

Un sistema experto difuso en la Web para diagnóstico de diabetes

Viridiana Cruz-Gutiérrez, Abraham Sánchez-López

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación,
México

`viricruz@rockkruz.net, asanchez@cs.buap.mx`

Resumen. La Inteligencia Artificial se ha proyectado a múltiples aspectos de la medicina, como la investigación, la asistencia clínica y la gestión sanitaria. Es por ello que en este artículo se propone un Sistema Experto Difuso para diagnóstico de diabetes mellitus usando la librería `jFuzzyLogic`, que ofrece la implementación de inferencia difusa y el API de Java para Servicios Web XML (JAX-WS). También, se describe el diseño de la base del conocimiento mediante técnicas de Ingeniería del Conocimiento e Inteligencia Artificial. El conocimiento plasmado contiene la información y experiencia de un especialista en medicina familiar del Instituto Mexicano del Seguro Social (IMSS). Con el procesamiento y uso de este conocimiento es posible diagnosticar distintos tipos de diabetes, tomando como base a los síntomas del paciente. Al final de este trabajo se presentan los resultados preliminares de la implementación.

Palabras clave: Base de conocimiento, diabetes mellitus, sistema experto.

A Web-based Fuzzy Expert System for Diabetes Diagnostics

Abstract. Artificial Intelligence is projected to multiple medicine aspects, including research, clinical care and health management. That is why this article proposes a Fuzzy Expert System for diagnosis of diabetes mellitus using the library `jFuzzyLogic` which offers implementation of fuzzy inference and Java API for XML Web Services (JAX-WS). Additionally, the design of the knowledge base is described by Knowledge Engineering techniques and Artificial Intelligence. The knowledge embodied in the system, contains the information and expertise of a specialist in family medicine at the Mexican Institute of Social Security (IMSS). With the processing and use of this knowledge is possible to diagnose different kinds of diabetes, based on the patient's symptoms. At the end of this paper the preliminary results of the implementation are presented.

Keywords: Knowledge base, diabetes mellitus, expert system.

1. Introducción

La medicina ha sido una disciplina que se ha apoyado en el uso de tecnologías y software. Surge así, la informática médica, la cual se relaciona con los sistemas de la información y de comunicación, los lenguajes médicos formales, las guías de práctica clínica, la Inteligencia Artificial y la cibernética [1].

Para dar un diagnóstico certero de diabetes mellitus, el médico se basa en los síntomas del paciente y en los criterios bioquímicos con o sin síntomas clínicos [2]. La poca experiencia de médicos principiantes, puede provocar que se realice un diagnóstico equívoco y que las recomendaciones no sean las adecuadas, poniendo en riesgo al paciente; por ello en esta investigación se muestra una propuesta de un Sistema Experto Difuso (SED) para apoyo a los médicos en el diagnóstico y prevención de diabetes, cuya base de conocimiento se diseñó mediante las técnicas de la Ingeniería del Conocimiento y que será posible consultar desde una Aplicación Web, para reducir el tiempo que les toma a los médicos dar un diagnóstico y para que complementen los consejos generales de cuidado con otros más específicos.

En la *Sección 2* se presenta una breve reseña de los conceptos más importantes acerca de la lógica difusa, la ingeniería del conocimiento y la representación del conocimiento, después, en la *Sección 3* se describen los pasos que se siguieron para el diseño de la base del conocimiento y la implementación en la Web, para que posteriormente en la *Sección 4* se presenten las pruebas de la propuesta, y los resultados obtenidos, finalmente en la *Sección 5* se dan a conocer las conclusiones y el trabajo futuro.

2. Marco teórico

Los Sistemas Expertos (SE) pertenecen a una de las áreas de mayor éxito en el campo de la Inteligencia Artificial; también son conocidos como sistemas expertos basados en conocimiento [4]. En los años setenta, en la Universidad de Stanford se desarrolló MYCIN que apoyaba a los médicos en la investigación y diagnóstico de enfermedades infecciosas de la sangre.

El conocimiento de los SE puede obtenerse por experiencia de un experto humano y/o consulta de conocimientos que están disponibles en bibliografía especializada; éste se puede documentar, revisar, complementar y utilizar en diferentes lugares y tiempos, permitiendo mejorar la toma de decisiones [5].

