

Trabajo de Investigación

Diseño de un Agente Inteligente Web basado en técnicas de Inteligencia Artificial

Autor: **Jorge Roperó Rodríguez**

Director: **Carlos León de Mora**

Departamento de Tecnología Electrónica, Universidad de Sevilla.

Avda. Reina Mercedes, s/n, 41012 - Sevilla (Spain)

Tlfno.: (+034) 954 55 43 25, Fax: (+034) 954 55 27 64.

jropero@dte.us.es

Fecha

28-Junio-2006

Índice general

Índice general	3
Índice de figuras	6
Índice de tablas	7
AGRADECIMIENTOS	8
PREFACIO	10
CAPÍTULO 1. INTRODUCCIÓN.	12
1. Motivación.	12
2. Conceptos de partida.	13
3. Estructura del trabajo.	14
CAPÍTULO 2. TRATAMIENTO DEL LENGUAJE NATURAL.	15
1. Introducción.	15
2. Técnicas NLP.	16
2.1. <i>Términos índice.</i>	16
2.2. <i>Sintáctica y semántica.</i>	18
2.3. <i>Modelo de espacio vectorial.</i>	18
2.3.1. Indexado de documentos	18
2.3.2. Introducción de pesos para los términos	19
2.3.3. Coeficientes de semejanza.	20
2.4. <i>TREC.</i>	20
2.5. <i>Técnicas actuales.</i>	21
2.5.1. Herramientas Monolingües.	21
2.5.2. Herramientas multilingües.	22
3. Sistemas basados en AI.	23
3.1. <i>Redes neuronales.</i>	24
3.2. <i>Lógica borrosa.</i>	25
CAPÍTULO 3. LÓGICA BORROSA.	27
1. Inteligencia artificial.	27
2. Lógica borrosa.	27

3. Conjuntos borrosos.	28
4. Reglas SI... ENTONCES.	30
5. Borrosificación (Fuzzyfication).	30
6. Desborrosificación (Defuzzyfication).	32
CAPÍTULO 4. DISEÑO DEL AGENTE INTELIGENTE.	35
1. Viabilidad del agente.	35
<i>1.1. El modelo a seguir: Bea.</i>	35
<i>1.2. Otros agentes inteligentes en internet.</i>	36
2. Punto de partida.	37
3. Asignación de coeficientes.	39
4. Sistema de lógica borrosa.	40
5. Administrador del sistema.	42
CAPÍTULO 5. PRUEBAS Y RESULTADOS.	43
1. Introducción.	43
2. Análisis de contenidos.	43
<i>2.1. Organización.</i>	44
<i>2.2. Estudios.</i>	44
<i>2.3. P.A.S.</i>	44
<i>2.4. Profesorado.</i>	45
<i>2.5. Recursos informáticos.</i>	45
3. Definición de las pruebas.	46
4. Funcionamiento de la herramienta para las pruebas.	46
4.1. <i>Consulta en lenguaje natural.</i>	47
4.2. <i>Asignación de vectores de peso.</i>	49
4.3. <i>Modificación de vectores de peso.</i>	53
4.4. <i>Entradas al sistema de lógica borrosa</i>	54
4.5. <i>Motor de inferencia.</i>	56
4.5.1. <i>Desarrollo del motor de inferencia.</i>	57
4.5.2. <i>Definición de los conjuntos borrosos.</i>	58
4.5.3. <i>Definición de las reglas.</i>	60
4.5.4. <i>Salidas del sistema de lógica borrosa.</i>	62
4.5.5. <i>Modificaciones al sistema.</i>	67

CAPÍTULO 6. CONCLUSIONES.	70
1. Conclusiones.	70
2. Líneas a seguir.	70
APÉNDICE A. HERRAMIENTAS DE LÓGICA BORROSA.	74
1. Toolbox de lógica borrosa de Matlab.	74
<i>1.1. El editor FIS.</i>	75
<i>1.2. El editor de funciones de pertenencia.</i>	75
<i>1.3. El editor de reglas.</i>	76
<i>1.4. El visor de reglas y el visor de superficies.</i>	77
2. El programa Un-fuzzy.	78
<i>2.1. Ventana principal.</i>	78
<i>2.2. Universo de entrada.</i>	80
<i>2.3. Universo de salida.</i>	81
<i>2.4. Base de reglas.</i>	82
<i>2.5. Simulación.</i>	84
<i>2.6. Generación de código.</i>	85
APÉNDICE B. RESULTADOS DE LAS PRUEBAS	87
<i>1. Relación de Preguntas-Tipo</i>	87
<i>2. Resultados con motor de 3 entradas en Matlab.</i>	90
<i>3. Resultados del motor de 3 entradas con umbrales fijos</i>	95
<i>4. Motor de 5 entradas con umbrales fijos</i>	100
<i>5.- Motor de 5 entradas con umbrales variables</i>	104
<i>6. Ejemplos de informes de las pruebas realizadas.</i>	108
6.1.- Informe de la consulta pregunta-tipo 122. Motor de 3 entradas con umbrales fijos.	108
6.2.- Informe de la consulta pregunta-tipo 122. Motor de 5 entradas con umbrales fijos.	111
6.3. Informe de la consulta pregunta-tipo 112. Motor de 5 entradas con umbrales variables.	112
6.4. Informe de la consulta pregunta-tipo 521. Motor de 3 entradas con umbrales fijos.	114
6.5. Informe de la consulta pregunta-tipo 521. Motor de 5 entradas con umbrales variables.	118
APÉNDICE C. ARTÍCULO PRESENTADO A MELECON '06	124
Bibliografía	129

Índice de figuras

Fig. 2.1. Términos índice.....	17
Fig. 2.2. Modelo de espacio vectorial.	20
Fig. 2.3. Asociaciones entre nodos en AIR.....	24
Fig. 3.1. Ejemplos de conjuntos.....	28
Fig. 3.2. Ejemplo. Grado de pertenencia de una persona al conjunto borroso gente joven.....	29
Fig. 3.3. Ejemplo de funciones de pertenencia para la altura de una persona.....	31
Fig. 3.4. Método de implicación para una regla (cena para dos).....	32
Fig. 3.5. Método de implicación para varias reglas (cena para dos).....	33
Fig. 3.6 Resultado de la desborrosificación.....	34
Fig. 4.1. Interfaz gráfica de Bea.	35
Fig. 4.2. Agentes inteligentes de Telefónica e Ikea.....	36
Fig. 4.3. Base de datos de palabras índice.....	38
Fig. 4.4. Base de datos de posibles respuestas.	38
Fig. 4.5. Modo de operación del sistema de lógica borrosa.	41
Fig. 5.1. Consulta de usuario.	48
Fig. 5.2. Selección de preguntas estándar.....	49
Fig. 5.4. Base de conocimiento del Asistente.	52
Fig. 5.5. Base de datos de palabras y coeficientes.	53
Fig. 5.5. Motor borroso manual.....	54
Fig. 5.6. Resultados para la herramienta de Windows con un motor de inferencia de tres entradas.	55
Fig. 5.7. Resultados para la herramienta de Windows con un motor de inferencia de cinco entradas... ..	56
Fig. 5.8. Definición de reglas con Un-Fuzzy.....	57
Fig. 5.9. Forma de los conjuntos borrosos de entrada.....	59
Fig. 5.10. Forma de los conjuntos borrosos de salida.	60
Fig. 5.11. Modo de operación del sistema de lógica borrosa.....	63
Fig. 5.12. Resultados para la herramienta de Windows con un motor de inferencia de cinco entradas y umbrales variables.	67
Fig. 5.13. Resultados para un número de entradas variable (II).....	69
Fig. A.1. Esquema del editor de sistemas de inferencia borrosos.....	74
Fig. A.2. Editor FIS.	75
Fig. A.3. Editor de funciones de pertenencia.....	76
Fig. A.4. Editor de reglas.	77
Fig. A.5. Ventana principal de Un-fuzzy.	79
Fig. A.6. Barra de botones de Un-Fuzzy.....	79
Fig. A.7. Cuadro de diálogo Definir variable de entrada.....	80
Fig. A.8. Cuadro de diálogo Variables del Universo de Salida.....	82
Fig. A.9. Cuadro de diálogo Definición Rápida de la Base de Reglas.	83
Fig. A.10 Cuadro de diálogo Definición de las Reglas de la Máquina de Inferencia.	83
Fig. A.11 Cuadro de diálogo Calcular Salidas.	84
Fig. A.12. Cuadro de diálogo Análisis Paso a Paso.....	85
Fig. A.13 Cuadro de diálogo Generación de Código Fuente.....	86

Índice de tablas

Tabla 5.1. Coeficientes de concreción.....	50
Tabla 5.2. Coeficientes de tema.....	51
Tabla 5.3. Coeficientes de apartado (I).....	51
Tabla 5.4. Coeficientes de apartado (II).....	51
Tabla 5.5. Coeficientes de apartado (III).....	51
Tabla 5.6. Coeficientes de pregunta (I).....	52
Tabla 5.7. Coeficientes de pregunta (II).....	52
Tabla 5.8. Coeficientes de pregunta (III).....	53
Tabla 5.9. Coeficientes de Tema. Ejemplo.....	.65
Tabla 5.10. Resultados para un número de entradas variable (I)..	70

AGRADECIMIENTOS

Nunca nadie es lo suficientemente importante como para realizar una tarea sin ayuda, sobre todo si esta es compleja. De pequeño me enseñaron que de bien nacido es ser agradecido, así que nunca está de más recordar a todos los que hicieron posible este trabajo de investigación.

Mi primer agradecimiento debe ir, como no, para el director de este trabajo, Carlos León, quien además permitió mi entrada en este apasionante mundo de la investigación. También querría dar las gracias a Ariel Gómez por su compañerismo y su trabajo en este proyecto.

Por supuesto, dar las gracias a todo el Departamento de Tecnología Electrónica, personificando en la persona de su director, José Ignacio Escudero, que siempre estuvo apoyando en todo momento, y en la de Marisa Delgado, siempre a disposición del que lo necesita. Así mismo, agradezco a mis compañeros de despacho en estos años Juan Antonio Rodríguez Artamendi y Octavio Rivera los buenos momentos que me han hecho pasar y su compañía y apoyo en los malos momentos y a mis compañeros en mi primer proyecto en el Departamento, Iñigo Monedero y José Luis de la Vega, con quienes compartí mis primeras experiencias en la investigación.

Y abro otro párrafo para dedicárselo a Julio Barbancho, porque esto nunca hubiera sido posible sin su amistad y su apoyo. Gracias, amigo.

Y por supuesto, a Rocío, por su apoyo vital, y a mis padres, sin los cuales no estaría escribiendo estas líneas.

PREFACIO

En algunas ocasiones, todos hemos entrado en una página web en la que era difícil moverse. Contenidos ambiguos, mala organización o, simplemente, temas complejos o gran cantidad de información difícil de gestionar. Todos estos inconvenientes se ven incrementados cuando el usuario en cuestión no es un habitual de Internet por lo que es deseable dotar al portal de una herramienta que facilite la navegación y el acceso a sus contenidos.

Este trabajo presenta un método para la clasificación de los contenidos de un conjunto de conocimientos con la finalidad de contestar a consultas de usuario por medio del lenguaje natural. El sistema está basado en un motor de lógica borrosa, que aprovecha su flexibilidad para gestionar conjuntos de conocimiento acumulado. Estos conjuntos pueden constituirse en niveles jerárquicos mediante una estructura arbórea.

Uno de los objetivos finales del sistema es la implementación de un agente inteligente para la gestión de la información contenida en un portal de internet. Para ello, se realizaron pruebas con información sobre la Universidad de Sevilla (US). Esta se encuentra inmersa en el proyecto de creación de un nuevo portal web. Aunque dicho portal ya incorpora entre sus recursos un buscador, en este trabajo, realizado por el Departamento de Tecnología Electrónica (DTE) en colaboración con el Servicio de Informática y Comunicaciones (SIC) de la Universidad de Sevilla, se trata de dotar al portal de un sistema inteligente que sea capaz de indicar al usuario el camino que debe llevar para acceder a la información, siendo realizado por medio de la Inteligencia Artificial (AI) y más concretamente, mediante la lógica borrosa.

CAPÍTULO 1. INTRODUCCIÓN.

1. Motivación.

El acceso a los contenidos de un conjunto extenso de conocimiento acumulado (una base de datos, un conjunto de documentos, contenidos web, etc.) es un problema de creciente interés en los últimos años. La búsqueda de información (*Information Retrieval*, IR) es un tema que trata con grandes colecciones de material textual, siendo su objetivo satisfacer las búsquedas y las necesidades del usuario [1]. Estas necesidades se ven incrementadas cuando el usuario en cuestión no es un habitual en la materia o cuando existen contenidos ambiguos, mala organización o, simplemente, temas complejos o una gran cantidad de información difícil de gestionar.

Finalmente, las tentativas infructuosas pueden llegar a ser frustrantes por no usar el término o los términos apropiados para realizar las consultas (una máquina solo responderá adecuadamente si se le pregunta de manera exacta), pudiendo terminar todo en una paradoja: mientras menos sabe uno, más difícil es encontrar las respuestas. En muchos casos, la solución es la búsqueda de una persona experta en el asunto y, en realidad, la ayuda demandada consiste en un intérprete que tenga la capacidad de generar una búsqueda sintáctica y semánticamente correcta que conduzca a la obtención de las respuestas deseadas.

Por lo tanto, existe la necesidad de un agente que interprete la información vaga de la que disponemos, dándonos respuestas concretas que estén relacionadas de alguna manera con los contenidos del conjunto de conocimientos. Esto debe estar basado en la estimación de la certeza de la relación entre lo que hemos expresado en lenguaje natural y los contenidos almacenados en este conjunto de conocimientos.

Con el fin de solucionar estos problemas, se desarrolla un método de clasificación de los contenidos por medio de la creación de índices basados en palabras clave y un método de consulta basado en una aplicación en lógica borrosa provista de una interfaz con la que se puede interactuar en lenguaje natural. Se propone, por tanto, una aplicación de la Inteligencia Artificial (*Artificial Intelligence*, AI) basada en el uso de la lógica borrosa.

2. Conceptos de partida.

Al efecto de clarificar ideas es conveniente especificar algunos conceptos de partida:

Buscador web.

Herramienta en la que el usuario introduce palabras relacionadas con los supuestos contenidos de la web. Estas palabras son comparadas con los índices de contenido de la web y se devuelven al usuario las coincidencias encontradas.

Millones de usuarios usan hoy motores de búsqueda web como medio primario (y a veces único) para localizar información. Los motores de búsqueda más populares actualmente (Google, Yahoo, MSNSearch...) mantienen un depósito local de la siempre creciente red. Cuando un usuario realiza una consulta, los motores de búsqueda van a sus enormes depósitos e identifican los documentos más relevantes.

En todos estos motores de búsqueda, el proceso de identificación de estos documentos implica generalmente la asociación de las palabras clave (*key words*) de la consulta con los documentos del depósito local. Por lo tanto, una página web se considera relevante cuando contiene las palabras clave de la consulta. Hasta ahora, esta aproximación particular ha funcionado muy bien para solucionar la mayoría de los requerimientos del usuario, pero existen consultas en las que la simple asociación de palabras clave no es válida [2].

Agente Inteligente web.

Herramienta en la que el usuario introduce frases y/o palabras aisladas, pudiendo utilizar estructuras sintácticas de alto nivel propias del lenguaje humano, las cuales son analizadas e interpretadas por el Agente de manera tal que extrae la parte concerniente a la demanda de información relativa a contenidos de la web y devuelve las coincidencias encontradas. Con el fin de gestionar de manera eficiente la información, surge la figura del agente inteligente web. El objetivo de dicho agente debe ser encargarse de mantener una conversación con el usuario, es decir, ser capaz de contestar a cualquier pregunta del usuario como si se tratase de una persona encargada de la ayuda y dirigirlo al enlace adecuado dentro de la web.

Consulta de usuario.

Consulta realizada por el usuario al Asistente Virtual. Formalmente puede ser una expresión lingüística con una construcción del más alto nivel.

3. Estructura del trabajo.

Para el diseño del agente inteligente objeto de este trabajo de investigación ha sido necesario profundizar en los dos temas principales que intervienen en este diseño: las técnicas de Procesado del Lenguaje Natural (*Natural Language Processing*, NLP) y la lógica borrosa (*Fuzzy Logic*). El capítulo 2 está enteramente dedicado al estudio del estado del arte del primero de ellos mientras que el capítulo 3 está referido a los fundamentos de la Inteligencia Artificial, centrándose en mayor medida en la lógica borrosa.

En el capítulo 4 se explica el modo de operación del método de clasificación e inferencia, es decir, el núcleo de la aplicación, basado en el motor de lógica borrosa. Por último, el capítulo 5 presenta las pruebas realizadas sobre el banco de preguntas-respuestas más frecuentes de la Universidad de Sevilla y los resultados obtenidos.

CAPÍTULO 2. TRATAMIENTO DEL LENGUAJE NATURAL.

1. Introducción.

Durante la década pasada, hemos contemplado una aceleración notable en el crecimiento de Internet, redes de comunicación, multimedia, etc. En esta nueva era, el vehículo principal para productos y servicios digitales es el lenguaje natural, aumentando la necesidad de sistemas de ingeniería con lenguajes robustos. Los recursos del lenguaje, tanto léxicos como gramáticos, constituyen el ingrediente principal de estos sistemas.

Por esta razón, existe una fuerte necesidad de desarrollar los recursos del lenguaje para que puedan ser explotados por las distintas aplicaciones del procesado del lenguaje natural (*Natural Language Processing*, NLP). Por ejemplo, la información morfológica y sintáctica es necesaria para el desarrollo de herramientas como los correctores ortográficos y sintácticos integrados en los procesadores de texto, así como para el desarrollo de los analizadores morfológicos y sintácticos explotados por aplicaciones NLP más complejas (motores de búsqueda, sistemas de filtrado y extracción de información, sistemas de traducción automática, etc.)[3].

Las técnicas de Tratamiento del Lenguaje Natural (*Natural Language Processing*, NLP) se pueden usar para la recuperación de información (*Information Retrieval*, IR) de varias formas. El objetivo principal de aplicar el tratamiento de lenguaje natural a IR es alcanzar una mejora en los resultados de recuperación en relación con mejores valores obtenidos para la memoria y la precisión, donde la memoria – *recall* – es el número de documentos relevantes recuperados dividido por el número de todos los documentos relevantes mientras que la precisión – *precision* – es el número de documentos relevantes recuperados dividido por el número de todos los documentos recuperados.

Otro aspecto para conseguir mejoras está relacionado con la utilización del sistema de búsqueda. Si un sistema presenta, por ejemplo, una frase significativa para cada documento buscado en una lista de aciertos, se puede dar al usuario la información esencial sobre cada documento [4].

Desde el punto de vista histórico, las técnicas NLP centradas en el Indexado Motivado Lingüísticamente (*Linguistically Motivated Indexing*, LMI) han sido el foco principal de investigación; LMI ha sido diseñado y evaluado en contraste con el Índice No Lingüístico (*Non-Linguistic Indexing*, NLI) [5]. Otros usos, por ejemplo la traducción automática (MT), son usados para recuperación multilingüística.

El empleo de recursos del lenguaje natural tales como redes semánticas o bases de datos léxicas (diccionarios bilingües, por ejemplo) sólo ha sido aplicado en un rango más amplio recientemente.

Hay dos aproximaciones diferentes para integrar técnicas NLP y recursos en la IR:

- El Indexado Motivado Lingüísticamente (LMI) se usa para crear términos índice para un modelo de espacio vectorial o para sistemas de búsqueda booleana. En el primer caso, los documentos o búsquedas son convertidas en vectores con un cierto peso, devolviéndose aquellos vectores similares a los de la consulta; el segundo caso se basa en el álgebra de Boole.
- Sistemas basados en Inteligencia Artificial (*Artificial Intelligence*, AI), que tratan de emparejar una consulta con las representaciones semánticas de los textos de entrada.

2. Técnicas NLP.

Las tentativas de aplicar módulos NLP para crear índices son casi tan viejas como la recuperación automática en general. Los primeros desarrollos trataron de imitar la clasificación humana, y bastantes de estos sistemas aplicaron la indexación asignada, basada en el razonamiento humano. Un ejemplo utilizado actualmente basado en estos sistemas es el MeSH (*Medical Subject Headings*), creado con el propósito de indexar los artículos y libros sobre las ciencias biológicas [6,7].

2.1. Términos índice.

Como las experiencias con motores de búsqueda en la www muestran, buscar en grandes bases de datos mediante un solo término a menudo devuelve muchos (demasiados) aciertos. Es por tanto importante restringir estas búsquedas mediante, por ejemplo, consultas multi-término. Esto requiere la posibilidad de comparar dicha consulta con los términos índice de los documentos, que puede estar basada en dos métodos de indexado [4]:

- Precoordinación de términos índice: es el proceso de usar términos compuestos para describir un documento. Por ejemplo, este capítulo puede ser incluido en un índice con el término "recuperación de información en lenguaje natural".
- Postcoordinación de términos índice: es el proceso de usar términos simples para describir un documento pudiendo ser combinados en base a una consulta dada.

