77–87, 2002.
8. G.J. Gordon, R.V. Jensen, et al. S. Ramaswamy, W.G. Richards, D.J. Sugarbaker, and R. Bueno. Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. Cancer Research, 17(62):4963–4967, 2002.
9. I. Guyon and A. Elisseeff. An introduction to variable and feature selection. Machine Learning Research, 3:1157–1182, 2003.
10. K-J. Kim and S-B. Cho. Prediction of colon cancer using an evolutionary neural network. Neurocomputing, 61:361–379, 2004.
11. R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, pages 1137–1143, 1995.
12. L. Li, C.R. Weinberg, T.A. Darden and L.G. Pedersen. Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method.. Bioinformatics, 17(12):1131–1142, 2001.
13. S. Lee. Mistakes in validating the accuracy of a prediction classifier in highdimensional but small-sample microarray data. Statistical Methods in Medical Research, 17:635–642, 2008.
14. G-Z. Li, X-Q. Zeng, J.Y. Yang, and M.Q. Yang. Partial least squares based dimension reduction with gene selection for tumor classification. In Proceedings of IEEE 7th International Symposium on Bioinformatics and Bioengineering, pages 1439–1444, 2007.
15. S. Li, X. Wu, and X. Hu. Gene selection using genetic algorithm and support vectors machines. Soft Comput., 12(7):693–698, 2008.
16. S. Pang, I. Havukkala, Y. Hu, and N. Kasabov. Classification consistency analysis for bootstrapping gene selection. Neural Computing and Applications, 16:527,539, 2007.
17. H. Park and C. Park. A comparison of generalized linear discriminant analysis algorithms. Pattern Recognition, 41(3):1083–1097, 2008.
18. Y. Peng, W. Li, and Y. Liu. A hybrid approach for biomarker discovery from microarray gene expression data. Cancer Informatics, 2:301–311, 2006.
19. E. F. Petricoin, A. M. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, E. C. Kohn, and L. A. Liotta. Use of proteomic patterns in serum to identify ovarian cancer. Lancet, 359:572–577, 2002.
20. S. L. Pomeroy, P. Tamayo, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature, 415:436–442, 2002.
21. D. Singh, P. Febbo, K. Ross, D. Jackson, J. Manola, C. Ladd, P. Tamayo, A. Renshaw, A. D'Amico, and J. Richie. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell, 1:203–209, 2002.
22. F. Tan, X. Fu, et al., Improving Feature Subset Selection Using a Genetic Algorithm for Microarray Gene Expression Data, CEC-IEEE, 2529–2534, 2006.

23. Z. Wang, V. Palade, and Y. Xu. Neuro-fuzzy ensemble approach for microarray cancer gene expression data analysis. In Proc. Evolving Fuzzy Systems., pages 241–246, 2006.
24. P. Yang and Z. Zhang. Hybrid methods to select informative gene sets in microarray data classification. In Australian Conference on Artificial Intelligence, pages 810– 814, 2007.
25. W-H. Yang, D-Q. Dai, and H. Yan. Generalized discriminant analysis for tumor classification with gene expression data. Machine Learning and Cybernetics., 1:4322–4327, 2006.
26. J. Ye. Characterization of a family of algorithms for generalized discriminant analysis on under sampled problems. Journal of Machine Learning Research, 6:483– 502, 2005.
27. J. Ye, T. Li, T. Xiong, and R. Janardan. Using uncorrelated discriminant analysis for tissue classification with gene expression data. IEEE/ACM Trans. Comput. Biology Bioinform., 1(4):181–190, 2004.
28. F. Yue, K. Wang, and W. Zuo. Informative gene selection and tumor classification by null space lda for microarray data. In ESCAPE'07, volume 4614 of Lecture Notes in Computer Science, pages 435–446. Springer, 2007.
29. L. Zhang, Z. Li, and H. Chen. An effective gene selection method based on relevance analysis and discernibility matrix. In PAKDD, volume 4426 of Lecture Notes in Computer Science, pages 1088–1095, 2007.