2.1. Lógica difusa

A diferencia de la lógica clásica, la lógica difusa [6], no tiene bien definidos los umbrales de decisión, por lo que proporciona un medio para enfrentar situaciones del mundo real que requieren razonamiento aproximado para manipular información cualitativa más que cuantitativa y situaciones complejas y dinámicas, caracterizadas en el lenguaje natural.

La lógica difusa se utiliza comúnmente para procesos difíciles de modelar matemáticamente, ya que el diseño y sintonización del controlador difuso se basa únicamente en la experiencia del experto en el proceso.

Los conceptos importantes en la lógica difusa son [8]:

- *Variable lingüística*: Concepto a calificar de forma difusa, por ejemplo: edad, altura, temperatura, etc.
- *Universo de discurso*: Rango de valores que pueden tomar los elementos que poseen la propiedad expresada por la variable lingüística.
- *Valor lingüístico*: Diferentes clasificaciones que se efectúan sobre la variable lingüística. En el caso de la temperatura, se puede dividir el universo de discurso en los diferentes valores lingüísticos: frío, templado, caliente.
- *Función de pertenencia*: Aplicación que se asocia a cada elemento del universo de discurso el grado con que pertenece al conjunto difuso. Las funciones más comunes son: Gamma, L, Lambda (triangular) y Pi (trapezoidal).
- *Conjunto difuso*: Valor lingüístico junto con una función de pertenencia.

2.2. Ingeniería del conocimiento

Al proceso de construir un SE se le denomina Ingeniería del Conocimiento [7], y consiste en la adquisición de conocimiento a partir de un especialista humano o de otra fuente y su codificación en el Sistema Experto.

El conocimiento extraído de los expertos y de la literatura, puede evolucionar durante el proceso de desarrollo, en el uso del sistema, por nuevos conocimientos del dominio o por consideraciones de los expertos humanos. Para desarrollo de un SE no es aplicable un ciclo de vida clásico ya que se requiere de un mantenimiento continuo de la base de conocimientos [9]. En la Figura 1, se muestran los tres procesos fundamentales de la Ingeniería del Conocimiento [10].

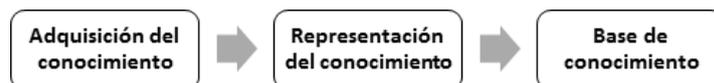


Fig. 1. Ingeniería del conocimiento de manera general.

En la *primera etapa* se tiene acercamiento con los expertos humanos para el dominio del conocimiento. La *segunda etapa* comprende la representación del conocimiento, en la cual el ingeniero del conocimiento codifica y se hacen explícitas las reglas para que el experto humano sea capaz de resolver problemas reales. En la *tercera etapa* la información puede ser representada por medio de reglas de producción (implicaciones lógicas).

2.3. Representación del conocimiento

La base del conocimiento es una colección de datos que representan conocimiento, dicha base se plasma utilizando un lenguaje de representación del conocimiento, y un programa para ampliar y/o consultar una base de conocimiento.

A partir del conocimiento existente puede crearse nuevo conocimiento utilizando la inferencia lógica [11].

$$\begin{aligned} \text{información} &= \text{datos} + \text{significado} \\ \text{conocimiento} &= \text{información} + \text{procesamiento} \end{aligned}$$

Además de su capacidad para hacer abstracciones a diferentes niveles, el conocimiento se relaciona con la acción. El uso efectivo del conocimiento conduce a la formación de planes de acción y finalmente a una profunda comprensión.

El lenguaje estructurado es el más popular para la realización de la representación del conocimiento; incluye una base en lógica de primer orden, reglas de producción, representaciones basadas en frames, combinaciones de frames y la lógica. Las reglas de producción han sido las estructuras más comunes para representar el conocimiento en los SE [12].

3. Diseño de la base de conocimiento

El diseño de la base de conocimiento para el SED propuesto, consiste en emplear las herramientas y métodos de la Ingeniería del Conocimiento, descritos en los siguientes pasos:

1. **Definición del alcance y adquisición del conocimiento:** Los dos elementos más importantes para la obtención del conocimiento son el experto humano y el ingeniero del conocimiento. Se respetaron las etapas de adquisición del conocimiento presentadas en [10], que pueden observarse en la Figura 2.
 - a) *Identificación:* Se realiza una entrevista no estructurada, en la que, el ingeniero de conocimiento plantea el problema y el experto (médico) presenta las características principales. Se aprende sobre la situación del problema y se plasman por escrito dichas características.
 - b) *Entendimiento:* Con la finalidad de deducir las posibles representaciones de la base de conocimiento, se continúa con la consulta de bibliografía médica: manuales del IMSS, artículos científicos y observaciones a personas con diabetes.
 - c) *Formalización:* Durante esta etapa, se organiza el conocimiento, se representa con una matriz de conocimiento, posteriormente con una red semántica y finalmente con reglas de producción. Estas representaciones se depuran a través de entrevistas estructuradas entre el ingeniero del conocimiento y el experto.
 - d) *Implementación:* Para la realización del prototipo, se emplea el lenguaje Java, se mapean las reglas de producción mediante jFuzzyLogic (librería de código abierto que permite diseñar controladores lógicos difusos)[16] y se generan los métodos del Servicio Web XML.
 - e) *Pruebas:* Se realizan ejecuciones del SED con los síntomas de 20 pacientes.

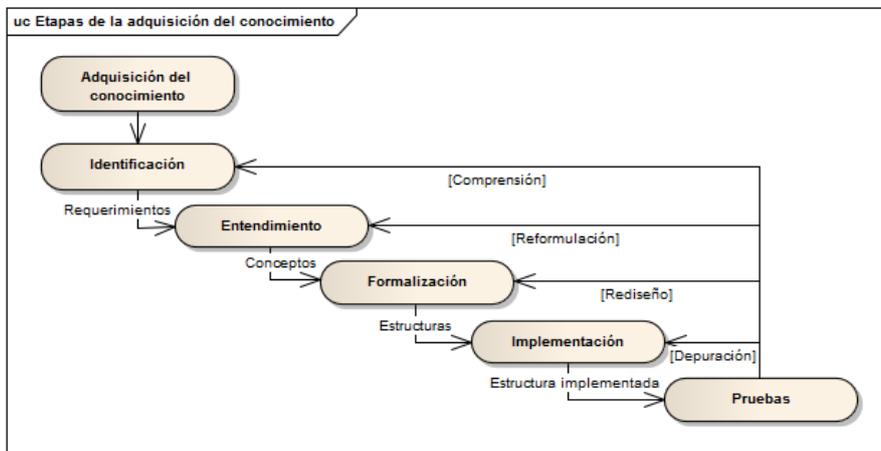


Fig. 2. Etapas de la adquisición del conocimiento.

Durante la fase de identificación al igual que en la fase de entendimiento, se definió el alcance que se tendría con el SE, y la formulación del conocimiento fundamental [13]. El dominio comprende el diagnóstico de cuatro tipos de diabetes: prediabetes, DM I, DM II y diabetes gestacional. La determinación del diagnóstico se realiza de acuerdo a las 21 variables establecidas por el experto humano junto con material bibliográfico; estas variables corresponden a los principales síntomas y características de las personas al momento de realizar un diagnóstico de diabetes.

2. **Matriz de conocimiento:** En la Tabla 1, se definen las 21 variables: binarias (B) y difusas (D), y se identifican cuáles son aplicables a cierto tipo de diabetes.
3. **Red semántica:** Para la representación del conocimiento, se empleó una red semántica, cuyos arcos están dirigidos. Los nodos representan las variables de la Tabla 1.
4. **Reglas de producción:** La codificación de la red semántica, se realizó a través de reglas de producción en jFuzzyLogic, supervisadas por el ingeniero de conocimiento y el experto, que consideran las condiciones y decisiones. Se emplea el método de encadenamiento hacia delante[4], ya que la determinación del tipo y nivel de diabetes se hace a través de un cuestionario. Las 10,000 reglas de producción consideran las variables difusas y las binarias. A continuación se muestra como ejemplo, el diseño de las reglas de producción para la determinación de la presión del paciente, se consideran la presión diastólica (pad) y la presión sistólica (pas).

```

RULE 1 : IF (pad IS normal AND pas IS normal) THEN resultado IS normal;
RULE 2 : IF (pad IS normal AND pas IS medio) OR (pad IS medio AND (pas
IS normal OR pas IS medio)) THEN resultado IS medio;
RULE 3 : IF ((pad IS normal OR pad IS medio OR pad IS alto) AND pas IS
alto) OR (pad IS alto AND (pas IS normal OR pas IS medio)) THEN
resultado IS alto;
    
```

Tabla 1. Variables definidas para cada tipo de diabetes.