Por ejemplo, este capítulo puede ser incluido en un índice mediante las palabras "natural", "lenguaje", "información" y "recuperación". Habría que combinar estos términos basados en una consulta como por ejemplo "natural AND lenguaje AND recuperación".

Las técnicas LMI permiten el tratamiento de términos índice multipalabra de una forma diferente y quizás más elaborada, aunque en ambos casos, precoordinación y postcoordinación son posibles.

Un término índice puede ser un término solo (una palabra sola o una palabra raíz) o uno compuesto: este último puede ser un término complejo (encontrado por cualquier LMI) o un término relacionado o similar. En la figura 2.1 vemos como pueden ser los términos índice.

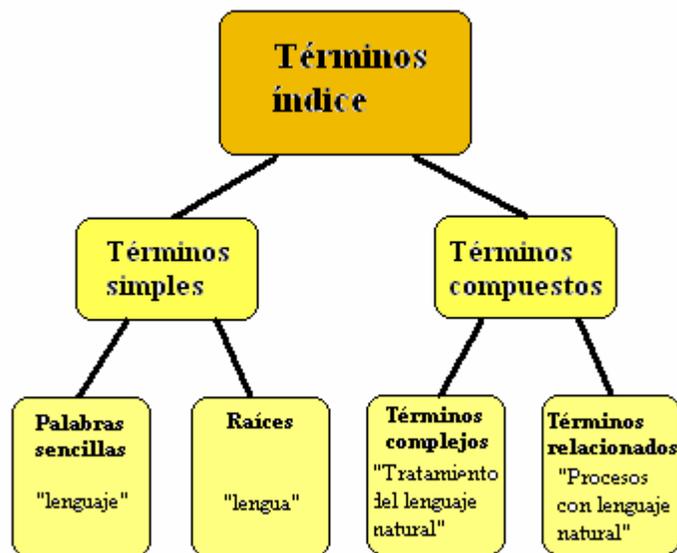


Fig. 2.1. Términos índice.

A finales de los años 60, empiezan a utilizarse los denominados tesauros (*thesaurus*). Estos tesauros (del griego, thesauros, almacén) suelen denominar un listado de palabras con significados similares, relacionados u opuestos. Por ejemplo, un libro de jerga para un campo especializado o, más técnicamente una lista de temas relacionados entre sí jerárquicamente utilizada para la indexación y recuperación de documentos.

En definitiva, un tesoro es un intermediario entre el lenguaje que encontramos en los documentos (lenguaje natural) y el que emplean los especialistas de un determinado campo del saber (lenguaje controlado) [7].

2.2. Sintáctica y semántica.

Aunque el tratamiento del lenguaje natural sea difícil, sus ventajas potenciales para la recuperación de documentos han hecho que varios investigadores investiguen el empleo tanto de un tratamiento sintáctico como de tratamiento semántico [8].

La sintaxis es la parte de la gramática que se encarga de estudiar las reglas que gobiernan la forma en que las palabras se organizan en sintagmas y, a su vez, estos sintagmas en oración, incluyendo el modo en que las oraciones se organizan en estructuras de texto. En cuanto a la semántica, esta se dedica al estudio del significado de los signos lingüísticos y de sus combinaciones [7].

Un método de aproximarse al contenido semántico de un texto es indexar con una sola palabra (*single word indexing*), que puede ser mejorado con métodos estadísticos, el tratamiento morfológico (generalmente la extracción de la raíz), y quizás algún tipo de clasificación para representar relaciones entre palabras. Esta aproximación "solo con palabras" ha disfrutado de un éxito considerable, sobre todo el llamado modelo de espacio vectorial (*vector space model*).

2.3. Modelo de espacio vectorial.

El procedimiento de modelo de espacio vectorial puede ser dividido en tres etapas. La primera etapa consiste en el indexado del documento, donde los términos más relevantes son extraídos del texto de documento. La segunda etapa se basa en la introducción de pesos para los términos indexados, con el fin de mejorar la búsqueda de lo que puede ser relevante para el usuario. La última etapa clasifica el documento en lo que concierne a la pregunta según una medida de semejanza [9].

2.3.1. Indexado de documentos

Es obvio que muchas de las palabras en un documento no describen el contenido, palabras como *el*, *ser* o *de*. Utilizando el indexado automático de documentos, estas palabras no significativas son eliminadas del vector correspondiente al documento, por lo que este sólo será representado por el contenido de las palabras relevantes. Este indexado puede estar basado en la frecuencia del término, donde los términos que tienen tanto frecuencia alta como baja dentro de un documento, se considera, son palabras función. En

la práctica, el empleo de esta técnica es difícil de poner en práctica en un indexado automático. En cambio, se utiliza una lista que contenga palabras comunes de alta frecuencia, lo que hace el indexado dependiente del lenguaje. En general, el 40-50 % del número total de palabras de un documento es suprimido con la ayuda de una de estas listas, denominadas *stop lists* [10,11].

Los métodos no lingüísticos de indexado (NLI) también han sido puestos en práctica. El indexado probabilístico está basado en la suposición de que hay alguna diferencia estadística en la distribución del contenido de las palabras relevantes y de las palabras función [11]. Excede de la finalidad de este documento una mayor profundidad en este tema, aunque se puede encontrar más información en artículos de Chakrabarti et al. y de Bookstein et al.[12,13].

2.3.2. Introducción de pesos para los términos

La introducción de pesos para los términos ha sido explicada mediante el control de la exhaustividad y de la especificidad de la búsqueda, donde la exhaustividad está relacionada con la memoria y la especificidad con la precisión. Los pesos en el modelo de espacio vectorial se basan completamente en la estadística para término simple, influyendo tres factores principalmente, el factor de frecuencia de término; el factor de frecuencia del conjunto y el factor de normalización de longitud. Estos tres factores se multiplican para conseguir el peso del término.

Un esquema para obtener los pesos de los términos dentro de un documento es usar la frecuencia de ocurrencia, mencionada en la sección anterior. La frecuencia de término es de alguna forma descriptiva del contenido de los documentos y generalmente se usa como base para la construcción de un vector para el documento. Es también posible usar vectores binarios, pero los resultados no han sido tan buenos comparados con el modelo de espacio vectorial.

Hay varios esquemas de introducción de pesos para distinguir un documento del otro. En general se denomina frecuencia de conjunto a este factor. La mayor parte de ellos, por ejemplo la frecuencia de documento inversa, asumen que la importancia de un término es proporcional al número de documentos en los que el término aparece. Experimentalmente, se ha demostrado que estos factores de discriminación de documento conducen a una recuperación más eficaz, mejorando la precisión y la memoria.

El tercer factor utilizado es el factor de normalización de longitud del documento. Los documentos largos tienen por lo general un conjunto de términos mucho más grande que los documentos cortos, lo que hace que los primeros sean más probablemente recuperados que los segundos [10, 11, 14].

2.3.3. Coeficientes de semejanza.

La semejanza en los modelos de espacio vectorial es determinada por medio de coeficientes asociativos basados en el producto escalar del vector del documento y el vector de la consulta, donde la superposición de palabras indica la semejanza. El producto escalar por lo general se normaliza. La medida de semejanza más popular es el denominado coeficiente del coseno, que mide el ángulo entre el vector del documento y el vector de la consulta [15].

Podemos representar el modelo de espacio vectorial de esta forma:

Modelo de espacio vectorial

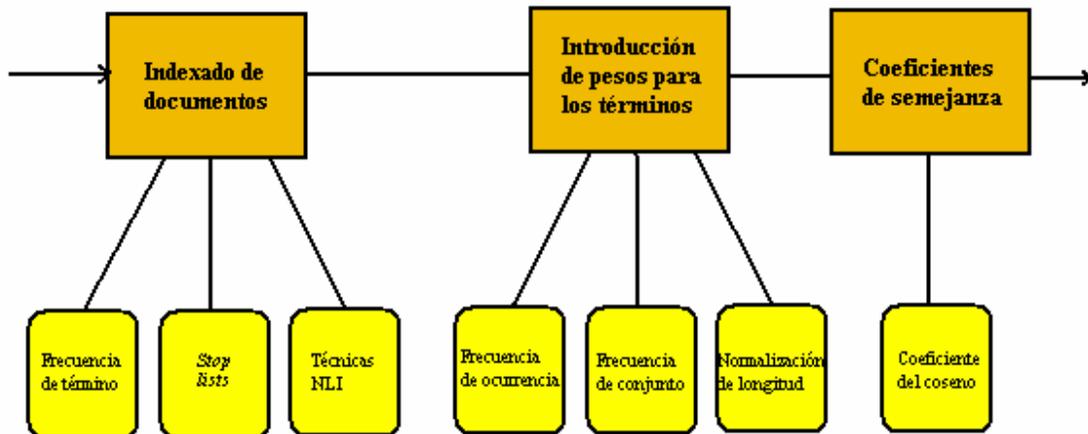


Fig. 2.2. Modelo de espacio vectorial.

2.4. TREC.

Los experimentos en IR (recordemos, recuperación de información – *Information Retrieval* -) a partir de los años 90 vienen marcados por la aparición de TREC, es decir, la Conferencia de Recuperación de Textos – *Text REtrieval Conference*.

La Conferencia de Recuperación de Textos (TREC) está copatrocinada por el Instituto Nacional de Normas y Tecnología (NIST) y el Departamento de Defensa de los Estados

Unidos, y fue puesta en marcha en 1992. Su objetivo era apoyar la investigación dentro de la comunidad dedicada a la recuperación de documentos proporcionando la infraestructura necesaria para la evaluación a gran escala de metodologías de recuperación de texto. En particular, los trabajos realizados en TREC tienen los objetivos siguientes:

- Apoyar la investigación en recuperación de documentos basada en grandes conjuntos de test.
- Fomentar la comunicación entre industria, academia, y gobierno con la creación de un foro abierto para el intercambio de ideas para la investigación.
- Acelerar la transferencia de tecnología desde los laboratorios de investigación a productos comerciales, habiéndose conseguido mejoras sustanciales de metodologías de recuperación sobre problemas del mundo real.
- Aumentar la disponibilidad de técnicas de evaluación apropiadas para su empleo industrial y académico, incluyendo el desarrollo de nuevas técnicas de evaluación más aplicables a sistemas actuales.

Este esfuerzo de evaluación ha crecido tanto en el número de sistemas participantes como en el número de tareas anuales. Noventa y tres grupos representando a 22 países participaron en 2003 TREC. TREC también ha patrocinado las primeras evaluaciones a gran escala de la recuperación de documentos en idiomas distintos al inglés (el español y el chino), la recuperación de grabaciones de voz y la recuperación a través de múltiples lenguas. Los conjuntos de prueba de TREC son bastante grandes de modo que se pueden ajustar bastante a la realidad. La mayor parte de motores de búsqueda comerciales actuales incluyen tecnología que ha sido desarrollada en primer lugar en TREC [16].

2.5. Técnicas actuales.

Las prácticas actuales para IR se están centrando sobre todo en técnicas estadísticas, aunque el objetivo de los proyectos de investigación recientes es ver si las técnicas de tratamiento del Lenguaje Natural (NLP) podrían mejorar los resultados en recuperación o añadir prestaciones adicionales a los usuarios.

2.5.1. Herramientas Monolingües.

El análisis morfológico (análisis de la estructura interna de las palabras) puede ser aplicado a los términos en las consultas y en los documentos. La idea es que los resultados de recuperación puedan ser mejorados si se reducen las variantes morfológicas de un término a una sola forma.

Una técnica simple es desprender el sufijo de la palabra basándose en una lista de finales de palabra frecuentes en la lengua considerada. Existen algoritmos eficientes para

este proceso de ‘enraízamiento’ (*‘stemming’*), dado que no se necesita consultar un diccionario. Sin embargo se producen multitud de errores debido a su carencia de información lingüística. Además, puede que las raíces producidas no correspondan a palabras existentes [4].

Un enraizador bien conocido es el enraizador Porter; otras herramientas son la enraizador-S y el enraizador Lovins. Todos ellos han sido desarrollados para el inglés. Un estudio hecho por Harman en 1991 mostró que ninguno de estos enraizadores mejora los resultados de recuperación de manera consistente [17]. Otro estudio posterior para enraizadores similares al Porter, pero para el esloveno, han mostrado que la precisión podría ser mejorada [18], concluyendo que este tipo de enraizadores puede mejorar los resultados de recuperación para lenguas morfológicamente más complejas.

Por otra parte, existen algoritmos que, desde el punto de vista lingüístico para el análisis morfológico, comprueban las formas resultantes comparándolas con las entradas de un diccionario. A pesar de la multitud de problemas como la inconsistencia o la posible incompletitud del diccionario, errores de escritura en las pruebas los nombres propios, variaciones en acentos o guiones, etc., esta técnica muestra mejoras comparadas con el algoritmo de Porter [19].

2.5.2. Herramientas multilingües.

La información multilingüe se utiliza para permitir a los usuarios el acceso a la información en una lengua distinta a su lengua materna. Algunos sistemas permiten a los usuarios formular preguntas en lenguas diferentes para luego recuperar documentos en las lenguas de la consulta. Estos sistemas se denominan multilingües aunque no realizan ninguna traducción. Los sistemas de Recuperación de Información en Lenguas Cruzadas (*Cross-language Information Retrieval, CLIR*) permiten recuperar documentos mediante preguntas en una lengua diferente de la lengua de documento. Hay varias posibilidades diferentes para conseguir este objetivo [20]:

- Traducción de documentos.
- Traducción de preguntas.
- Empleo de una interlingua para indexar. Una interlingua es una lengua internacional auxiliar (IAL), que combina un vocabulario latino románico con una gramática románica simplificada [7].

La traducción es posible durante el indexado (off-line) o como un proceso previo al proceso de recuperación (on-line). El proceso de traducción puede estar basado en tres fuentes de conocimiento de transferencia:

- Traducción automática (*Machine Translation, MT*)
- Sistemas diccionarios bilingües o tesauros

- Cuerpos paralelos (*parallel corpa*). Incluyen textos en lenguas diferentes que son traducciones el uno del otro.

Las búsquedas, en general, se traducen en on-line. Si se usan diccionarios son con este fin, cada palabra de la búsqueda es sustituida por sus posibles traducciones. La desventaja de este método es que en algunos casos la ambigüedad es difícil de eliminar, sobre todo en el caso de preguntas cortas. Esto puede llevar a la disminución de la precisión. Por otra parte, la Memoria podría aumentar, dado que los sinónimos son añadidos a la consulta. Las expresiones multipalabra o los términos técnicos son un problema en CLIR. Las traducciones basadas en palabras fallan en estos casos. Los sistemas basados en MT pueden dar resultados buenos para preguntas más largas, ya que estos sistemas pueden explotar mejor los aspectos sintácticos y semánticos del contexto para mejorar la traducción. Para preguntas cortas, no hay bastante contexto para ser explotado durante el proceso de traducción. Por último, los cuerpos paralelos pueden usarse para conseguir diccionarios bilingües. El dominio de la terminología específica puede así ser tenido en cuenta, y expresiones multipalabra pueden ser detectadas [10, 20].

3. Sistemas basados en AI.

Dado que el sistema que es objeto de estudio en este trabajo de investigación utiliza la Inteligencia Artificial (AI), es de capital importancia saber como se encuentra el estado del arte en este aspecto.

La representación del conocimiento en AI y el tratamiento de este ha tenido una influencia profunda sobre el desarrollo de 'sistemas inteligentes de recuperación de datos'. Es importante reconocer el hecho de que las estructuras de conocimiento AI representan en esencia relaciones asociativas importantes entre los datos, y que los distintos métodos de tratamiento, tanto los basados en reglas como los intuitivos, tienen una característica fundamental asociativa. Todo esto se aplica también al tratamiento del lenguaje natural práctico y, en particular, a herramientas como verificadores de ortografía, diccionarios electrónicos y tesauros [21].

Por otra parte, la inteligencia artificial (AI) se divide fundamentalmente en dos campos principales: redes neuronales y lógica borrosa. A continuación se ofrecen algunos ejemplos de aplicaciones de estas a la recuperación de información.

3.1. Redes neuronales.

Los primeros trabajos destacables que combinan redes neuronales con la recuperación de información son los realizados por R.K Belew y K.L. Kwok en 1989.

En el primero de ellos, Belew construye un sistema de recuperación adaptativo (*Adaptative Information Retrieval*, AIR). El objetivo del sistema es la recuperación de los documentos que más se parecen a aquellos que busca el usuario. Belew usa para ello una red neuronal de tres capas, correspondiendo cada capa a distintas “características”: palabras clave, documentos y autores. Cada uno de estos representa un nodo de la correspondiente capa [22].

En una línea parecida está K.L. Kwok, que también construye una red neuronal de tres capas, correspondiendo cada capa a las búsquedas, los términos índice y los documentos, respectivamente. Cada documento será evaluado en relación con los pesos activados por encima de un determinado umbral [1].

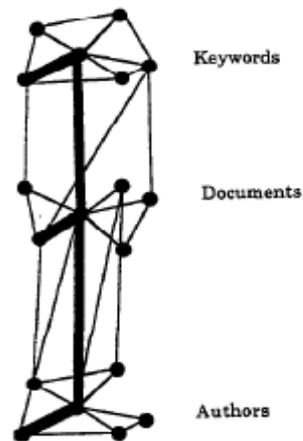


Fig. 2.3. Asociaciones entre nodos en AIR [23].

Posteriores avances en la materia se han centrado en la posible reducción de nodos por asociación de palabras, la utilización de redes neuro-semánticas, el uso de tesauros o la modificación de características para obtener mejores resultados [23-25]. Cabe también mencionar el uso de mapas de Kohonen para la recuperación de información mediante mapas semánticos autoorganizados [26].

Un ejemplo interesante es la utilización de redes neuronales por parte de Ruiz y Srinivasan para la asignación de frases MeSH (ver la primera página de este capítulo)

basada en la frecuencia de término de palabras simples del título y el abstract. Para ello se basaron en una colección de documentos Medline (2344 documentos). Cada documento de esta colección incluye título, autores, información de citas, abstract y un conjunto de términos MeSH asignado manualmente. El proceso consistió en señalar las palabras de los títulos y *abstracts*, eliminar las palabras más comunes usando una *stop list*, realizar un enraizado de palabras y, por último, computar la frecuencia de cada raíz en cada documento.

Más de 12000 palabras-raíz y 4000 frases MeSH fueron encontradas. Dada la ingente cantidad de neuronas que serían necesarias, se utilizó un umbral mínimo de frecuencia para poder pertenecer a la colección, con lo que se redujeron mucho estos términos. Aún así, la red neuronal construida tenía 1016 nodos de entrada (palabras-raíz) y 180 nodos de salida (términos MeSH). A esto hay que sumar 540 nodos de una capa intermedia. El problema principal está en el tiempo de entrenamiento (más de un día), con el consiguiente problema cada vez que haya que actualizar la red. A pesar de todo, los resultados obtenidos fueron buenos [23].

3.2. Lógica borrosa.

En cuanto a la lógica borrosa, la recuperación de información se data en multitud de fuentes como una de sus mayores aplicaciones, pero lo cierto es que no muchos autores ahondan en estas aplicaciones. Como muestra, de los 126 artículos presentados en TREC 2005, solo uno versaba sobre lógica borrosa, en concreto sobre el grado borroso de proximidad de las ocurrencias de un término de búsqueda en un documento con el fin de comprobar su relevancia con respecto a la búsqueda [27].

Cabe destacar la aplicación de la lógica borrosa a las denominadas búsquedas persistentes, en las que los usuarios necesitan una información específica durante un cierto periodo de tiempo, propuesta por Cordón et al. En esta aplicación, se realizó un indexado automático sin tener en cuenta las *stop words* y se aplicó un algoritmo de enraizado. Posteriormente se utilizó la teoría de conjuntos borrosos para la clasificación [28].

Otra aplicación interesante desde el punto de vista de la investigación en este campo es la propuesta de Subasic y Huettner, que fusiona el procesado del lenguaje natural y las técnicas de lógica borrosa para analizar el contenido afectivo de un texto [29].

En todo caso, todas estas aplicaciones tienen en común con el trabajo que se propone dos aspectos primordiales:

- La gran cantidad de información a manejar.
- La estructura jerárquica de esta información.

Por tanto, la idoneidad de la lógica borrosa y la novedad en el enfoque de este problema, nos hace decidirnos por ella para el diseño del agente inteligente.

CAPÍTULO 3. LÓGICA BORROSA.

1. Inteligencia artificial.

El camino hacia la construcción de máquinas inteligentes comienza en la Segunda Guerra Mundial con el diseño de ordenadores analógicos ideados para controlar cañones antiaéreos o para navegación. Algunos investigadores observaron entonces que existía una semejanza entre el funcionamiento de estos dispositivos de control y los sistemas reguladores de los seres vivos. El estudio unificado del control y de la comunicación en animales y máquinas fue conocido entonces como cibernética.

En torno a esta fecha comienza también la construcción de los primeros ordenadores electrónicos que, unidos a la concepción de Von Neumann, quien modeló una computadora basada en lógica digital que opera en serie las instrucciones que componen un algoritmo que se codifica en forma de programa y se almacena en memoria. El programa es ejecutado en una máquina secuencial. Como desarrollo natural de esta tendencia, se pensó en la capacidad de que estas máquinas tuvieran la capacidad de pensar racionalmente. En 1960, John McCarthy acuña el término Inteligencia Artificial (*Artificial Intelligence*, AI), para definir los métodos algorítmicos capaces de hacer pensar a los ordenadores.

Un cuarto de siglo más tarde, sin embargo, los ordenadores eran miles de veces más potentes que los de la época de los pioneros en AI y no resultaban mucho más inteligentes. El problema es que este modelo, pese a su gran potencia, presenta problemas a la hora de abordar ciertas tareas, sobre todo aquellas denominadas *del mundo real*, donde la información que se presenta es masiva, imprecisa y distorsionada. Para abordar este tipo de tareas, desde la década de los 80 se han retomado o surgido algunos paradigmas alternativos, como las redes neuronales, los sistemas borrosos, los algoritmos genéticos o la computación evolutiva, de los cuales los dos primeros son quizás los más relevantes y empleados [30].

2. Lógica borrosa.

La lógica borrosa, también llamada lógica difusa, nace en 1965 de la mano del profesor Lofti A. Zadeh [31], básicamente con el fin de enfrentarse al problema de que la lógica binaria es a veces inadecuada para la descripción del razonamiento humano. La lógica borrosa usa el intervalo completo comprendido entre 0 (Falso) y 1 (Verdadero) para caracterizar dicho razonamiento [32]. El uso de la matemática borrosa permite razonar en términos lingüísticos como pequeño, medio o rápido en vez de en términos numéricos, de manera que las ambigüedades y contradicciones pueden ser manejadas cómodamente.

La lógica borrosa surge como respuesta a la rigidez de la lógica binaria habitual. En ésta, no existe el término medio: o hace calor o hace frío. La lógica borrosa ofrece la posibilidad de tener estados intermedios: fresco, templado, cálido... además de, mediante una serie de funciones, dar un grado de flexibilidad a estos calificativos: lo que para un sevillano puede ser fresco, puede ser templado para un berlinés.

La lógica borrosa resulta útil para:

- a) Tratar la incertidumbre.
- b) Tratar la información precisa de que se dispone junto con la incertidumbre.

En este enfoque se dispone simultáneamente, por tanto, de información precisa y de incertidumbre. Utilizando lógica borrosa, se sacrifica cierta cantidad de precisión en función de la incertidumbre con la esperanza de obtener conclusiones que, aunque sean más vagas, sean más robustas [33].

Lo normal es que los conceptos imprecisos se apoyen o basen en alguna medida precisa. Una posible medida precisa para, por ejemplo, el concepto difuso “soleado” podría ser “% de cobertura de nubes”. El “% de cobertura de nubes” es absolutamente preciso.

3. Conjuntos borrosos.

Uno de los conceptos lógicos más importantes de la lógica borrosa es el de conjunto borroso (*fuzzy set*). Los conjuntos lógicos no son más que un desarrollo posterior del concepto matemático de conjunto. El pionero en el estudio de los conjuntos fue el matemático alemán Georg Cantor, a finales del siglo XIX. Según su teoría, un conjunto es una colección de objetos que puede ser tratado como un todo. Un conjunto se especifica por sus miembros, caracterizando estos por completo a dicho conjunto [32]. Un ejemplo de conjuntos puede ser observado en la figura 3.1.

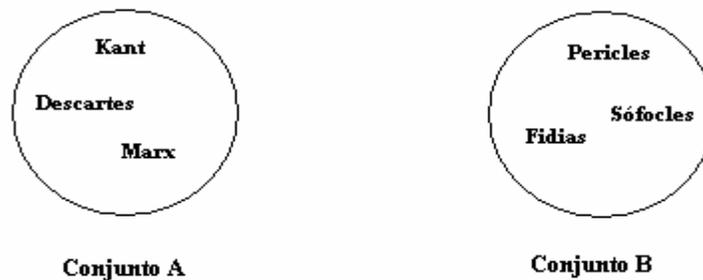


Fig. 3.1. Ejemplos de conjuntos.

Es evidente que el conjunto A representa un conjunto de filósofos, mientras que el conjunto B representa un conjunto de personajes importantes de la Grecia clásica. Si escogiéramos a Aristóteles, este estaría presente en ambos conjuntos. La lógica borrosa va, sin embargo, más allá de esta definición: los objetos pueden pertenecer a un conjunto borroso en diferentes grados, llamados *grados de pertenencia*.

Tomemos, por ejemplo, un conjunto denominado *gente joven*. Es evidente que un bebé de un año pertenece a él con grado 1 (completamente), mientras que una persona de 100 años pertenece al conjunto con grado 0. Pero, ¿qué ocurre con una persona de 20, 30, 40 o 50 años? El grado de pertenencia describe un conjunto borroso: dicho grado oscila entre 0 y 1, no existiendo una base formal para determinarlo. La pertenencia de un hombre de 50 años al conjunto borroso *gente joven* dependerá del punto de vista de cada uno, es algo preciso pero subjetivo, y está en función del contexto [32].

Un ejemplo de esto puede verse en la figura 3.2. Un observador fija el grado de pertenencia de una persona al conjunto borroso *gente joven* según su edad.

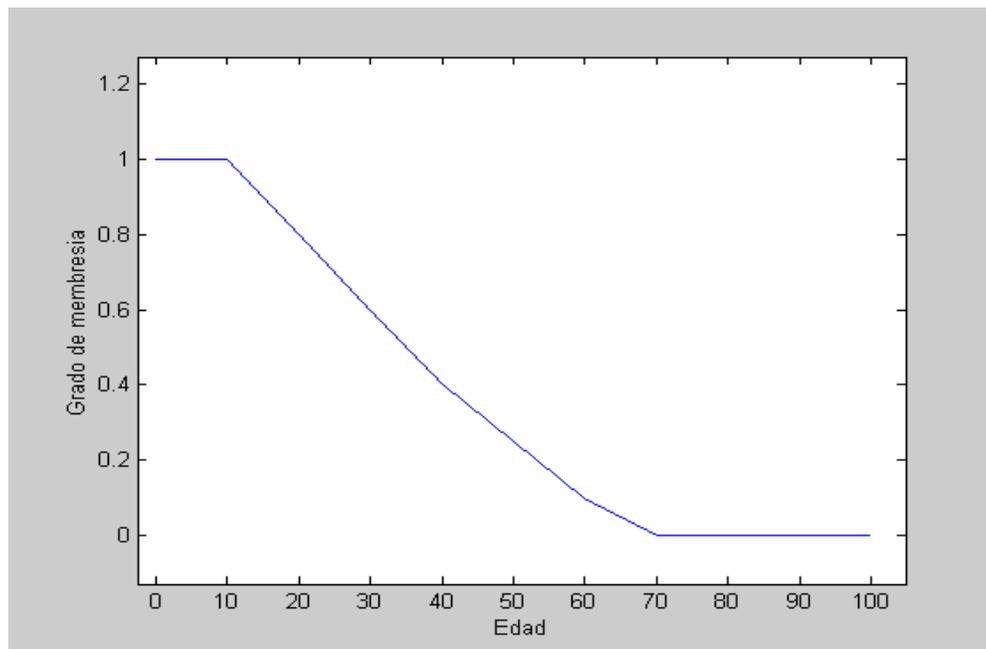


Fig. 3.2. Ejemplo. Grado de pertenencia de una persona al conjunto borroso *gente joven*.

Es evidente que esta representación del conjunto borroso no es única. En este caso, está hecha desde la visión subjetiva de una persona de 30 años (el autor). Con toda probabilidad,

esta visión sería muy diferente para una persona de 10 años u otra de 80. La gráfica dada en la figura 3.2 representa una *función de pertenencia*.

4. Reglas SI... ENTONCES.

Los conjuntos borrosos y los operadores borrosos son los sujetos y verbos de la lógica borrosa. Las reglas SI-ENTONCES (IF-THEN) son las afirmaciones condicionales que dan sentido a esta. Una regla SI-ENTONCES borrosa es del tipo “Si x es A , ENTONCES y es B ”, donde A y B son valores lingüísticos en los rangos (*universos de discurso*) X e Y , respectivamente. La parte “ x es A ” se denomina antecedente o premisa, mientras que la parte “ y es B ” se denomina consecuente o conclusión [33].

Un ejemplo sería:

SI “el tamaño del coche” es “grande”, ENTONCES “aparcar” es “difícil”

donde “grande” estaría representado por un número entre 0 y 1 (0 es diminuto y 1 enorme, por ejemplo) y “difícil” estaría representado por un conjunto borroso

Estas reglas SI-ENTONCES pueden ser compuestas (si “ x es A ” y “ z es B ”, entonces “ y es C ” y, además, admiten modificadores tales como bastante, casi, muy, etc. (si “ x es bastante A ”, entonces “ y es B ”).

5. Borrosificación (Fuzzyfication).

Para expresar un número en palabras, necesitamos una manera de traducir los valores numéricos de entrada en un conjunto borroso de descriptores lingüísticos: este es el proceso de *borrosificación* [34]. En matemática borrosa, esto es realizado por las funciones de pertenencia, de las que ya se habló en anteriores epígrafes. Estas funciones de pertenencia convierten un valor numérico en un valor lingüístico, como podemos ver en la figura 3.3. En esta figura, se comparan los conjuntos de lógica clásica con los conjuntos borrosos. En la lógica clásica, únicamente se pertenece o no a un conjunto lógico. En la lógica borrosa, se pertenece a un conjunto borroso en un cierto grado, pudiendo además un elemento pertenecer a varios conjuntos borrosos.

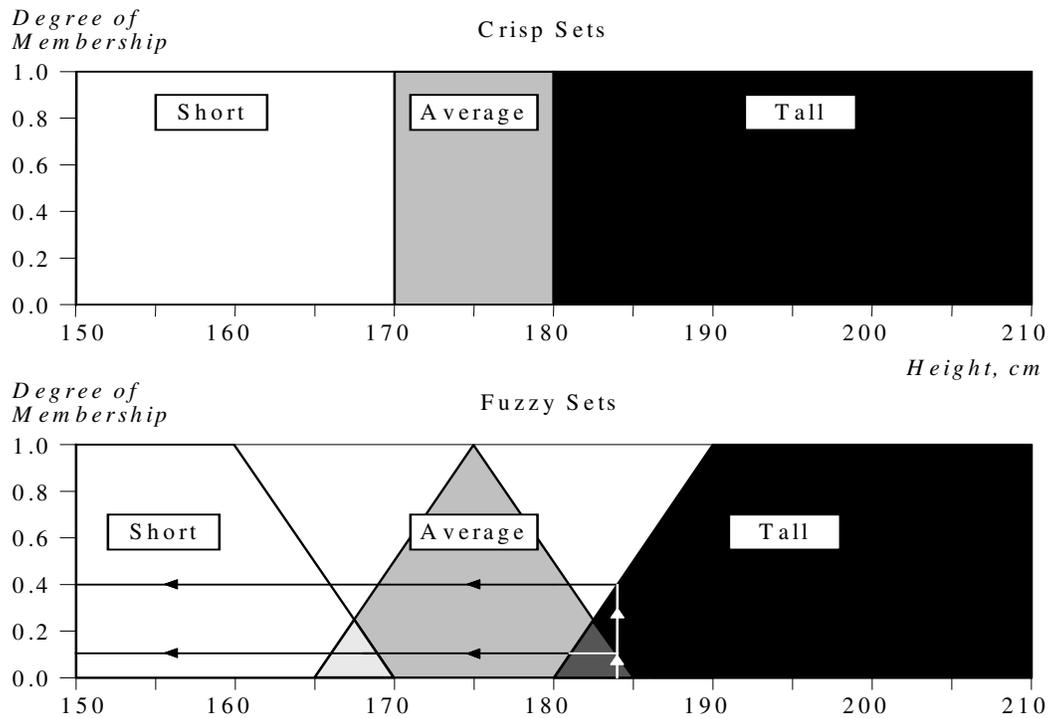


Fig. 3.3. Ejemplo de funciones de pertenencia para la altura de una persona [35].

El uso de funciones de pertenencia que se solapan, como en el ejemplo anterior, es extremadamente importante en los problemas de razonamiento borroso, ya que esto es lo que hace las fronteras entre conjuntos borrosos no estén delimitadas, pudiendo aprovechar así las ventajas de la lógica borrosa.

Por otra parte, la definición de estas funciones de pertenencia puede hacerse de múltiples formas distintas. Para ilustrarlas, baste el siguiente ejemplo.

Supongamos que estamos evaluando la altura de las personas y tenemos cinco conjuntos borrosos: MUY BAJO, BAJO, MEDIO, ALTO y MUY ALTO. ¿Qué peso tendría en cada conjunto, por ejemplo, la altura de Michael Jordan? Hay dos maneras principales de evaluarlo [36]:

- Por el método de frecuencia. Este método tiene en cuenta la respuesta afirmativa de un número de personas a una pregunta. Por ejemplo, ante la pregunta “¿Es Michael Jordan MUY ALTO?” la mayoría contestaríamos que si, pero posiblemente Pau Gasol contestaría que no. Si preguntamos a 100 personas y 99 nos dicen que si, el índice de pertenencia de

Michael Jordan (y el de todos los que midan lo mismo que él) al conjunto borroso MUY ALTO es de 0.99.

- Por estimación directa. Este método tiene la ventaja de ser mucho más sencillo de aplicar que el anterior, dado que depende de la percepción del autor del sistema. Por ejemplo, el autor de este trabajo podría asignarle a Michael Jordan un coeficiente de 0.98 en el conjunto borroso MUY ALTO. Podríamos pensar que este método tiene la desventaja de su subjetividad, pero es precisamente esta la que dota a los sistemas de lógica borrosa de una gran flexibilidad, propia del pensamiento humano.

Aplicando uno u otro método se llega a la definición de las funciones de pertenencia que, a fin de cuentas, no son más que el resultado de aplicar estas técnicas a un rango continuo de valores.

6. Desborrosificación (Defuzzification).

La desborrosificación es, como su nombre indica, el proceso opuesto a la borrosificación [34]. En el método de implicación (reglas SI... ENTONCES), cada regla es aplicada al número dado por el antecedente, construyéndose un conjunto borroso para el consecuente, dado por la aplicación del método del mínimo, que trunca el conjunto borroso de salida o del producto, que lo escala. En la figura 3.4, vemos el clásico ejemplo propuesto en el interesantísimo manual de usuario del toolbox de lógica borrosa de MATLAB, conocido como “Cena para dos”. En el gráfico, se ha aplicado el método del mínimo a una de las reglas (si el servicio es excelente o la comida es deliciosa entonces la propina es generosa). Este método debe aplicarse a todas las reglas definidas para el sistema.

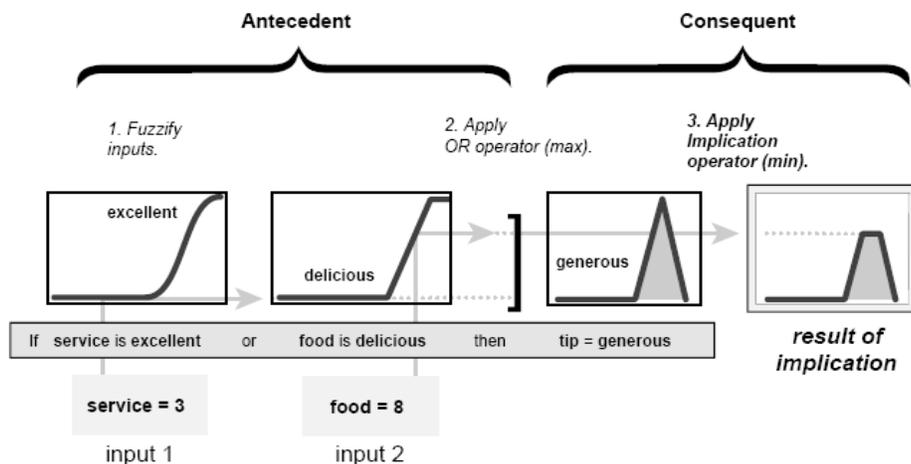


Fig. 3.4. Método de implicación para una regla (cena para dos) [33].

Aplicando todas las reglas, el resultado de la implicación para este mismo ejemplo se muestra en la figura 3.5.

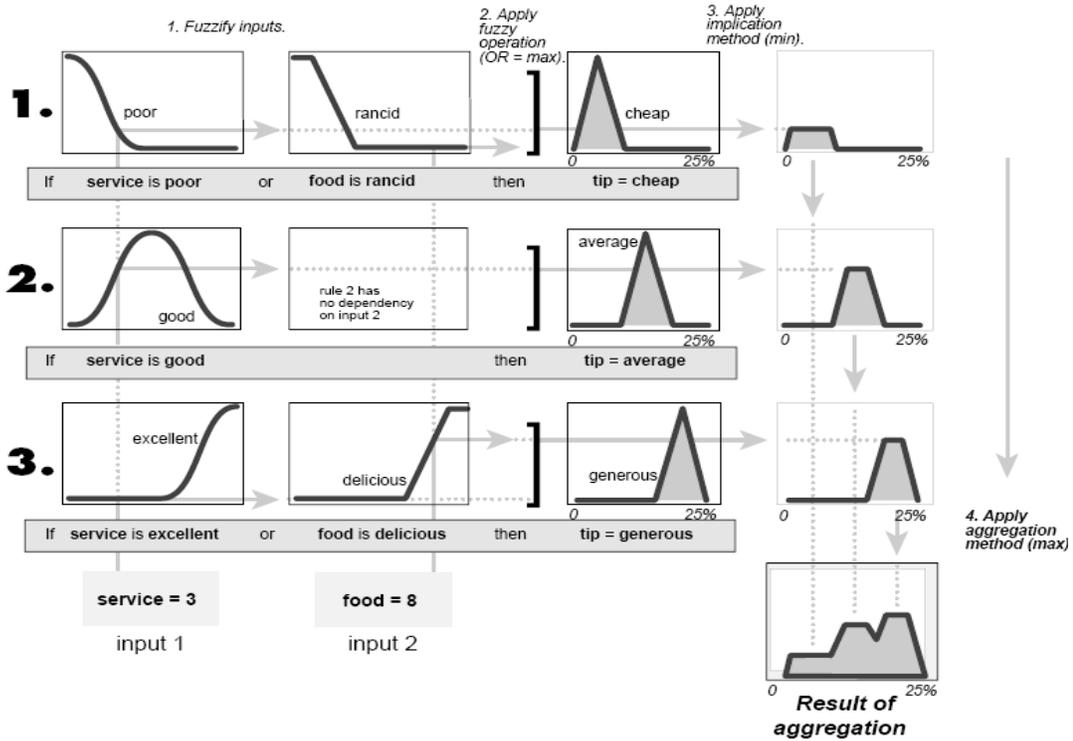


Fig. 3.5. Método de implicación para varias reglas (cena para dos) [33]

Una vez se hayan aplicado todas las reglas, se obtiene un conjunto borroso de salida, pero, por mucho que la lógica borrosa nos ayude durante los pasos intermedios en la definición de reglas, la salida deseada suele ser un valor numérico. Para hallar este valor numérico existen diversos métodos, aunque el más popular es el del centroide, consistente en hallar el centro de gravedad de la superficie que se encuentra bajo la curva de la función correspondiente al conjunto borroso de salida. Este método se ilustra en la figura 3.6.

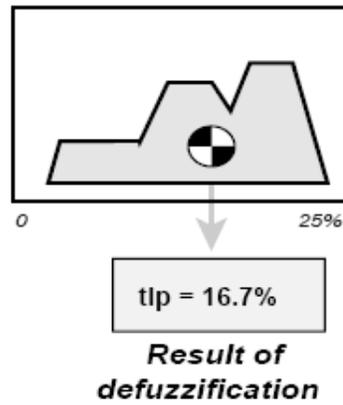


Fig. 3.6 Resultado de la desborrosificación [33].

CAPÍTULO 4. DISEÑO DEL AGENTE INTELIGENTE.

1. Viabilidad del agente.

1.1. El modelo a seguir: Bea.

En ofimática es posible encontrar muchas utilidades que pueden ayudar a la utilización de programas, la búsqueda de información o la resolución de problemas, por citar algunos ejemplos. Estas utilidades van desde bases de datos que contienen comandos a buscadores web, pasando por los asistentes que algunos programas nos proporcionan.

. Existen ya algunos agentes inteligentes en páginas web siendo el mejor de los existentes en español, sin ninguna duda, el asistente de Cajamadrid, llamado Bea. Este asistente, aparte de tener una interfaz gráfica agradable, responde a muchas de las dudas correspondientes a su ámbito (e incluso a algunas que no lo son) en forma de conversación y redireccionando al usuario a la página que considera más parecida a la consulta.