No.	Criterio	Tipo	Prediabetes	DMI	DM2	Gestacional
1	Sexo	B	X	X	X	X
2	IMC	D	X		X	
3	Edad	D	X	X	X	X
4	Embarazo	B				X
5	Glucosa en ayunas	D	X	X	X	X
6	Tolerancia a glucosa	D	X	X	X	X
7	Antecedentes familiares	B		X	X	X
8	Sedentaria	B		X	X	
9	Consume alimentos grasosos	B		X	X	
10	Cansancio	B		X	X	X
11	Pérdida de peso	B		X		
12	Aumento apetito	B		X	X	X
13	Colesterol	D		X	X	
14	Presión sistólica	D		X	X	X
15	Presión diastólica	D		X	X	X
16	Hidratación	D		X	X	X
17	Micción	D		X	X	X
18	D. gestacional previa	B				X
19	Parto previo con sobrepeso	B				X
20	Bebé previo con malformaciones	B				X
21	Vómito	B				X

4. Implementación

Esta sección describe el desarrollo del SE difuso para el diagnóstico de diabetes. En la Figura 4, se presenta la arquitectura del SED propuesto. A continuación se describen los elementos de esta arquitectura:

- *Aplicación Web*: Es el mecanismo mediante el cual, el médico introducirá los datos y síntomas de sus pacientes, esta información la podrá consultar posteriormente con su usuario y contraseña. La aplicación enviará su solicitud al servicio web, que se comunicará con el SED y se realizará el fuzzificado de las entradas en el motor de inferencia para posteriormente retornar una respuesta de diagnóstico. La aplicación está desarrollada en PHP. La interfaz de la aplicación para el diagnóstico de DMI se muestra en la Figura 3; para cada tipo de diabetes se muestra una interfaz correspondiente con sus síntomas.
- *Servidor Web*: Se usó el Servidor GlassFish para el funcionamiento de los Servicios Web implementados con JAX-WS.
- *Base de datos*: Almacenará la información de entrada y salida del Servicio Web (síntomas y diagnósticos).
- *Servicio Web*: Emplea el estándar XML para realizar un intercambio de datos con los clientes desarrollados en diferentes plataformas: Aplicación Web y/o móvil. El servicio web cuenta con siete métodos (login, datosPersonales,

glucosaAyunas, toleranciaGlucosa, DMI, DMII y DGestacional) que reciben los datos y síntomas para dar una respuesta según el análisis del motor de inferencia.

- *Sistema Experto Difuso*: Sistema que cuenta con el conocimiento definido en la Sección 3.

The screenshot shows a web application titled "Sistema Experto para Diagnóstico de Diabetes". It consists of five main sections, each with a "Continuar" button:

- 1 Ingrese sus datos personales**: Fields for Nombre, Género (Femenino), Estatura (en metros), Peso (en kilos), Edad, and No embarazada.
- 2 Ingrese el valor de glucosa en ayunas**: A text input field for "Glucosa en ayunas".
- 3 Ingrese el valor de tolerancia a la glucosa**: A text input field for "Tolerancia oral a la glucosa".
- 4 Diagnóstico de Diabetes Tipo 1**: Dropdown menus for "No antecedentes familiares", "Persona no sedentaria", "Sin consumo alto de grasas y carbohi", "Sin cansancio durante el día", "Sin pérdida de peso sin razón", and "Sin aumento de apetito". Input fields for "Nivel de colesterol", "Presión sistólica", "Presión diastólica", "Litros de agua al día", and "Orina durante el día (litros)".
- 5 Resultado del diagnóstico**: Displays "Diabetes Mellitus tipo 1 en un 89%" and provides recommendations: "Recomendaciones: Repartir sus calorías en: Carbohidratos 50%, proteínas 15 - 20% y grasas (aceites vegetales) 30 - 35%. Incorporar fibra en su alimentación. Consumir fruta fresca. Realizar al menos 30 minutos de ejercicio, como trotar. Hidratarse al menos con 2.5 litros de agua."

Fig. 3. Interfaz de la Aplicación Web, (1) Datos generales del usuario, (2) Valor de glucosa en ayunas, (3) Valor de tolerancia a la glucosa, (4) Datos para diagnóstico de DMI, (5) Resultado del diagnóstico y recomendaciones.

4.1. Sistema experto difuso

En esta sección se describen los elementos para modelar el SED.