En la figura 4.1 se observa la interfaz gráfica de Bea.

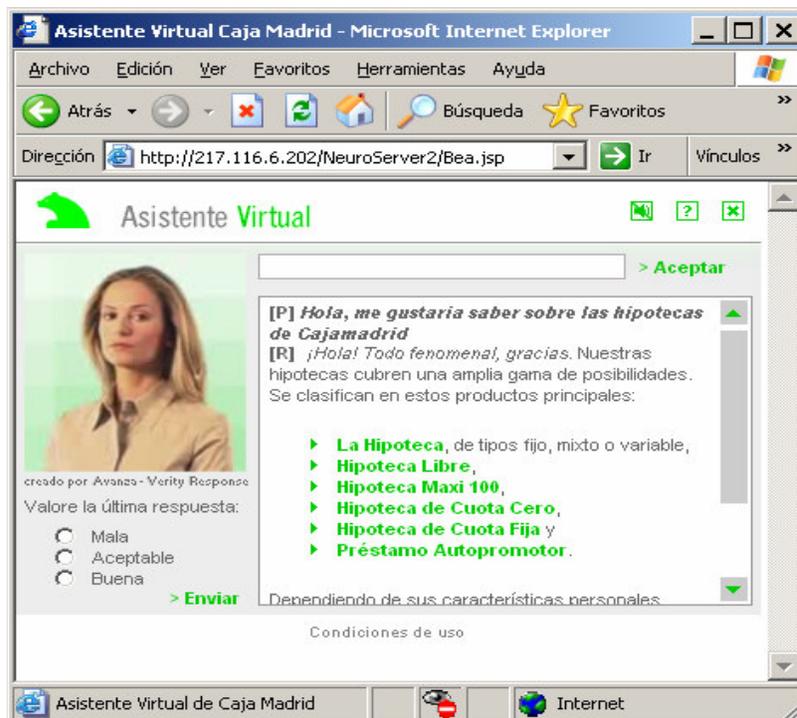


Fig. 4.1. Interfaz gráfica de Bea.

En la propia página web de Cajamadrid definen algunas características de Bea. Nos dicen que el asistente entiende mejor preguntas realizadas en lenguaje natural con una sintaxis sencilla y siempre de una en una. Si en algún caso no conoce el dato solicitado, dirigirá al usuario a páginas del Grupo Caja Madrid donde se contesta la pregunta. Si por el contrario es la pregunta lo que no entiende, le dice que no ha entendido la pregunta [37].

Por otra parte, también según la página web, el asistente ha sido sometido a más de 150.000 conversaciones y 350.000 preguntas lo que, evidentemente, es un entrenamiento importante (y trabajoso).

1.2. Otros agentes inteligentes en internet.

Existen otros agentes inteligentes en español en la red, entre los cuales podríamos destacar los de Telefónica e Ikea, los cuales podemos ver en la figura 4.2.

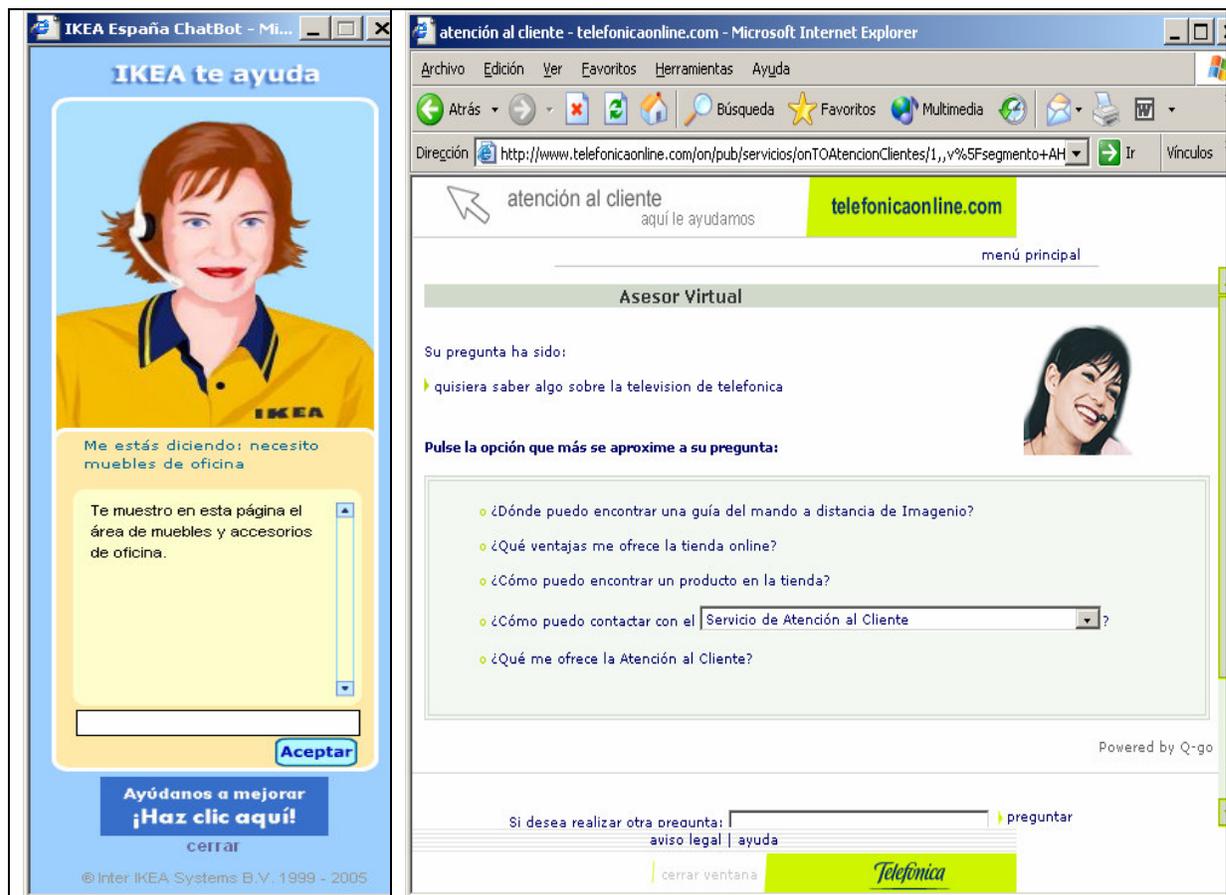


Fig. 4.2. Agentes inteligentes de Ikea y Telefónica.

Estos asistentes son sensiblemente inferiores al de CajaMadrid, aunque posiblemente se encuentren en fase de mejora.

2. Punto de partida.

Nuestro objetivo principal debe ser el diseño de un agente inteligente que permita una fácil navegación de un usuario por un portal de internet. Esto será posible mediante una interfaz que permita preguntas de los usuarios en lenguaje natural acerca de los contenidos y que también responda a estas preguntas en lenguaje natural. Así mismo se contempla la posibilidad de que el agente inteligente redirija al usuario a la dirección que estime más probable como respuesta a la consulta del usuario.

En principio, se pensó en utilizar la propia página web de la Universidad para realizar las pruebas del sistema, pero no lo hicimos por las siguientes razones:

- En el momento del comienzo de este trabajo, aún se estaba desarrollando el nuevo portal de la Universidad, por lo que aun en el caso de haber hecho las pruebas con el portal antiguo, habría que actualizar toda la información.
- La información contenida en el portal de la Universidad es ingente y, por tanto, difícil de manejar, sobre todo teniendo en cuenta que tendremos que asignar coeficientes a todas las palabras que puedan ofrecer algún tipo de información, además de tener que ser añadidas a una base de datos.

Por estas razones, se decidió trabajar con el banco de preguntas-respuestas del que disponía la Universidad de Sevilla como relación de preguntas más frecuentes y que nos ofrecía la ventaja de su mayor manejabilidad y, sobre todo, una gran similitud con la estructura de una página web, en tanto en cuanto las preguntas estaban organizadas de manera jerárquica, organizándose en grupos jerárquicos. Para más información acerca de esta estructura, se remite al capítulo 5 de este trabajo, en el punto referente al análisis de contenido.

El primer paso del procesado consiste en distinguir las denominadas *palabras clave* de la consulta de usuario. Estas palabras se buscan en una base de datos que debe contener las palabras que estén relacionadas de alguna manera con el contenido de la materia a tratar (Figura 4.3). Otra base de datos con las posibles respuestas es necesaria (Figura 4.4).

	Campo1	Campo2	Campo3	Campo4
▶	colegiados	0,8	0	0
	equipo	0	0,5	0
	gobierno	0	0,8	0
	organos	0,9	0	0
	puestos	0	0	0,7
	relacion	0	0	0,7
	sevilla	0,2	0,2	0,2
	trabajo	0	0	0,7
	universidad	0,3	0,3	0,3

Registro: 1 de 9

Fig. 4.3. Base de datos de palabras índice

Indice	Nivel	Orden	Subniveles	Descripción	Tipo
0	0	1	5	Los niveles se organizan como Tema->Apartado->Pregunta	General 08/09/2005
1	1	1	4	Tema 1.- Organización	Tema
2	1	2	7	Tema 2.- Estudios	Tema
3	1	3	4	Tema 3.- Personal de Administración y Servicios	Tema
4	1	4	5	Tema 4.- Profesorado	Tema
5	1	5	10	Tema 5.- Recursos Informáticos	Tema
6	2	1	6	A1.- Departamentos, Centros y Servicios del Tema Organización	Apartados
7	2	2	3	A2- Gobierno y Administración en la Universidad de Sevilla del Tema Organización	Apartados
8	2	3	3	A3.- Información General sobre la Universidad de Sevilla del tema Organización	Apartados
9	2	4	6	A4.- Normativas del Tema Organización	Apartados
10	2	1	8	A1.- Acceso y planes de estudio	Apartados
11	2	2	5	A2.- Alumnos extranjeros	Apartados
12	2	3	6	A3.- Becas y ayudas al estudio	Apartados
13	2	4	2	A4.- Empleo y prácticas	Apartados
14	2	5	3	A5.- Estudios de tercer ciclo	Apartados
15	2	6	9	A6.- Información general para estudiantes	Apartados
16	2	7	4	A7.- Matriculación y exámenes	Apartados
17	2	1	1	A1.- Ayudas de acción social	Apartados
18	2	2	1	A2.- Convocatoria de oposiciones y concursos PAS	Apartados

Registro: 1 de 37

Fig. 4.4. Base de datos de posibles respuestas.

Las palabras clave son asignadas a cada posible respuesta con el fin de ser identificadas. Estas palabras se eligen entre aquellas que puedan aparecer en una posible consulta de usuario. Todo el conjunto del conocimiento está agrupado en niveles jerárquicos. La pertenencia de estas palabras a cada uno de estos niveles se determina por medio de una serie de coeficientes numéricos que indican como de significativa es la palabra considerada en el nivel en cuestión. Con este objetivo, se asignan *vectores de peso* a cada palabra, conteniendo cada vector la certeza o probabilidad de pertenencia de la palabra a cada

conjunto borroso. Es importante señalar que la misma palabra puede pertenecer a distintos conjuntos.

3. Asignación de coeficientes.

El éxito del método propuesto depende en gran medida de una correcta asignación de los coeficientes de las palabras clave. El proceso consta de 3 etapas:

- 1.- Elección de las palabras clave.
- 2.- Asignación de coeficientes a las palabras clave elegidas.
- 3.- Modificación de los valores de los índices para conseguir las certezas mínimas deseadas.

1ª etapa: Elección de palabras clave.

Para cada elemento a determinar, se enuncia una pregunta en lenguaje natural cuya respuesta identifique dicho elemento. A esta pregunta se le denomina *pregunta-tipo*. Para cada una de estas preguntas-tipo se eligen las palabras clave siendo éstas las que servirán para establecer la correlación entre la consulta del usuario y la pregunta-tipo, y por tanto, para señalar el elemento buscado como respuesta.

La elección de una palabra de la pregunta-tipo como palabra clave se basa en su concreción entendiéndose por tal el grado de relación unívoca de la palabra con el elemento a identificar. Esto deja fuera a los artículos, conjunciones, verbos, etc, a no ser que por la naturaleza de la pregunta sean fuertemente significativos. Por ejemplo, en la pregunta-tipo “¿Cómo podría obtener la dirección de correo electrónico de un miembro del departamento de matemática aplicada?” serían seleccionadas dirección, correo, electrónico, departamento, matemática, y aplicada, como palabras clave.

2ª etapa: Asignación de coeficientes a las palabras clave elegidas.

Los coeficientes indican el grado de relación de la palabra clave con la respuesta (elemento a identificar). Hay que asignar a cada palabra un coeficiente por cada nivel del conjunto del conocimiento (por ejemplo, una base de datos). Para ello se utilizará un diseño bottom-up empezando por asignar los coeficientes desde el nivel más bajo (el más próximo al elemento a buscar) siguiendo hacia niveles superiores. Es necesario contemplar que en cada nivel se busca una respuesta diferente: en el nivel más bajo se busca como respuesta el elemento concreto objeto de la consulta, en el nivel anterior se busca el grupo al cual pertenece, en el nivel precedente se busca determinar el conjunto de grupos al que pertenece, y así sucesivamente.

Al estar el conjunto de conocimiento estructurado de manera jerárquica, las palabras con mayor grado de relación con la respuesta en un nivel no tienen porqué ser las mismas que en otro nivel. Dicho de otro modo, los índices asignados a una palabra varían con el nivel en que se produzca la identificación.

Para el último nivel, se usa como criterio el que hemos denominado *grado de concreción* (es decir, lo concreta que es una palabra en cuanto a su significado) de la palabra clave mientras que para niveles superiores se utiliza el criterio de “exclusividad de relación” que consiste en asignar coeficientes más altos a palabras que aparezcan en menos conjuntos a determinar. El objetivo de la determinación de nivel N (el más próximo al objeto del conjunto de conocimiento) es señalar el elemento del conjunto de conocimiento relacionado con la consulta del usuario, mientras que el objetivo de la determinación de nivel N-1 y anteriores es identificar el grupo al que pertenece el elemento del conjunto de conocimiento buscado por el usuario.

Como se menciona en el capítulo 2, la asignación de pesos para una palabra puede estar basada en la frecuencia de ocurrencia de un término, aunque este método presenta serios inconvenientes en la práctica: una palabra mencionada en muchas ocasiones podría ser muy importante para la identificación de una parte del conocimiento y a una palabra mencionada en pocas ocasiones podría ocurrirle lo contrario. La solución propuesta será la asignación de coeficientes basada en el estudio del conjunto del conocimiento por parte del administrador del sistema. Una vez asignados estos coeficientes, son susceptibles de ser modificados en la tercera etapa, como se indica en el apartado siguiente.

3ª etapa: modificación de los índices inicialmente asignados.

Una vez asignados los coeficientes en los vectores de peso, se realizan pruebas de reconocimiento de preguntas-tipo, consistente en introducir como consulta de usuario cada una de las preguntas-tipo definidas. El fin de estas pruebas es modificar los coeficientes para obtener mejores resultados.

4. Sistema de lógica borrosa.

En este apartado se describe el método de inferencia utilizado para encontrar el conocimiento demandado por la consulta de usuario.

Las palabras clave se extraen por comparación con las contenidas en la base de datos correspondiente a estas. Las entradas al sistema son los coeficientes de pertenencia a cada nivel.

La pertenencia a cada conjunto de nivel 1 se analiza considerando el valor devuelto por el motor de inferencia borroso. Si el nivel de certeza es menor que un valor prefijado (umbral), el contenido del conjunto correspondiente es rechazado. El hecho de empezar por el primer nivel y usar una estructura jerárquica posibilita el rechazo de una gran cantidad de contenido, que ya no deberá ser considerado en futuras búsquedas.

Para cada conjunto que supere el umbral mínimo de certeza, se repite el proceso, analizándose los coeficientes de pertenencia correspondientes al nivel 2. Los conjuntos en los que el grado de certeza devuelto por el motor borroso no supera un umbral mínimo son nuevamente rechazados. Si sobrepasan dicho umbral, el proceso se repite y así sucesivamente hasta el último nivel. Las respuestas definitivas corresponden a los elementos de último nivel cuya certeza supere el umbral definitivo, existiendo la posibilidad de que haya varias respuestas. Mientras más vagas sean las preguntas, más respuestas se obtienen. El proceso se observa en la figura 4.5.

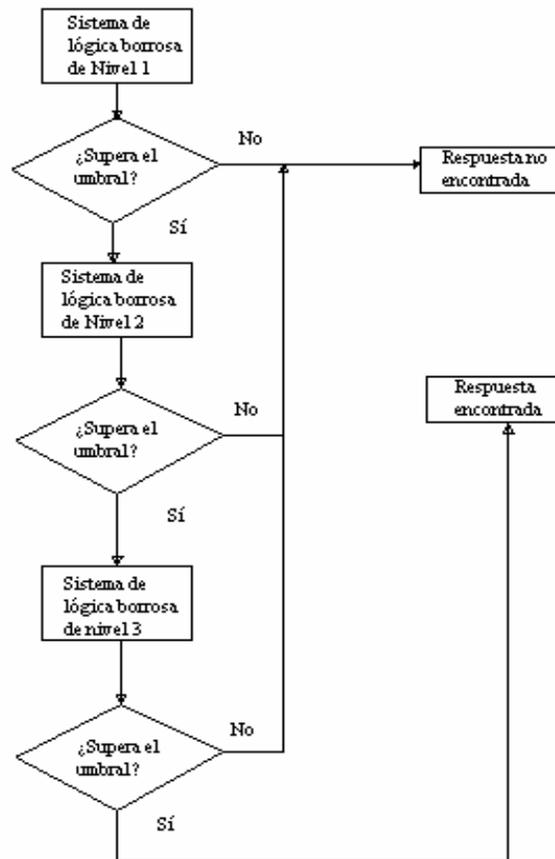


Fig. 4.5. Modo de operación del sistema de lógica borrosa.

El núcleo del sistema de lógica borrosa es el motor borroso. Este motor es el responsable de determinar la probabilidad de que las palabras clave contenidas en la consulta de usuario pertenezcan a un cierto conjunto borroso en un nivel concreto. El motor debe evaluar la pertenencia a cada conjunto para el nivel correspondiente. Por esta razón, el motor toma los vectores de pesos de las palabras clave como entradas. La salida del motor borroso quedará determinada por las reglas que sean definidas. Estas reglas son del tipo SI ... ENTONCES. Un ejemplo de una regla de este tipo sería:

SI palabra_índice 1 es ALTO Y palabra_índice2 es MEDIO y palabra_índice3 es BAJO, ENTONCES salida es ALTO.

Al igual que ocurría con la definición de los vectores de peso de las palabras clave, la definición de estas reglas concierne al administrador del sistema. Igualmente, la definición de reglas estará basada en el estudio del conjunto del conocimiento por parte del administrador del sistema y en las eventuales pruebas que pueda realizar sobre este.

5. Administrador del sistema.

Como se ha comentado en secciones anteriores, el sistema creado deberá tener un administrador del sistema. Las funciones de este administrador deben ser básicamente tres:

- 1.- Definir y modificar los coeficientes de pertenencia a un Tema, Apartado o Pregunta.
- 2.- Añadir nuevas palabras a la base de datos cuando sea necesario.
- 3.- Realizar una realimentación del sistema preguntando a los usuarios finales su opinión acerca de las respuestas dadas por el asistente, con el fin de tomar las medidas necesarias en cada caso.

CAPÍTULO 5. PRUEBAS Y RESULTADOS.

1. Introducción.

El objetivo de este trabajo de investigación es el desarrollo de un sistema para la clasificación de los contenidos en un conjunto de conocimientos con el fin de contestar a posibles consultas de un usuario en lenguaje natural. Particularmente, esto se aplicará a un portal web, aunque podría ser aplicado a cualquier otro conjunto de conocimiento con una estructura jerárquica. Para poder probar la eficacia del sistema propuesto y mejorar las prestaciones, nos basamos en el portal web de la Universidad de Sevilla, el cual cumple todas las exigencias que se habían propuesto en cuanto a que constituye un extensísimo conjunto de conocimiento y está estructurado jerárquicamente.

Como se indicó en el apartado 2 del capítulo 4 de este trabajo, para realizar las pruebas se recurrió al banco de preguntas-respuestas más frecuentes de la Universidad de Sevilla realizadas al administrador del portal. Este banco de preguntas-respuestas se constituye como el conjunto de conocimiento, consistente en un total de 117 preguntas. Los elementos a encontrar son las respuestas a estas preguntas.

2. Análisis de contenidos.

Como es lógico, para poder conducir a los usuarios por un portal de internet, es necesario conocer dos cosas: el contenido del portal y lo que los usuarios van a demandar de dicho portal.

Por otra parte, por la propia definición de lógica borrosa y su utilización como herramienta para el diseño del agente inteligente, también es necesario realizar una clasificación de estos contenidos con vistas a la posterior definición de conjuntos borrosos.

Es conveniente tomar algún punto de partida para poder desarrollar nuestra herramienta y ese es el banco de preguntas-respuestas que el portal de la Universidad ha puesto a disposición de sus usuarios hasta la fecha de comienzo de este trabajo. Este banco de preguntas-respuestas tiene una estructura jerárquica en la que las preguntas están clasificadas en distintos temas, que a su vez se subdividen en apartados que contienen las preguntas propiamente dichas y sus correspondientes respuestas. Esta clasificación es de la siguiente manera:

- A. Organización
- B. Estudios
- C. P.A.S.
- D. Profesorado
- E. Recursos informáticos

2.1. Organización.

Este apartado contiene todo lo referente a la organización de la Universidad de Sevilla, dividiéndose a su vez en cuatro subapartados:

1. Departamentos, centros y servicios (6 preguntas).
2. Gobierno y administración de la Universidad de Sevilla (3 preguntas).
3. Información general sobre la Universidad de Sevilla (3 preguntas).
4. Normativas (6 preguntas).

2.2. Estudios.

Este apartado contiene toda la información referente a los estudios que se pueden cursar en la Universidad de Sevilla, dividiéndose a su vez en siete subapartados:

1. Acceso y planes de estudio (8 preguntas).
2. Alumnos extranjeros (5 preguntas).
3. Becas y ayudas al estudio (6 preguntas).
4. Empleo y prácticas (2 preguntas).
5. Estudios de tercer ciclo (3 preguntas)
6. Información general para estudiantes (10 preguntas).
7. Matriculación y exámenes (4 preguntas).

2.3. P.A.S.

Este apartado trata de todo lo relacionado con el Personal de Administración y Servicios (P.A.S). El apartado se subdivide en cuatro categorías:

1. Ayudas de acción social (1 pregunta).
2. Convocatoria de oposiciones para el P.A.S. (1 pregunta).
3. Formación P.A.S. (2 preguntas).
4. Recursos humanos (3 preguntas).

2.4. Profesorado.

En este apartado se encuentra la información relacionada con el Personal Docente e Investigador (P.D.I.) de la Universidad de Sevilla. Se subdivide en cinco apartados:

1. Becas de investigación (2 preguntas).
2. Concursos y oposiciones PDI (2 preguntas).
3. Docencia (8 preguntas).
4. Formación P.D.I. (3 preguntas).
5. Investigación (4 preguntas).

2.5. Recursos informáticos.

Este apartado contiene la información acerca de los recursos informáticos de la Universidad, tales como el correo electrónico o el alojamiento web. Se subdivide en 10 apartados:

1. Acceso a bases de datos bibliográficas. (4 preguntas).
2. Alojamiento de páginas web en el portal de la Universidad (10 preguntas).
3. Consulta de datos académicos (5 preguntas).
4. Correo electrónico (5 preguntas).
5. Foros (4 preguntas).
6. Incidencias (2 preguntas)
7. Localizar a personas, Centros, Departamentos o Servicios de la Universidad (5 preguntas).
8. Otros servicios informáticos de la Universidad (5 preguntas).
9. PAUTA (6 preguntas).
10. Tarjeta universitaria (6 preguntas).

Puede resultar llamativo el hecho de que los apartados no sean homogéneos pero tampoco es homogéneo el conocimiento humano, por lo que este hecho no debe perjudicar al sistema de lógica borrosa. Por otra parte, esta división de contenidos nos debe servir, en principio, y como se ha mencionado anteriormente, como punto de partida, pudiendo esta clasificación ser cambiada con el fin de obtener un mejor comportamiento del sistema o bien un mejor nivel conversacional del agente inteligente.

Por último, cabe recordar que esta división no es la definitiva, sino que nos servirá de punto de partida para estudiar el diseño del agente inteligente. Una vez diseñado este, se procederá al estudio de la clasificación de los contenidos de la web real de la Universidad.

3. Definición de las pruebas.

Como se ha dicho en apartados anteriores, las preguntas elegidas para las pruebas se denominan *preguntas-tipo*. Las pruebas se basan en la utilización de estas preguntas estándar como consultas de usuario.

La primera meta de estas pruebas es comprobar que el sistema hace una identificación correcta de estas preguntas-tipo con un índice de certeza mayor que 0.7. El hecho de utilizar la lógica borrosa posibilita la identificación, no solo de la pregunta-tipo correspondiente, sino de otras. Este concepto está relacionado con el de *memoria – recall* -, mencionado en el capítulo 2 de este trabajo, aunque no corresponde a su definición exacta.

El segundo objetivo es comprobar si la respuesta a la pregunta-tipo requerida está entre las tres respuestas con mayor grado de certeza. Estas tres respuestas son las que deben ser presentadas al usuario, debiendo estar la respuesta correcta entre estas tres posibilidades. Este aspecto está relacionado con el concepto de *precisión*, aunque de nuevo no corresponde a su definición exacta.

Para realizar estas pruebas, se construyó una herramienta con la posibilidad de realizar consultas de usuario y utilizar el motor borroso para obtener las respuestas deseadas. El proceso realizado por esta herramienta es descrito en el apartado siguiente.

4. Funcionamiento de la herramienta para las pruebas.

Como se menciona en el apartado anterior, se ha desarrollado una herramienta para la realización de las pruebas. Para ello se introducirán las denominadas preguntas-tipo en lenguaje natural.

Para construir la herramienta se propone utilizar Borland C++ Builder, debido a la facilidad de construcción de la interfaz visual, la robustez de la aplicación generada y la posibilidad de compilar el código para Sistemas Operativos Windows o UNIX, posibilidad que permite integrar la aplicación en el Sistema Operativo deseado.

Por otra parte, para el motor de lógica borrosa se eligió en principio el programa Matlab, con su toolbox de lógica borrosa, debido a su facilidad de uso, la posibilidad de integrar otros toolboxes y la potencia de la herramienta. El problema principal de esta aplicación es que no exporta código C de la totalidad de lo implementado, cuestión necesaria para integrar el desarrollo en una aplicación independiente, como es nuestra intención con el Agente Inteligente. Este problema de exportación de código es resuelto con la herramienta Un-Fuzzy, que es un software para diseño de sistemas de lógica difusa de la Universidad de Colombia, de libre distribución y que exporta código C.

Por tanto, el programa Un-Fuzzy fue el elegido finalmente para el desarrollo del motor borroso para la herramienta de pruebas.

Los pasos que deberá realizar la herramienta para llevar a cabo las pruebas son los siguientes:

1. Recibir por teclado las preguntas-tipo como consulta de usuario. La consulta deberá estar hecha en lenguaje natural y será descompuesta en palabras.
2. Comparar estas palabras con un catálogo (base de datos), en el que cada palabra lleva aparejado un índice indicativo del grado de pertenencia a ciertos conjuntos borrosos definidos.
3. Seleccionar de un número de palabras con el más alto grado de referencia a contenidos.
4. Aplicar la lógica difusa para obtener una estimación (grado de certeza) del grupo de contenidos (tema) al que se refiere la pregunta mediante el motor de inferencia de lógica borrosa.
5. Repetir este proceso para todos los niveles de la jerarquía de manera que las palabras con un mayor grado de pertenencia a un nivel se vuelven a enviar al motor de inferencia, es decir, al sistema de lógica difusa.
6. El motor de inferencia determina en última instancia cual o cuáles son las preguntas tipo a las que se refiere la consulta de usuario mostrando las respuestas correspondientes ordenadas en base a la probabilidad de acierto inferida y que superen un umbral previamente fijado.

Estos puntos serán ampliados en los siguientes apartados siendo adornados además con ejemplos que pretenden clarificar el funcionamiento del sistema y justificar las decisiones que se tomaron en cuanto a la configuración de los parámetros correspondientes al sistema de lógica borrosa.

4.1. Consulta en lenguaje natural.

Para realizar las pruebas, lo primero que debemos hacer es introducir cada una de las preguntas-tipo en la herramienta. Dentro de cada pregunta-tipo se determinan una serie de palabras claves (palabras más significativas) que servirán para identificar esa pregunta como relacionada con la consulta. La misma palabra clave puede encontrarse en más de una pregunta. En cada consulta, estas palabras deben ser buscadas en una base de datos.

Recordemos que el conocimiento total del que disponemos, entendido como la información contenida en el banco de preguntas-respuestas mencionado anteriormente, se clasifica en tres niveles:

- Nivel 1: Tema.
- Nivel 2: Apartado.
- Nivel 3: Pregunta.

De esta manera, cada consulta de usuario se identifica como una parcela de conocimiento que pertenece a un Tema, a un Apartado y a una Pregunta. También puede pertenecer a varios temas, apartados o preguntas.

Tomemos como ejemplo la pregunta tipo “¿Cómo puedo localizar información sobre un Centro de la Universidad de Sevilla?” Esta pregunta es la primera de la que disponemos en el banco de preguntas – respuestas de la Universidad de Sevilla. (Tema1, Apartado1, Pregunta1) y en ella se definen las siguientes palabras clave: localizar, información, centro, Sevilla, universidad.

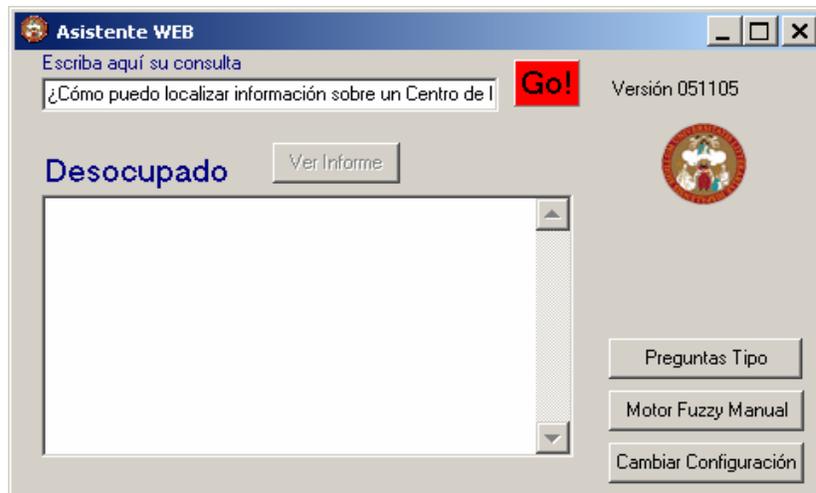


Fig. 5.1. Consulta de usuario.

Las palabras clave que deben ser añadidas a la base de datos serán todas aquellas tengan alguna relevancia en el conjunto de conocimiento, en este caso, en el conjunto de todas las preguntas estándar definidas en el apartado 2 de este capítulo. Así mismo, en el futuro habrán de considerarse aquellos aspectos de Tratamiento del Lenguaje Natural descritos en el capítulo 2, tales como el enraizado (p.e., localizo por localizar), los sinónimos (p.e., encontrar por localizar) y las expresiones compuestas (p.e., Universidad de Sevilla).

En la herramienta, es posible enviar al campo consulta las preguntas tipo contenidas en la base de datos de conocimiento. De esta manera se efectúan las pruebas de identificación de preguntas-tipo de forma rápida y cómoda, como se observa en la figura 5.2.

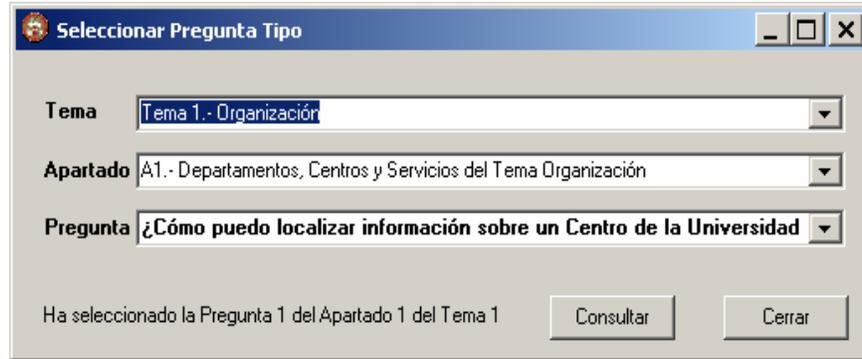


Fig. 5.2. Selección de preguntas estándar.

4.2. Asignación de vectores de peso.

Al conjunto de palabras clave se le debe asignar un vector de peso, tal y como se describe en el capítulo 2 de este trabajo. En nuestro caso, el primero de los coeficientes viene a indicar el grado de significación de esa palabra, es decir, cuánto identifica un contenido. Este coeficiente ha sido denominado coeficiente de concreción. El resto de coeficientes hacen referencia al grado de pertenencia a cada elemento de cada nivel.

Tal y como se describe en el capítulo 4, la asignación de coeficientes podría realizarse basándose en la frecuencia de ocurrencia de un término. Las dificultades para su aplicación descritas en dicho capítulo, nos hacen decantarnos por otra solución. La solución propuesta será la asignación de coeficientes basada en el estudio del conjunto del conocimiento por parte del administrador del sistema. Una vez asignados estos coeficientes, son susceptibles de ser modificados basándose en los resultados obtenidos.

Por ejemplo, en el caso anterior podríamos considerar las palabras centro, Sevilla y Universidad. La asignación inicial de coeficientes se basa en la percepción del administrador del sistema. En el caso del coeficiente de concreción, es decir, el grado de significación de una palabra, variará entre 0 y 1, siendo 0 una palabra que no aporta ningún significado y 1 una palabra que aporta el máximo de significado. Un ejemplo lo tenemos en la tabla 5.1.

Palabra	Coficiente de concreción
centro	0,6
sevilla	0.3
universidad	0.3

Tabla 5.1. Coeficientes de concreción.

De la misma manera, se definen los vectores de peso que contienen el grado de pertenencia entre 0 y 1 a los distintos niveles, en el caso del ejemplo Tema, Apartado y Pregunta.

Nivel 1: Grado de pertenencia a cada Tema

Palabra	Tema1	Tema2	Tema3	Tema4	Tema5
centro	0,8	0,5	0	0.1	0.5
sevilla	0.2	0.5	0	0.25	0.2
universidad	0.4	0.3	0.3	0.35	0.3

Tabla 5.2. Coeficientes de tema.

Nivel2: Grado de Pertenencia a cada Apartado

Del Tema 1:

Palabra	Ap. 1	Ap. 2	Ap. 3	Ap. 4
centro	0.8	0	0	0
sevilla	0.3	0.3	0.3	0.3
universidad	0.4	0.4	0.4	0.4

Tabla 5.3. Coeficientes de apartado (I).

Del Tema 2:

Palabra	Ap.1	Ap.2	Ap.3	Ap.4	Ap.5	Ap.6	Ap.7
centro	0	0	0	0	0	0.5	0
sevilla	0.2	0.2	0.2	0	0.2	0.25	0.2
universidad	0.4	0.2	0.3	0.3	0.4	0.3	0.2

Tabla 5.4. Coeficientes de apartado (II).

Del Tema 5:

Palabra	Ap.1	Ap.2	Ap.3	Ap.4	Ap.5	Ap.6	Ap.7	Ap.8	Ap.9	Ap.10
centro	0	0.2	0	0	0	0	0.7	0	0	0
sevilla	0.1	0	0	0.1	0	0	0	0	0	0
universidad	0.3	0.3	0	0.4	0.2	0.3	0	0.3	0.2	0

Tabla 5.5. Coeficientes de apartado (III).

Nivel 3: Grado de pertenencia (identificación) a cada Pregunta

Del Apartado1 del Tema1 (6 Preguntas):

Palabra	P.1	P.2	P.3	P.4	P.5	P.6
centro	0.9	0.5	0	0	0	0
sevilla	0.2	0.1	0.2	0.1	0.2	0.1
universidad	0.35	0.35	0.35	0.35	0.35	0.35

Tabla 5.6. Coeficientes de pregunta (I).

Del Apartado 2 del Tema1 (3 Preguntas y no aparece centro):

Palabra	P.1	P.2	P.3
sevilla	0.25	0.25	0.25
universidad	0.35	0.35	0.35

Tabla 5.7. Coeficientes de pregunta (II).

Del Apartado 6 del Tema2 (10 Preguntas):

Palabra	P.1	P.2	P.3	P.4	P.5	P.6	P.7	P.8	P.9	P.10
centro	0	0	0	0	0	0	0	0	0.6	0
Sevilla	0	0.25	0.25	0	0.25	0	0	0	0.25	0.55
Universidad	0	0.35	0.35	0	0.4	0	0	0	0.4	0.4

Tabla 5.8. Coeficientes de pregunta (III).

Por lo tanto, todas las palabras clave presentes en el conjunto de conocimiento deben ser almacenadas en una base de datos junto con sus vectores de peso, como se puede ver en las figuras 5.4 y 5.5.

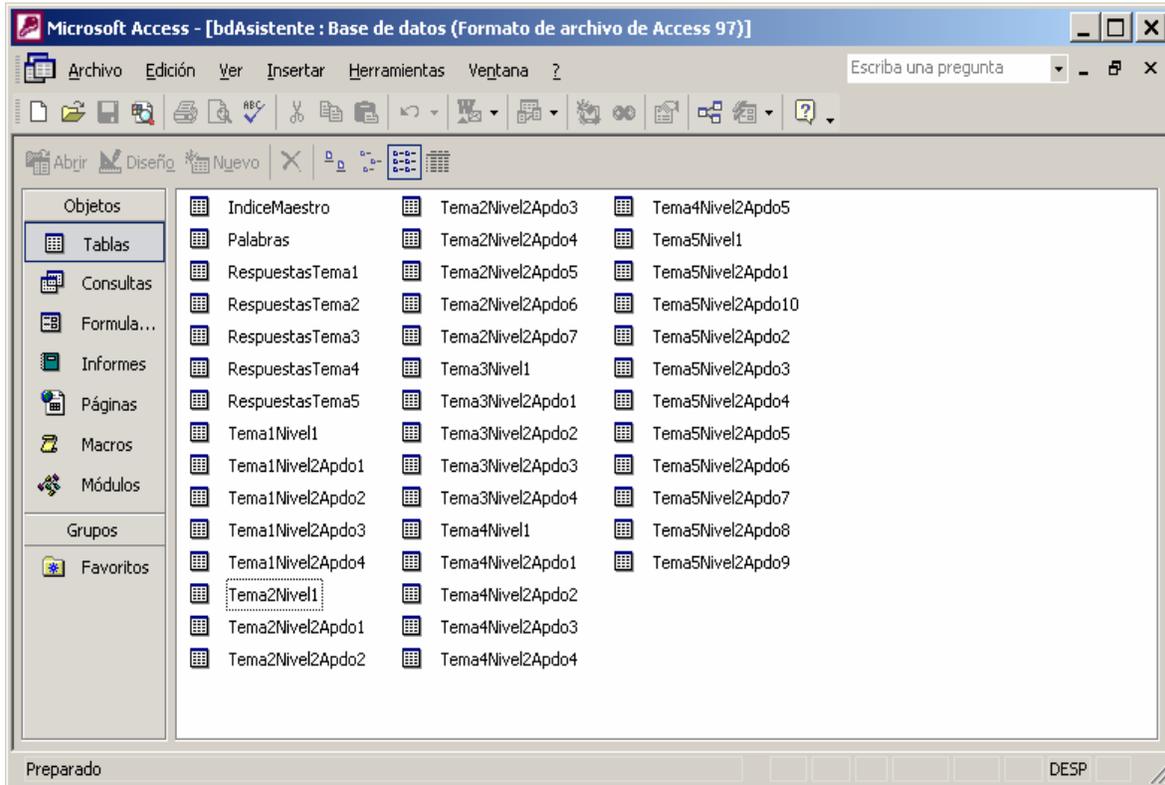


Fig 5.4. Base de conocimiento del Asistente.

The screenshot shows a Microsoft Access window titled 'Microsoft Access - [Tema2Nivel1 : Tabla]'. The window displays a data table with 8 columns labeled 'Campo1' through 'Campo8'. The data is as follows:

	Campo1	Campo2	Campo3	Campo4	Campo5	Campo6	Campo7	Campo8
becas		0	0	0,8	0	0,8	0	0
calendario		0	0	0	0	0	0,7	0,7
cadus		0	0	0	0	0	0,9	0
carnet		0	0	0	0	0	0,7	0
centro		0	0	0	0	0	0,5	0
centros		0	0	0	0	0	0,5	0
ciclo		0,5	0	0	0	0,4	0	0
ciencia		0	0	0,4	0	0	0	0
colaboracion		0	0	0,3	0	0	0	0
comunidad		0	0	0	0	0	0,4	0
configuracion		0,7	0	0	0	0	0	0
consejo		0	0	0	0	0	0,5	0
corte		0,6	0	0	0	0	0	0
cursos		0	0,5	0	0	0,5	0,5	0
departamento		0	0	0	0	0	0,4	0
departamentos		0	0	0	0	0	0,4	0
doctorado		0	0	0	0	0,8	0	0
educacion		0	0	0,4	0	0	0	0
empleo		0	0	0	0,8	0	0	0

The interface includes a menu bar (Archivo, Edición, Ver, Insertar, Formato, Registros, Herramientas, Ventana), a search bar, and a status bar at the bottom showing 'Registro: 14 de 94' and 'Vista Hoja de datos'.