1. *Fuzzificación*: Proceso para convertir las variables en valores difusos, para ello, se utilizó la función de pertenencia triangular (1). Los parámetros de las funciones se muestran en la Tabla 2.
2. *Proceso de inferencia*: Las operaciones lógicas aplicables a los conjuntos difusos, estas son utilizadas por el motor de inferencias para derivar un resultado, tal y como lo hace el cerebro humano, utilizando el razonamiento. El motor de inferencias es implementado usando jFuzzyLogic; las variables de entrada son definidas en reglas y almacenadas en un archivo FCL que será procesado por la librería.

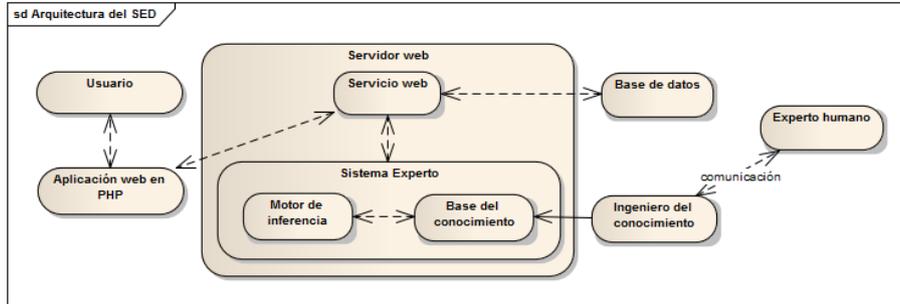


Fig. 4. Arquitectura del sistema experto difuso.

3. *Defuzzificación*: Es el proceso contrario a la fuzzificación; convierte el valor difuso en un valor nítido. Los métodos más habituales de defuzzificación son: centro de gravedad, el centro de sumas y la media de los máximos [3]. El SED utiliza método del centro de gravedad (2).

$$\text{triangle}(x; a, b, c) = \begin{cases} 0, & x \leq a. \\ \frac{x-a}{b-a}, & a \leq x \leq b. \\ \frac{c-x}{c-b}, & b \leq x \leq c. \\ 0, & c \leq x. \end{cases} \quad (1)$$

$$z^{\text{COA}} = \frac{\int_z \mu_A(z)z dz}{\int_z \mu_A(z) dz} \quad (2)$$

5. Experimentos y resultados

El sistema fue probado en un grupo de veinte pacientes de la Unidad de Medicina Familiar del ISSSTEP: 3 hombres con DMI, 3 mujeres con DMI, 5 hombres con DMII, 3 mujeres con DMII, 2 mujeres con PD, 3 mujeres adolescente con DG y 1 mujer adulta con DG. La Figura 5 a) muestra los tiempos que les tomó a los pacientes en su consulta tradicional y con el SED. La Figura 5 b) muestra el número de síntomas solicitados por el SED y los de una consulta tradicional.

Con los resultados mostrados en la Figura 5, se puede observar que el SED ayuda considerablemente a los médicos a determinar un diagnóstico en un menor tiempo al que les toma comúnmente y que además, se consideran más síntomas por parte del SED promoviendo un mejora en la productividad de los médicos. Los médicos lograron complementar las recomendaciones proporcionadas por el SED y aconsejaron a sus pacientes en tratamiento farmacológico para un mejor control de glucosa.

Tabla 2. Parámetros de funciones de pertenencia triangular.