Fig 5.5. Base de datos de palabras y coeficientes.

4.3. Modificación de vectores de peso.

Lógicamente y, a pesar de que el conocimiento del administrador del sistema sobre el conjunto de conocimiento sea muy bueno, habrá que modificar los vectores de peso con el fin de que se consigan mejores resultados. Aunque ya se hablará de los criterios por los que unos resultados se han considerado buenos o malos, digamos como adelanto que los coeficientes se han modificado de tal forma que todas las preguntas-tipo introducidas en las pruebas sean detectadas con un nivel de certeza mayor o igual a 0.7.

Habida cuenta de que la herramienta permite introducir de forma manual valores numéricos al motor difuso escogido y comprobar el valor devuelto, se pueden comprobar de forma manual situaciones concretas.

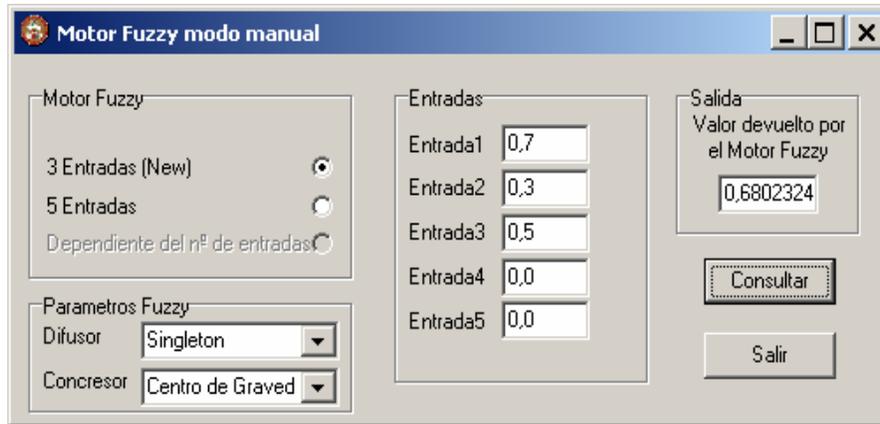


Fig. 5.5. Motor borroso manual.

4.4. Entradas al sistema de lógica borrosa

Como se ha visto en los apartados anteriores, la herramienta debe ocuparse de seleccionar las palabras clave introducidas por el usuario en su consulta en lenguaje natural. De entre las seleccionadas elige aquellas N palabras (más tarde analizaremos el número óptimo de estas) con mayor índice de significación (mayor grado de concreción) y usa sus coeficientes de pertenencia a los diferentes conjuntos difusos (empezando en nuestro caso por los de Nivel 1: Tema) como entradas del motor de inferencia.

Sin entrar de momento en la forma de estos conjuntos borrosos ni en las reglas que gobiernan el motor de lógica borrosa, de los cuales hablaremos más adelante, es necesario determinar cual es el número de entradas idóneo para el sistema de lógica borrosa.

Para ello, se realizaron pruebas sobre el motor de inferencia, utilizando las preguntas-tipo del conjunto de conocimiento. Los resultados de la prueba de reconocimiento de preguntas-tipo se han agrupado en cinco categorías, a saber:

- 1.- La pregunta correcta es la única identificada o la que tiene mayor certeza.
- 2.- La pregunta correcta está entre las 2 con mayor certeza o es la que tiene la segunda mayor certeza.
- 3.- La pregunta correcta está entre las 3 con mayor certeza o es la que tiene la tercera mayor certeza.
- 4.- La pregunta correcta está entre las 4 o más con mayor certeza o es la que tiene la cuarta mayor certeza o menos (aparece en cuarto lugar o posterior).

5.- La pregunta correcta no aparece entre las identificadas.

El usuario puede incluir de una a cinco palabras clave en su consulta. Definir un motor con pocas entradas provoca la rápida saturación del sistema. Este es un gran inconveniente para la precisión: el 90 % de las respuestas correctas son detectadas pero solo la mitad de ellas lo hacen como primera opción, tal y como se puede ver en la figura 5.6, en la que se representan los resultados para un motor de 3 entradas.

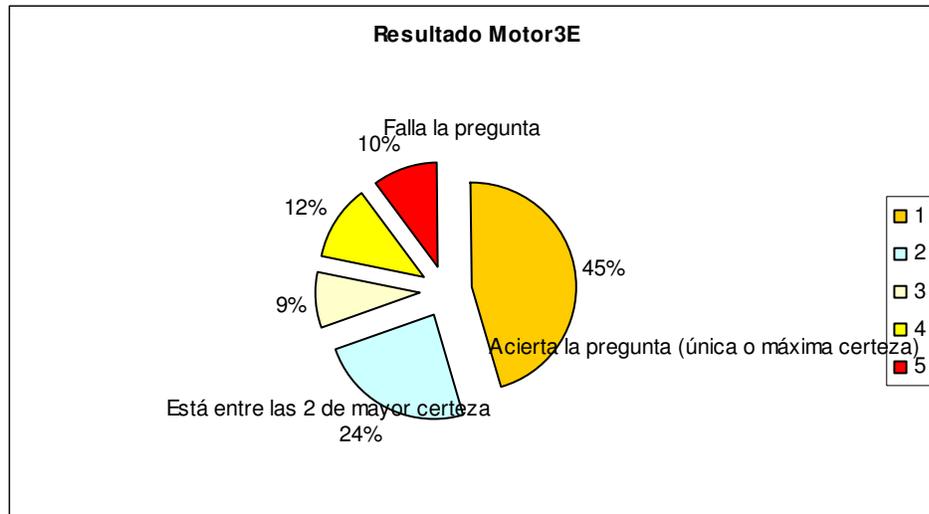


Fig. 5.6. Resultados para la herramienta de Windows con un motor de inferencia de tres entradas.

Si se define un motor de inferencia con cinco entradas, esto produce valores muy bajos en el grado de certeza, como es lógico por otra parte. La precisión sube al 55 %, pero la memoria baja, como podemos ver en la figura 5.7.

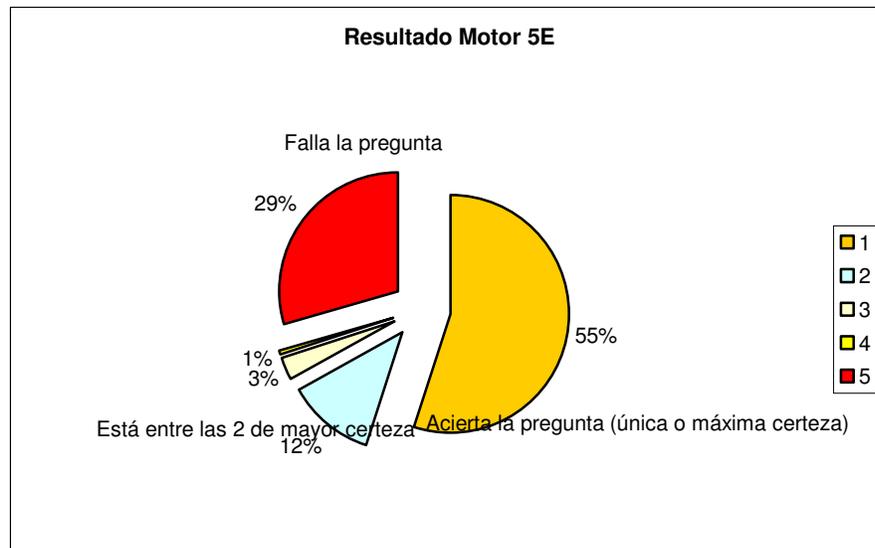


Fig. 5.7. Resultados para la herramienta de Windows con un motor de inferencia de cinco entradas.

Por lo tanto, concluimos que un número pequeño de entradas influye negativamente en la precisión mientras que un gran número de entradas lo hace en la memoria. No obstante, se pueden efectuar mejoras a este respecto, basándose en utilizar un número de entradas variables. Este aspecto será comentado en un apartado posterior.

4.5. Motor de inferencia.

El elemento del sistema que determina el grado de certeza acerca de que unas determinadas palabras clave correspondan a cada uno de los posibles temas, apartados y/o preguntas es el motor de inferencia.

El motor de inferencia implementado necesita unas entradas y un conjunto de reglas y devuelve unas salidas. Debido a que la información de entrada con la que se alimenta al motor es imprecisa, es interesante el uso de lógica difusa para su tratamiento, como se explicó en capítulos anteriores. Al tratarse de lógica difusa, tanto los conjuntos difusos de las entradas y salidas como las reglas, pueden ser expresados y manejados mediante valores lingüísticos (poco, bastante, mucho, alto, medio, bajo, etc.).

4.5.1. Desarrollo del motor de inferencia.

Como se indica al principio de este apartado, el motor de inferencia se desarrolla a partir del código C generado por el programa Un-Fuzzy. Es posible construir en Un-Fuzzy un motor con los mismos parámetros que un motor que fuera desarrollado en Matlab. Se comprobó que el funcionamiento de un motor implementado en Un-Fuzzy es igual que el implementado en Matlab a pesar de que los resultados numéricos no son exactamente iguales. Esto es debido a que, mientras que en Un-Fuzzy es posible definir diferentes congresores y difusores, en Matlab no se tiene acceso a esa información por lo que es imposible generar un motor exactamente igual.

El código generado por Un-Fuzzy es utilizable después de realizar unas mínimas modificaciones, aunque la definición de las reglas de inferencia es larga y tediosa.

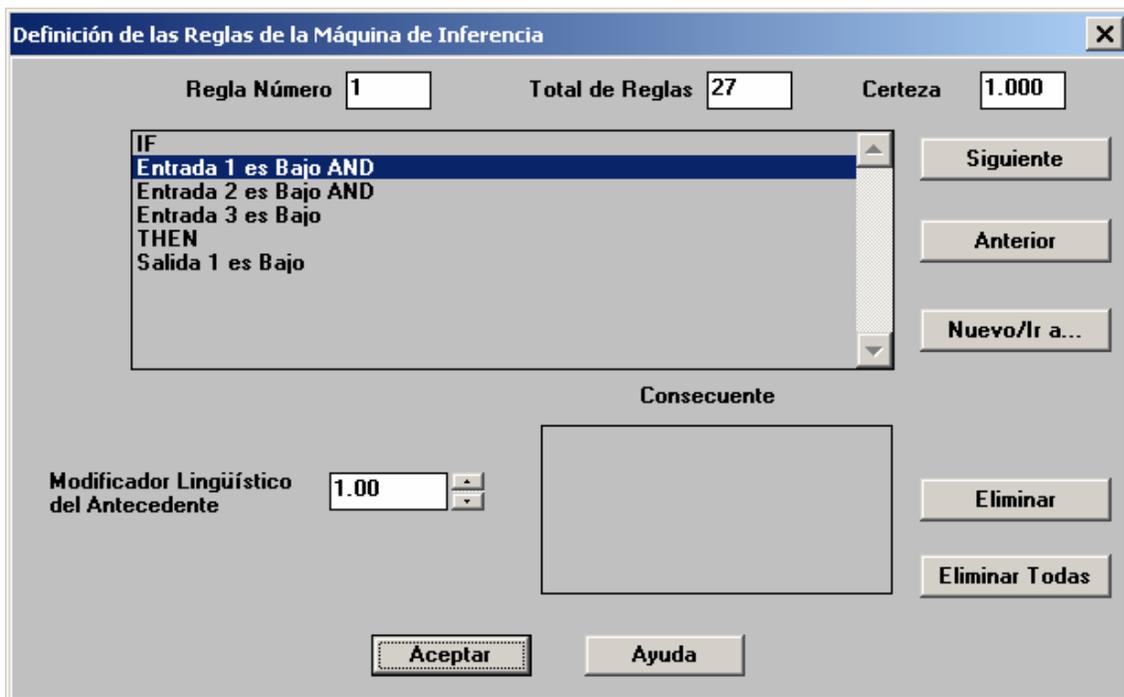


Fig. 5.8. Definición de reglas con Un-Fuzzy.

El hecho de modificar el número de conjuntos de las entradas o salidas obliga a redefinir las reglas. Por ejemplo, un motor de tres entradas con tres conjuntos borrosos cada una y una salida con cuatro conjuntos borrosos tiene 27 reglas pero un motor de cinco entradas con tres conjuntos borrosos y una salida con cuatro conjuntos borrosos tiene 243

reglas. Se puede observar que es apreciable el esfuerzo que supone realizar una modificación en el motor de inferencia.

En definitiva, insertar el código generado por Un-Fuzzy en la aplicación desarrollada es prácticamente inmediato si:

- a.- Sólo se maneja un motor difuso
- b.- La configuración de dicho motor es rígida

En caso contrario (más de un motor difuso), es necesario renombrar las funciones de cada motor para que no exista conflicto entre funciones diferentes con igual nombre. Si, además se desea permitir la modificación de la configuración del motor de forma opcional (p.e. cambiar el tipo de conector) es necesario realizar tareas de ingeniería inversa para interpretar el código generado por Un-Fuzzy y sus opciones.

No obstante, y dado que se utilizará el mismo motor para todos los niveles del conjunto del conocimiento, las ventajas que ofrece Un-Fuzzy son superiores a sus inconvenientes, por lo que fue el programa utilizado finalmente para el desarrollo de la herramienta.

4.5.2. Definición de los conjuntos borrosos.

Las entradas del motor de inferencia están incluidas en un universo lingüístico que toma los valores Bajo, Medio, Alto. La salida, que indica el grado de certeza, tiene un universo lingüístico que puede tomar los valores Bajo, Medio-Bajo, Medio-Alto, o Alto. Estos valores corresponden a los conjuntos borrosos de entrada y salida.

El hecho de que las entradas tomen estos valores es simplemente debido al sentido común (lo que a fin de cuentas tiene mucho que ver con la lógica borrosa). Son el número suficiente como para que los resultados puedan ser coherentes y no son tantas opciones como para que el número de reglas aumente de manera alarmante. De hecho, las salidas también se definieron así en un principio. El haber introducido un posible valor más en las salidas se debió a una mejora considerable de los resultados con esta modificación.

Las entradas numéricas, que corresponden a los valores contenidos en los vectores de pesos, se encuentran por tanto acotadas entre 0 y 1. Si la entrada es 0, es evidente que el valor es Bajo con un grado de certeza 1, mientras que si la entrada es 1, el valor es Alto con certeza 1. Los valores de entrada intermedios toman valores entre 0 y 1 para los conjuntos borrosos Bajo, Medio y Alto, tal y como se puede ver en la figura 5.9.

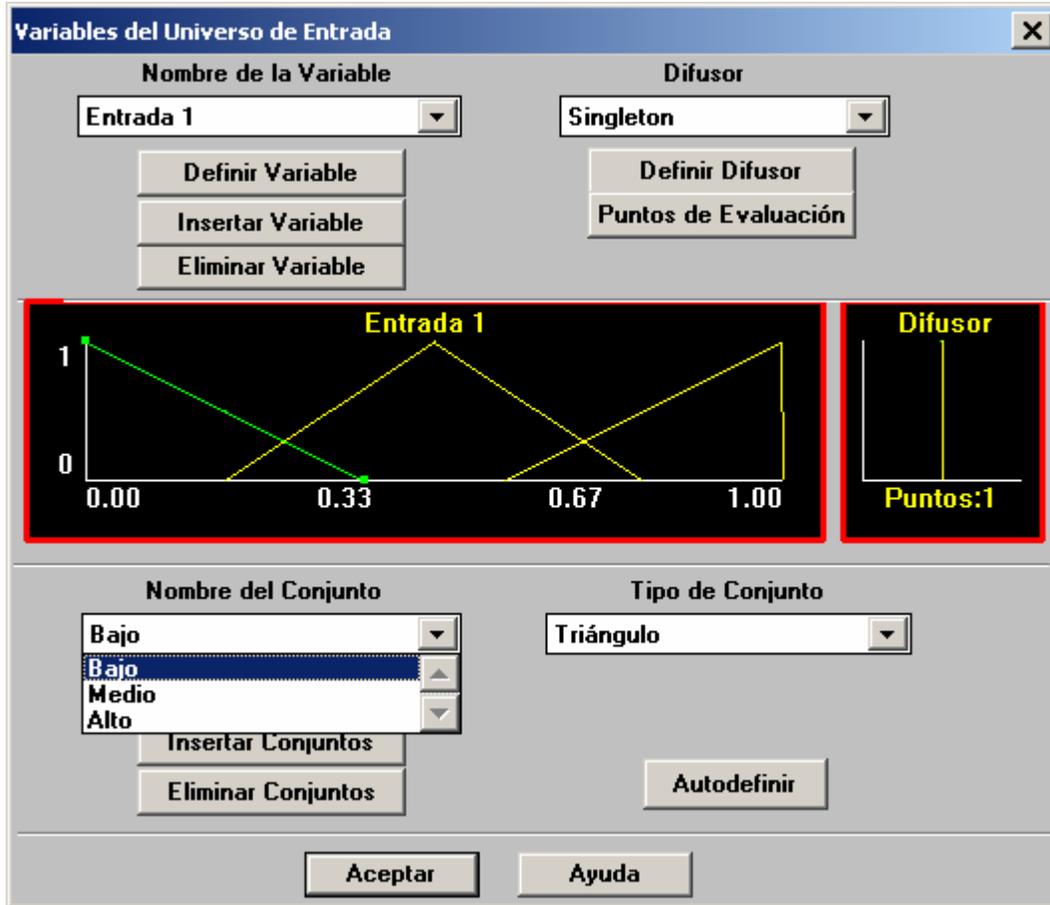


Fig. 5.9. Forma de los conjuntos borrosos de entrada.

La forma de los conjuntos es triangular, por una parte por simplicidad y por otra parte porque probados otros tipos de conjuntos más complejos de definir (Campana, tipo Pi, etc), los resultados no experimentaron ninguna mejoría.

En cuanto a los conjuntos borrosos de salida, se definen de forma análoga, aunque con la diferencia de que existe un conjunto borroso más. Esto puede verse en la figura 5.10.

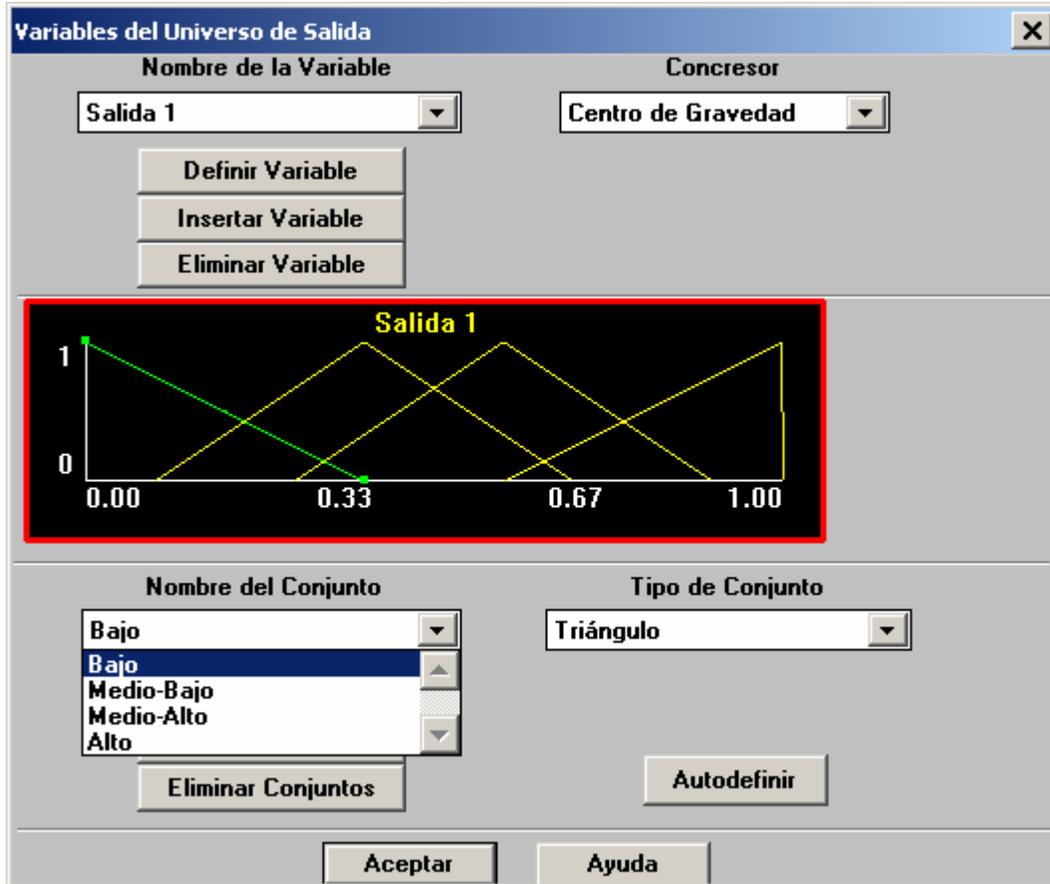


Fig. 5.10. Forma de los conjuntos borrosos de salida.

4.5.3. Definición de las reglas.

Tal y como se indicaba en el capítulo anterior, la definición de reglas está basada en el estudio del conjunto del conocimiento por parte del administrador del sistema y en las pruebas que pueda realizar sobre este. Tras estudiar dicho conjunto de conocimiento y realizar pruebas con el sistema, las reglas de inferencia que se definieron fueron las siguientes:

1. Para el motor de 3 entradas.

Las reglas definidas para la inferencia son:

- a.- Con todo Bajo la salida es Bajo
- b.- Con 1 Medio, la salida es Medio-Bajo

- c.- Con 2 Medio, la salida es Medio-Alto
- d.- Con 3 Medio ó 1 Alto, la salida es Alto

El desarrollo de esto da lugar a:

- Si todas las entradas son Bajo, la salida es Bajo
- Si una de las entradas es Medio y las demás Bajo, la salida es Medio-Bajo
- Si dos de las entradas es Medio y las demás Bajo, la salida es Medio-Alto
- Si tres entradas o más son Medio y las demás Bajo, la salida es Alto
- Si 1 entrada o más es Alto, la salida es Alto

Y sus posibles combinaciones generan 27 reglas para el motor difuso.

2. Para el motor de 3 entradas.

Las reglas definidas para la inferencia son:

- a.- Con 1 Medio o ninguno la salida es Bajo
- b.- Con 1 Alto ó 2 Medio, la salida es Medio-Bajo
- c.- Con 1 Alto y 1 Medio (3/2 Alto) o equivalente, la salida es Medio-Alto
- d.- Con 2 entradas Alto o equivalente, la salida es Alto
- e.- Con 3 Altos o equivalente la salida es Muy Alto

Nótese que se igualan 2 Medios a 1 Alto.

El desarrollo de esto da lugar a:

- Si todas las entradas son Bajo, la salida es Bajo
- Si una de las entradas es Medio y las demás Bajo, la salida es Bajo
- Si una de las entradas es Alto y las demás Bajo, la salida es Medio-Bajo
- Si dos de las entradas es Medio y las demás Bajo, la salida es Medio-Bajo
- Si una de las entradas es Alto, otra Medio y las demás Bajo, la salida es Medio-Alto
- Si tres de las entradas es Medio y las demás Bajo, la salida es Medio-Alto
- Si cuatro de las entradas es Medio y la otra Bajo, la salida es Alto
- Si tres de las entradas es Medio, otra Alto y la otra Bajo, la salida es Alto
- Si tres de las entradas es Medio, y las otras Alto, la salida es Alto
- Si dos entradas son Medio, otra Alto y las demás Bajo, la salida es Alto
- Si dos entradas son Medio, dos Alto y la otra Bajo, la salida es Alto
- Si dos de las entradas es Alto y las demás Bajo, la salida es Alto
- Si dos de las entradas es Alto, una Medio, y las demás Bajo, la salida es Alto
- Si todas las entradas son Medio, la salida es Alto
- Si cuatro de las entradas es Medio y la otra Alto, la salida es Muy Alto
- Si tres de las entradas es Alto y las demás Bajo, la salida es Muy Alto
- Si tres de las entradas es Alto, otra Medio y la otra Bajo, la salida es Muy Alto

Si cuatro de las entradas es Alto y la otra Bajo, la salida es Muy Alto
Si cuatro de las entradas es Alto y la otra Medio, la salida es Muy Alto
Si todas las entradas son Alto, la salida es Muy Alto

Sus posibles combinaciones generan 243 reglas para el motor difuso. Vemos lo tediosa que resulta la definición de todas estas reglas en Un-Fuzzy a pesar de su aparente facilidad (se han resumido todas las reglas en un principio solo en 5 puntos, pero hay que definir 243 reglas diferentes).

Nótese además que el motor desarrollado sólo tiene 4 conjuntos de salida por lo que, en la codificación de sus reglas, las que deberían devolver una salida Muy Alto devuelven una salida Alto.

4.5.4. Salidas del sistema de lógica borrosa.

Dadas las entradas, los conjuntos borrosos y las reglas definidas anteriormente, el motor de inferencia debe encargarse de calcular la salida de cada sistema borroso. Existe un sistema de lógica borrosa por cada nivel, como ya se indicó en el apartado 4 del capítulo 4.

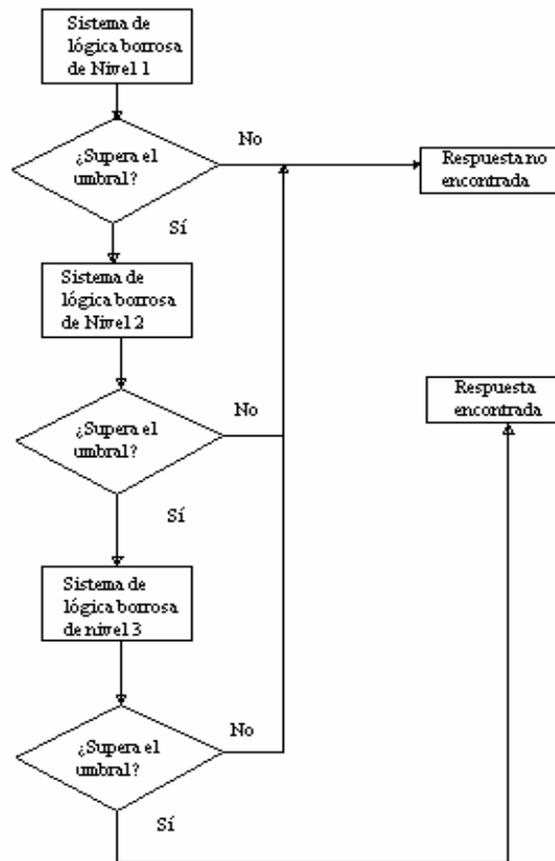


Fig. 5.11. Modo de operación del sistema de lógica borrosa.

En este caso, el nivel 1 corresponde a cada uno de los temas del conjunto del conocimiento. Las entradas son los vectores de peso correspondientes a las palabras seleccionadas para cada uno de los temas. Si la salida supera un cierto umbral fijado, se consideran los vectores de peso correspondientes al nivel 2 (en este caso, cada uno de los apartados), pero solo se consideran los apartados de los temas cuya salida supere el umbral. Este proceso se repite para el nivel 3 (Pregunta).

En las pruebas se han considerado umbrales de 0.5 para todos los niveles, aunque estos umbrales pueden ser modificados con el fin de obtener mejores resultados.

En el ejemplo que se ha venido considerando, la consulta de usuario “¿Cómo puedo localizar información sobre un Centro de la Universidad de Sevilla?”, ocurriría lo siguiente:

Las palabras clave identificadas son:

Palabra1: localizar; Concreción: 0,3
 Palabra2: informacion; Concreción: 0,3
 Palabra3: centro; Concreción: 0,7
 Palabra4: universidad; Concreción: 0,3
 Palabra5: sevilla; Concreción: 0,3

Las palabras seleccionadas para evaluar los temas son las cinco con mayor grado de concreción (en este caso, las cinco identificadas). Los vectores de peso correspondientes a cada una de estas palabras para cada tema son las entradas al primer motor de lógica borrosa.

Palabra	Coefficiente Tema 1	Coefficiente Tema 2	Coefficiente Tema 3	Coefficiente Tema 4	Coefficiente Tema 5
Localizar	0,1	0	0,1	0,1	0,5
Información	0,3	0,1	0,1	0,5	0,1
Sevilla	0,2	0,5	0	0,25	0,2
Centro	0,8	0,5	0	0,1	0,5
Universidad	0,4	0,3	0,3	0,35	0,3

Tabla 5.9. Coeficientes de Tema. Ejemplo.

Tras atravesar dicho motor, los índices de certeza de que la consulta pertenezca a cada tema son:

Índice de Certeza del Tema 1: 0,680232465267181
 Índice de Certeza del Tema 2: 0,517307579517364
 Índice de Certeza del Tema 3: 0,155263155698776
 Índice de Certeza del Tema 4: 0,425446450710297
 Índice de Certeza del Tema 5: 0,517307579517364

El umbral de Certeza para aceptar el tema como posible es 0,5, por lo que toda la información correspondiente a los Tema 3 y 4 queda descartada.

Para cada tema que ha superado el umbral, se tomarán los vectores de peso correspondientes a cada apartado como nuevas entradas para el motor de lógica difusa. Por ejemplo, para el Tema 1:

Las palabras seleccionadas para evaluar los Apartados del Tema 1 son:

Palabra 1: informacion
 Palabra 2: centro
 Palabra 3: universidad

Palabra 4: sevilla

La certeza del Apartado1 es de: 0,819387793540955

La certeza del Apartado2 es de: 0,331999927759171

La certeza del Apartado3 es de: 0,331999927759171

La certeza del Apartado4 es de: 0,331999927759171

Dado que el umbral vuelve a ser de 0.5, la información correspondiente a los Apartados 1, 2 y 3 del tema 1 queda descartada. Este proceso se realiza también con el resto de temas que habían pasado el primer umbral. Lo mismo ocurrirá con el tercer nivel, correspondiente a las preguntas. En el caso del Tema 1, Apartado 1, ocurre lo siguiente:

La pregunta nº 1 del Apartado 1 del Tema 1 ha sido identificada como probable

La pregunta tipo identificada es:

"¿Cómo puedo localizar información sobre un Centro de la Universidad de Sevilla?"

El grado de certeza obtenido es de: 0,713486790657043

Esta pregunta es la que habíamos introducido en la consulta de usuario. Además, también se han recogido otras preguntas de otros Temas y Apartados relacionadas de alguna forma con la consulta. Son las siguientes:

La pregunta nº 9 del Apartado 6 del Tema 2 ha sido identificada como probable

La pregunta tipo identificada es:

"¿Cómo puedo obtener información sobre Centros, Departamentos y Servicios de la Universidad de Sevilla?"

El grado de certeza obtenido es de: 0,62736839056015

La pregunta nº 10 del Apartado 6 del Tema 2 ha sido identificada como probable

La pregunta tipo identificada es:

"¿Existe alguna página informativa sobre Sevilla en el portal de la Universidad?"

El grado de certeza obtenido es de: 0,5625

Estas son las preguntas que han superado el umbral y que están relacionadas en mayor o menor medida con la consulta de usuario. Las tres respuestas correspondientes a las preguntas con mayor grado de certeza serían presentadas al usuario ordenadas en base a dicho grado de certeza. La información final quedaría de esta manera:

La pregunta nº 1 del Apartado 1 del Tema 1 ha sido identificada como probable

El grado de certeza obtenido es de: 0,713486790657043

La pregunta tipo identificada es:

"¿Cómo puedo localizar información sobre un Centro de la Universidad de Sevilla?"

La respuesta correspondiente es:

"Puede localizar información sobre los Centros Universitarios a través del portal WEB de la Universidad de Sevilla <http://www.us.es> siguiendo las opciones de menú desplegable:

Centros y departamentos -> Centros

La Universidad de Sevilla dispone además de un directorio en el que usted puede localizar información de los distintos Centros de la Universidad. Opciones de menú desplegable:

Presentación y Organización->Directorio

Opción: Directorio Institucional (Alfabético)"

La pregunta nº 9 del Apartado 6 del Tema 2 ha sido identificada como probable

El grado de certeza obtenido es de: 0,62736839056015

La pregunta tipo identificada es:

"¿Cómo puedo obtener información sobre Centros, Departamentos y Servicios de la Universidad de Sevilla?"

La respuesta correspondiente es:

"La Universidad de Sevilla dispone de un directorio, donde usted puede localizar los distintos Centros, Departamentos y Servicios de la Universidad.

Puede acceder al Directorio a través de nuestra página WEB: www.us.es, siguiendo las opciones de menú desplegable:

Presentación y Organización -> Directorio -> Directorio Institucional (Alfabético)"

La pregunta nº 10 del Apartado 6 del Tema 2 ha sido identificada como probable

El grado de certeza obtenido es de: 0,5625

La pregunta tipo identificada es:

"¿Existe alguna página informativa sobre Sevilla en el portal de la Universidad?"

La respuesta correspondiente es:

**"Puede obtener información sobre Sevilla, a través de nuestra página WEB:
www.us.es, siguiendo las opciones de menú desplegable:
Universidad virtual -> Enlaces de interés**

Consulte en el apartado "Enlaces sobre Sevilla".

4.5.5. Modificaciones al sistema.

En el apartado 4.4 de este capítulo referente a las entradas del sistema, se había concluido que un número pequeño de entradas influye negativamente en la precisión mientras que un gran número de entradas lo hace en la memoria. Sin embargo, del análisis de los resultados fallidos se observó que en buena parte de ellos no se obtenía respuesta porque la certeza fijada como umbral es mayor que la mínima obtenida. Esto plantea bajar los umbrales de certeza para aceptar el resultado como correcto. No obstante, al aplicar esta modificación se dan por válidas muchas respuestas erróneas estropeando los resultados anteriores.

La solución propuesta es modificar el procedimiento de manera que, sólo en el caso de que ningún resultado supere el umbral fijado, el Asistente baje automáticamente el umbral de aceptación del resultado. Aplicando este método, los resultados mejoran notablemente. En la figura 5.12 se observan dichos resultados.

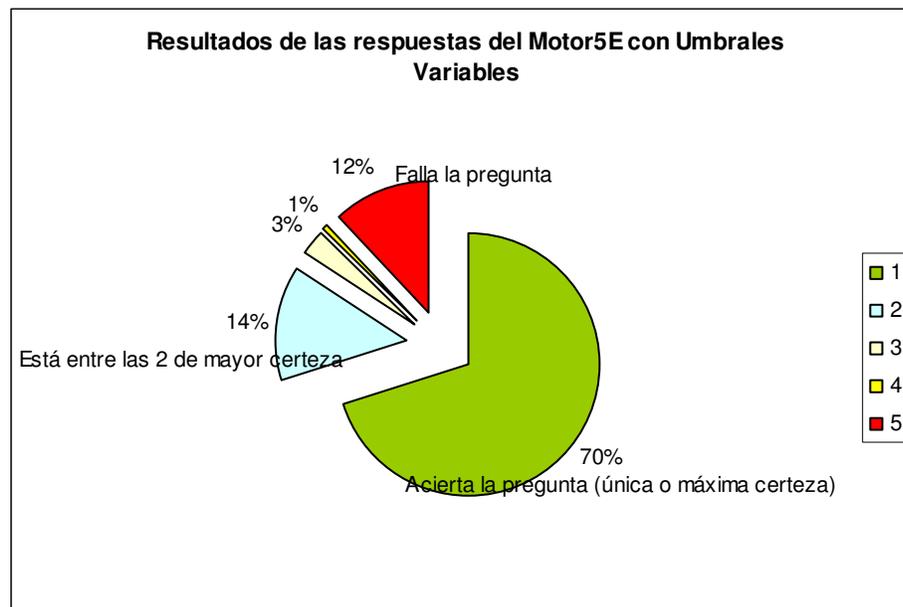


Fig. 5.12. Resultados para la herramienta de Windows con un motor de inferencia de cinco entradas y umbrales variables.

Viendo como funciona la modificación anterior, cabe la tentación de subir mucho los umbrales (incluso partir del 100%) de los índices de certeza para filtrar aún mejor los resultados. Observando los informes se aprecia claramente que no siempre el resultado correcto es el que tiene mayor certeza (recordemos que se trata de aplicar lógica difusa) por lo que si se suben mucho los índices es posible que no se llegue a la respuesta correcta. Además, recordemos que al agente no se le formulará una consulta tipo sino una consulta parecida por lo que elevar mucho el umbral de exigencia podría inducir a repuestas erróneas.

Por otra parte, hasta este momento se había visto que fijar los umbrales a 0.5 es un compromiso que proporcionaba buenos resultados (la respuesta se encontraba entre las tres primeras) y permite menos rigor en la adjudicación de los coeficientes de pertenencia a cada nivel. Es tarea del administrador el ajustar los vectores de pesos y los valores de los umbrales para conseguir unos valores óptimos.

En resumen, y, en base a las pruebas realizadas, se concluye que si se ofrecen al usuario los tres posibles resultados con mayor índice de certeza, la respuesta correcta aparece el 87% de las veces, y es la primera el 70% de las veces, lo que, a todas luces, constituye un buen resultado.

En cuanto al número de entradas y, teniendo en cuenta que hay ocasiones en las que es mejor utilizar el motor difuso de tres entradas, mientras que en otras ocasiones es mejor utilizar el de cinco entradas, se llega a un compromiso por el que se utilizará un motor variable dependiente del número de palabras clave identificadas en la consulta. En el caso de que se identifiquen entre una y tres palabras clave, se usa el motor de tres entradas, mientras que si se identifican más de cinco palabras, es el motor de cinco entradas el utilizado. Los resultados se pueden observar en la Tabla 5.10 y la Figura 5.13.

Categoría	1	2	3	4	5
%	77.45	15.79	4.51	0.75	1.50

Tabla 5.10. Resultados para un número de entradas variable (I).

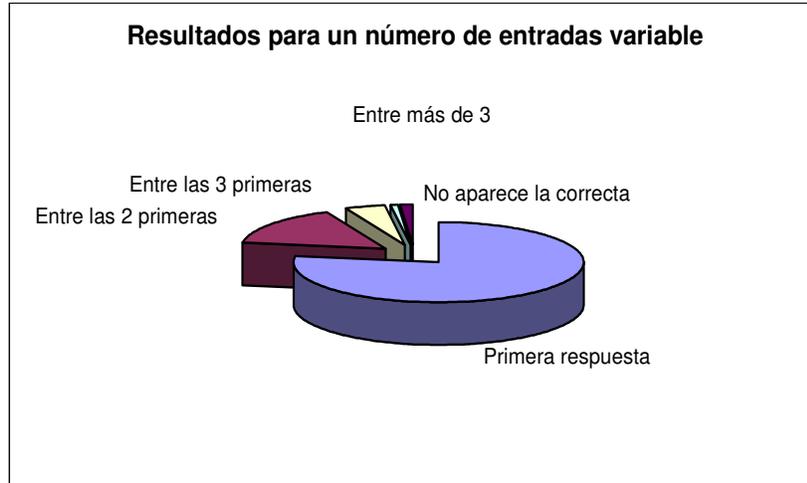


Fig. 5.13. Resultados para un número de entradas variable (II).

Se puede ver que ofreciendo las tres respuestas identificadas como probables con mayor certeza el usuario obtiene la correcta el 97.75% de las veces y el 77.45% de las veces es la primera opción.

CAPÍTULO 6. CONCLUSIONES.

1. Conclusiones.

El Departamento de Tecnología Electrónica de la Universidad de Sevilla y, en concreto, el grupo de investigación “*Tecnología e Informática e Industrial*” lleva años desarrollando técnicas de Inteligencia Artificial y Sistemas Expertos.

Este trabajo de investigación se enmarca dentro de esta línea y, en particular, dentro de una de las ramas de la Inteligencia Artificial: la lógica borrosa. A pesar de que el grupo de investigación ya ha trabajado ampliamente con la otra rama principal de la Inteligencia Artificial, las redes neuronales, este trabajo nos da la oportunidad también de utilizar la lógica borrosa como una potente herramienta para el diseño de un agente inteligente para la clasificación de contenidos, que se describe en este trabajo. Hasta el momento, se ha desarrollado una herramienta basada en la lógica borrosa para localizar contenidos en una base de datos estructurada jerárquicamente por niveles mediante consultas en lenguaje natural que arroja unos buenos resultados.

Además, el trabajo ya ha dado pie a la aceptación de un artículo en la 13ª Conferencia Electrotécnica Mediterránea, organizada por el IEEE en Benalmádena (Málaga) y que tuvo lugar en Mayo de 2006. Este artículo se titula *A fuzzy logic system for classifying the contents of a database and searching consultations in natural language* (Un sistema de lógica borrosa para la clasificación de los contenidos de una base de datos y la búsqueda de consultas) y se incluye en el apéndice C.

2. Líneas a seguir.

Varias líneas de investigación se abren para continuar la senda abierta por este trabajo.

Por un lado, debe continuar el análisis de los parámetros del sistema de lógica borrosa, así como las posibles modificaciones a introducir ante un eventual aumento del tamaño del conjunto del conocimiento. Dentro de esta línea de investigación, se enmarcan las siguientes tareas a realizar:

- Modificación de los valores de los vectores de pesos para identificar correctamente las preguntas que no salen entre las posibles en las pruebas realizadas.

- Modificación de la aplicación para permitir variar los tipos de conjuntos difusos de las entradas y de la salida.
- Realización de las pruebas modificando el difusor y el congresor elegido.
- Realización de pruebas modificando los tipos de conjuntos de entrada y salida.
- Repetición de las pruebas realizadas hasta ahora con bancos de preguntas similares y con otros conjuntos de conocimiento mayores. Entre estos últimos se enmarcaría el nuevo portal web de la Universidad de Sevilla, el cual constituirá una gran prueba de fuego para el sistema.