Num	Variable lingüística	Valor lingüístico	Universo de discurso
1	IMC	Deseable	[18.5, 25, 30]
		Intermedio	[25, 30, 35]
		Elevado	[30, 35, 45]
2	Edad	Niño	[5, 9, 12]
		Adolescente	[10, 15, 20]
		Adulto	[20, 30, 50]
		Adulto mayor	[44, 60, 100]
3	Glucosa en ayunas	Normal	[, 50, 79]
		Óptimo	[60, 98, 126]
		Prediabetes	[100, 125, 130]
		Diabetes	[112, 146, 162]
4	Tolerancia a glucosa	Normal	[56, 138, 145]
		Intolerancia	[138, 170, 202]
		Alto	[195, 202, 250]
5	Colesterol	Deseable	[, 50, 200]
		Intermedio	[180, 216, 245]
		Elevado	[235, 280, 300]
6	Presión sistólica	Normal	[, 90, 130]
		Medio	[120, 130, 160]
		Alto	[130, 250,]
7	Presión diastólica	Normal	[, 60, 85]
		Medio	[80, 85, 100]
		Alto	[85, 140,]
8	Hidratación	NormalNiña	[1.3, 1.6, 2]
		AltoNiña	[1.9, 2.5, 3]
		NormalNiño	[1.3, 1.8, 2.2]
		AltoNiño	[2.1, 2.7, 3.2]
		NormalAdolescenteMujer	[1.5, 1.9, 2.3]
		AltoAdolescenteMujer	[2.1, 2.7, 3.2]
		NormalAdolescenteHombre	[2.1, 2.6, 3]
		AltoAdolescenteHombre	[2.8, 3.4, 3.8]
		NormalAdultoMujer	[2, 2.2, 2.6]
		AltoAdultoMujer	[2.5, 3, 3.4]
		NormalAdultoHombre	[2.5, 3, 3.4]
		AltoAdultoHombre	[3.2, 3.7, 4.1]
9	Micción	NormalNiña	[0.6, 0.7, 1]
		AltoNiña	[0.9, 1.75, 2.1]
		NormalNiño	[0.6, 0.8, 1.1]
		AltoNiño	[1.0, 1.8, 2.2]
		NormalAdolescenteMujer	[0.7, 1.1, 1.5]
		AltoAdolescenteMujer	[1.4, 2.5, 3.0]
		NormalAdolescenteHombre	[0.8, 1.5, 1.9]
		AltoAdolescenteHombre	[1.4, 3.5, 4.5]
		NormalAdultoMujer	[1.1, 1.5, 2]
		AltoAdultoMujer	[1.9, 3, 4.5]
		NormalAdultoHombre	[1.1, 1.7, 2.1]
		AltoAdultoHombre	[2, 4, 5]

3. Márquez-Hernández, F.A.: Cooperación entre sistemas de inferencia, métodos de defuzzificación y aprendizaje de sistemas difusos lingüísticos. Tesis doctorado, Benemérita Universidad Autónoma de Puebla, México (2013)
4. Giarratano, J., Riley, G.: Sistemas expertos: principios y programación. International Thomson, México (2001)
5. Quiroz-Hernández, J.L.: Prototipo de un sistema experto en el diagnóstico de acné. Tesis licenciatura, Benemérita Universidad Autónoma de Puebla, México (2000)
6. Cardona, C., Restrepo, C., Padilla, S.: Diseño y desarrollo de un sistema experto con lógica difusa para diagnosticar el nivel psicológico, social y criminalístico de individuos de acuerdo a su prospección a cometer ciertos delitos. Proyecto Curso Inteligencia Artificial, Universidad Nacional de Colombia, Medellín (2006)
7. Michie, D: Knowledge Engineering. In: Kybernetes. Gordon and Breach Science Publishers Ltd, Vol. 2, pp. 197–200, Inglaterra (1973)
8. Conde-Ramírez, J.C.: Modelado cognitivo en videojuegos. Tesis maestría, Benemérita Universidad Autónoma de Puebla, México (2013)
9. Fernández-Fernández, G.: Representación del conocimiento en sistemas inteligentes, disponible en: <http://www.gsi.dit.upm.es/~gfer/ssii/rcsi/>. Universidad Politécnica de Madrid, España (2004)
10. Haces-Álvarez, J.A.: Sistema para generar micromundos para la asignatura de álgebra. Tesis licenciatura, Universidad Nacional Autónoma de México, México (2011)
11. Free On-Line Dictionary of Computing (FOLDOC): Knowledge. Disponible en: <http://foldoc.doc.ic.ac.uk/foldoc/foldoc.cgi?query=knowledge>
12. Biondo, S.: Fundamental of expert systems technology Principles and concepts. Ablex Publishing Corporation, EUA (1990)
13. Aristizábal-Mejía, N., Torres-Moreno, M.E.: Técnicas de levantamiento de requerimientos con innovación. En: Cuarto Congreso Colombiano de Computación 4CCC, Sociedad Colombiana de Computación S(Co)2. Colombia (2009)
14. Osorio-Vega, C.F.: Sistema experto para determinar el tipo de diabetes. Tesis licenciatura, Universidad Tecnológica Equinoccial, Ecuador (2009)
15. Rey-Salazar, G., García-Araya, A.: Sistema experto para determinar tipo de diabetes. Ingeniería Civil Electrónica, Departamento de Ingeniería Eléctrica, Universidad de La Frontera. Chile (2007)
16. Cingolani, P., Alcalá-Fdez, J.: jFuzzyLogic: a Java Library to Design Fuzzy Logic Controllers According to the Standard for Fuzzy Control Programming. In: International Journal of Computational Intelligence Systems, Vol. 6, Supplement 1, pp. 61–75 (2013)