Por otra parte, será necesario evolucionar en el procesado del lenguaje natural. Hasta ahora, se ha trabajado con vectores de pesos y asignación de coeficientes para palabras simples, pero será necesario introducir otra serie de tareas:

- Introducción de sinónimos o expresiones similares en la base de datos de palabras clave.
- Gestión del enraizado, es decir, la posibilidad de incluir de masculinos, femeninos, plurales, singulares, tiempos verbales distintos, abreviaturas de palabras, etc.
- Gestión de expresiones compuestas.
- Humanización de la interfaz: tratamiento de la parte de la consulta del usuario que no corresponda a contenidos de la web creando un modelo conversacional capaz de simular las respuestas que daría una persona. Por ejemplo, ante las frases “hola”, “buenos días” o “buenas tardes” se podría responder “Hola. Bienvenido a la web de la Universidad de Sevilla. ¿En qué puedo ayudarle?”.

La tercera vía de investigación es el estudio de la viabilidad en la utilización de una de las herramientas menos conocidas y más prometedoras de la Inteligencia Artificial, los sistemas neuro-borrosos (*neurofuzzy systems*). De hecho, se ha propuesto un proyecto fin de carrera con el Título “*Implementación de un sistema buscador de información mediante técnicas de Inteligencia Artificial*”, en el que se pretende estudiar el estado del arte en los sistemas neuro-borrosos y su posible aplicación a la búsqueda de información.

Por último, será necesario gestionar la información que no pertenezca propiamente al conjunto del conocimiento. Dentro de las tareas a realizar se encontrarían las siguientes:

- Gestión de errores: un “bunos días” puede ser entendido como un “buenos días”.

- Gestión de temas no relacionados, como el uso de palabras obscenas o simplemente no contenidos en la web.
- Acceso a otros contenidos relacionados de alguna manera con el conjunto de conocimiento, pero que no se encuentren dentro de este. Por ejemplo, en el caso del portal de la Universidad, habría que considerar las páginas web de las facultades y escuelas.

APÉNDICE A. HERRAMIENTAS DE LÓGICA BORROSA.

1. Toolbox de lógica borrosa de Matlab.

Con el fin de comprobar la viabilidad del método propuesto, se realizaron pruebas utilizando el potente toolbox de Matlab. Dicho toolbox consta de una interfaz gráfica sencilla de usar y que nos permite configurar todos los parámetros de un sistema de lógica borrosa de una manera fácil y cómoda.



Fig. A.1. Esquema del editor de sistemas de inferencia borrosos [33].

La interfaz gráfica se divide en cinco interfaces, que se describen a continuación.

1.1. El editor FIS.

La interfaz principal es el editor FIS. En este editor se definen los parámetros principales para comenzar la implementación de un sistema de lógica borrosa.

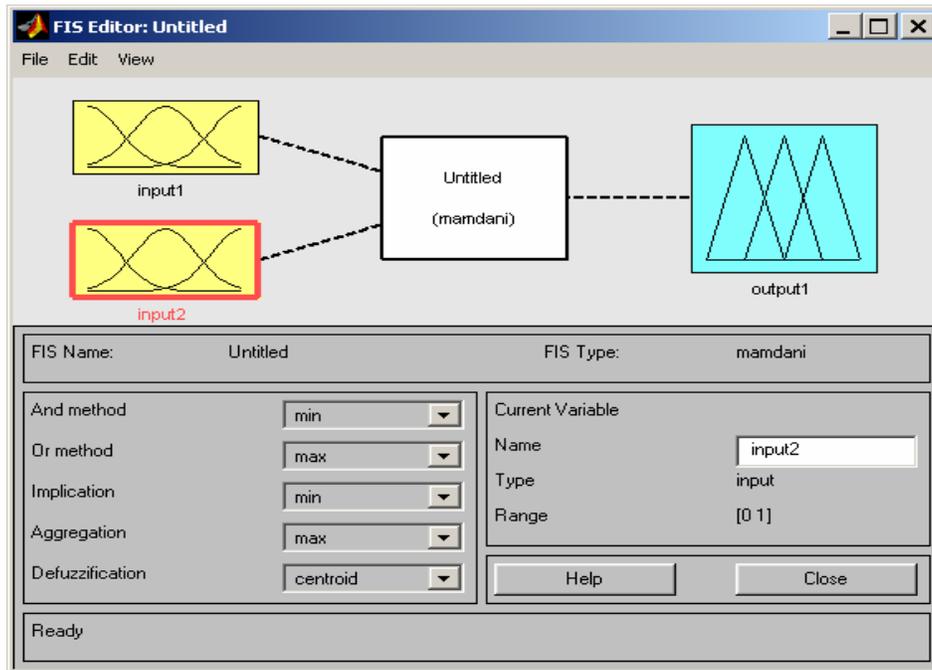


Fig. A.2. Editor FIS.

El editor FIS permite definir:

- El número y nombre de las entradas.
- El número y nombre de las salidas.
- Otros parámetros del sistema, tales como las funciones de inferencia borrosa y el método de desborrosificación.

1.2. El editor de funciones de pertenencia.

El editor de funciones de pertenencia comparte algunas características con el editor FIS. De hecho es accesible desde el editor simplemente pinchando en la entrada o salida correspondiente. Este editor permite:

- Definir el número de conjuntos lógicos para cada variable (entrada o salida).
- Definir el nombre de cada conjunto.
- Definir la función de cada conjunto lógico (triángulo, gaussiana, sigmoidea, ...).
- Definir el rango de valores de cada función.

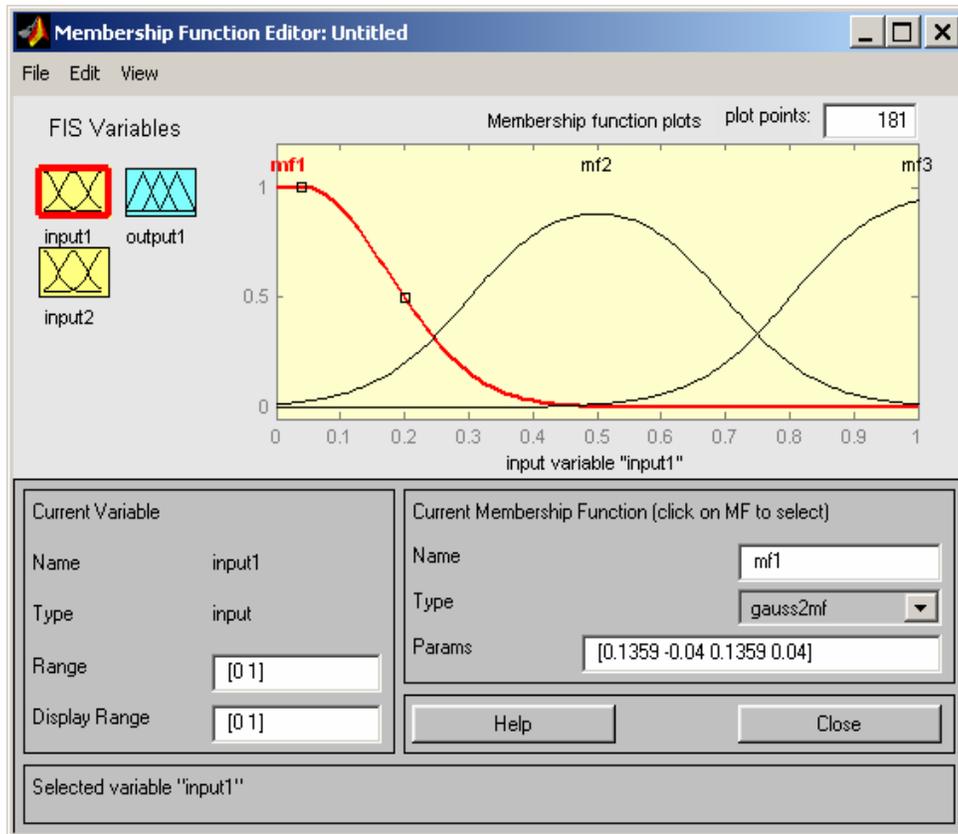


Fig. A.3. Editor de funciones de pertenencia.

1.3. El editor de reglas.

La construcción de reglas usando el editor de reglas gráfico es trivial. No hay más que ver la propia interfaz para comprender esta facilidad.

Se pueden usar operaciones AND y OR para construir reglas del tipo SI... ENTONCES. Además, existe la posibilidad de asignar pesos diferentes a cada regla.

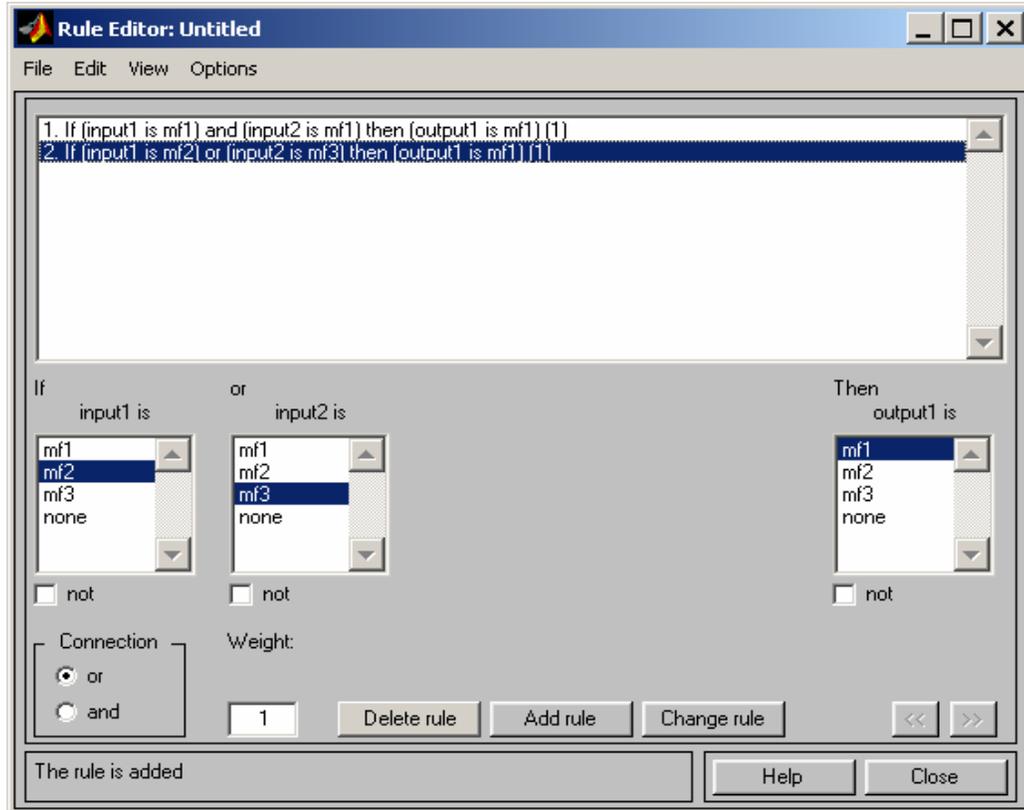


Fig. A.4. Editor de reglas.

1.4. El visor de reglas y el visor de superficies.

Estas dos interfaces son de solo lectura, con lo cual no interactuamos directamente con ellas y tienen un valor puramente informativo.

El visor de reglas despliega un “mapa” de todo el proceso de inferencia, representándose el antecedente y la consecuencia de cada regla de manera gráfica.

En cuanto al visor de superficies, este representa un diagrama en n-dimensiones en el que se relacionan las variables de entrada con las de salida, pudiendo observar las distintas posibilidades sobre el gráfico.

Dado que no vamos a actuar sobre ellas y, como se ha mencionado antes, solo tienen valor informativo, el estudio ulterior de estas interfaces es irrelevante para el objeto de este trabajo.

2. El programa Un-fuzzy.

Un-Fuzzy, actualmente en la versión 1.2, es una herramienta para el análisis, diseño, simulación e implementación de Sistemas de Lógica Difusa creado por Oscar G. Duarte para el departamento de Tecnología Eléctrica de la Universidad de Colombia, distribuyéndose gratuitamente a través de la Web de dicho departamento [38].

Es destacable de este programa la poca cantidad de recursos que necesita para funcionar, sin necesitar ni siquiera una instalación previa, y que tanto el programa como el manual y la ayuda en línea se encuentran en español, algo no muy común en este tipo de herramientas.

Las interfaces más importantes con las que cuenta Un-fuzzy se detallan en los siguientes apartados:

2.1. Ventana principal.

Esta es la ventana que aparece nada más arrancar la utilidad y desde la que podemos acceder a todas las opciones en el diseño, simulación e implementación de problemas en lógica difusa que nos da el programa.

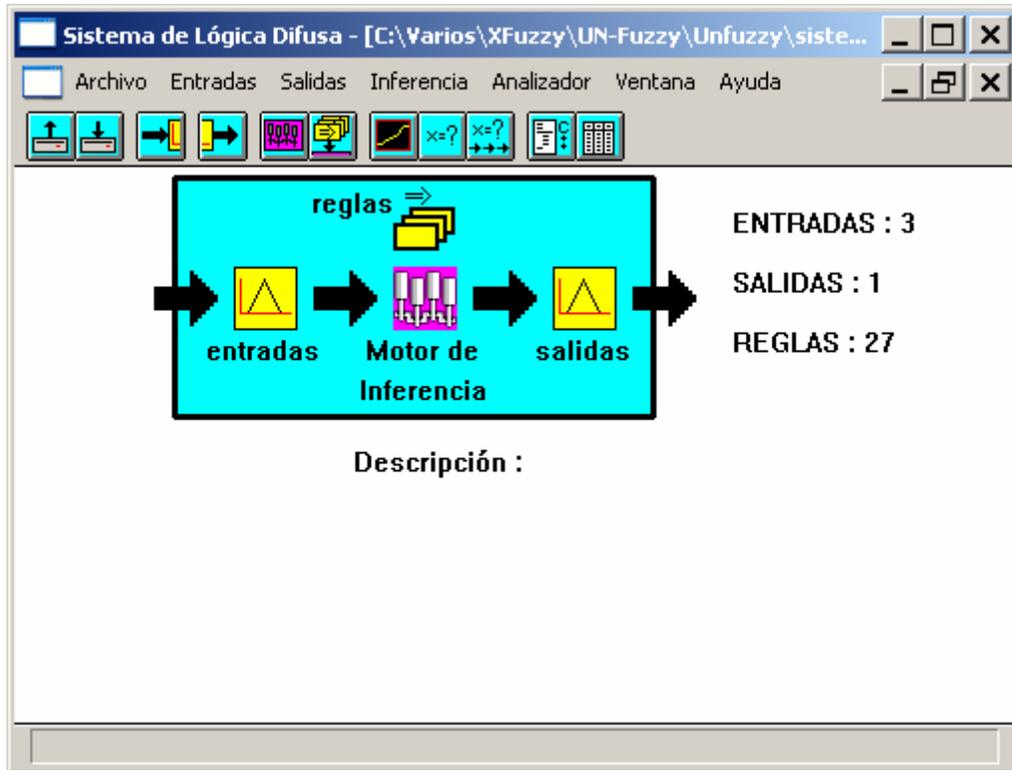


Fig. A.5. Ventana principal de Un-fuzzy.

Como puede apreciarse en la figura A.5, el programa tiene la estructura típica de cualquier programa en Windows, donde cabe destacar la barra de botones que resulta cómoda e intuitiva a la hora de utilizar el programa. El significado de estos botones aparece en la figura A.6.



Fig. A.6. Barra de botones de Un-Fuzzy

En esta ventana se hace una representación gráfica de una máquina de inferencia en la que aparecen una serie de bloques que representan al universo de entrada, al de salida, a la

base de reglas y al motor de inferencia. Si quisiéramos modificar cualquiera de estas partes, solo tendríamos que hacer doble clic sobre el bloque que lo representa.

Este programa ofrece la posibilidad de abrir un archivo, crear uno nuevo ó guardar nuestros proyectos. Todas estas opciones están accesibles desde el menú Archivo. Sólo tenemos la posibilidad de guardar a disco nuestro trabajo si tenemos definido el universo de entrada, el de salida y al menos una regla. Esto puede suponer un grave problema para aplicaciones en las que haya muchas variables de entrada y salida ya que obliga a realizar todos estos pasos en una sola sesión.

2.2. Universo de entrada.

Para la definición del Universo de entrada es escogen las variables lingüísticas del sistema, sus correspondientes conjuntos y los difusores correspondientes.

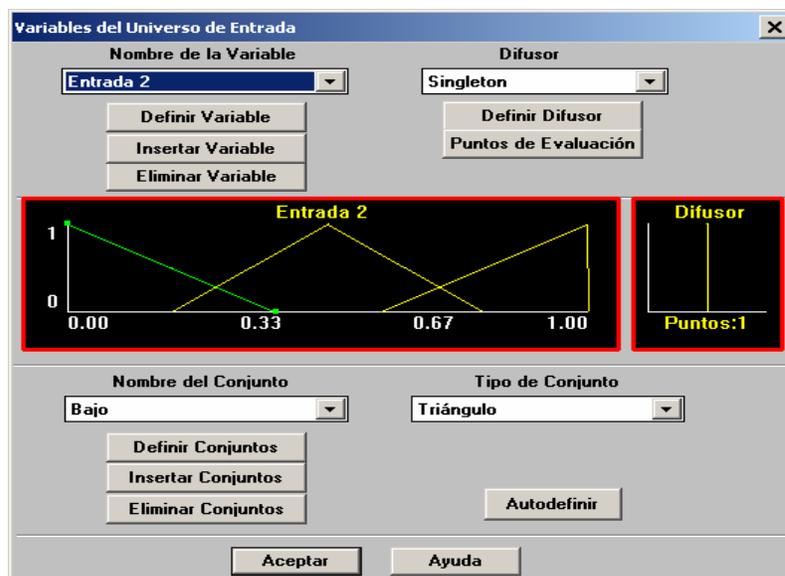


Fig. A.7. Cuadro de diálogo Definir variable de entrada.

Desde este cuadro de diálogo se tiene acceso a las opciones para eliminar, crear ó modificar las variables lingüísticas del universo de entrada.

En el caso de definir una nueva o modificar una existente, aparece una ventana donde podemos escoger el nombre, el universo de discurso de la variable (pudiendo tomar tanto valores positivos como negativos) y el número de puntos de evaluación para la generación de las tablas.

Este cuadro de diálogo también permite crear, modificar ó eliminar los distintos valores lingüísticos de cada una de las variables, así como escoger el tipo, es decir, la forma de su función de pertenencia correspondencia.

Las funciones de pertenencia para los conjuntos difusos que podemos utilizar en este programa son:

- Tipo L.
- Triángulo.
- Tipo Pi.
- Tipo Gamma.
- Tipo Z.
- Campana.
- Tipo S.
- Pi.Campana.
- Singleton.

Por último, este cuadro de diálogo nos da la opción de definir el difusor, el número de puntos de evaluación y su intervalo de discurso. Los distintos tipos de difusores que podemos escoger son:

- Singleton.
- Triángulo.
- Campana.
- Tipo Pi.
- Pi-Campana.

2.3. Universo de salida.

En la generación del universo de salida nos encontramos con las mismas opciones que los comentados para el universo de entrada.

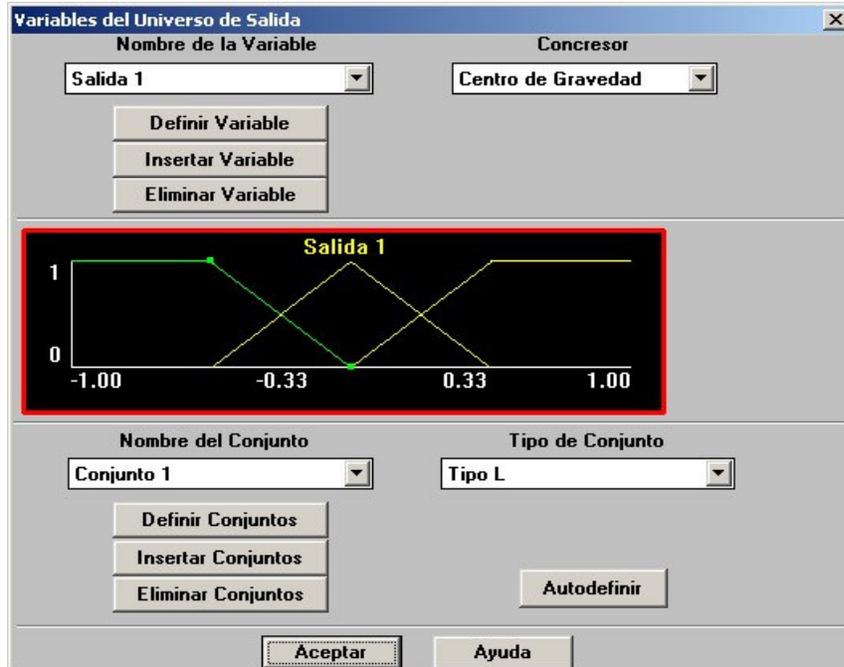


Fig. A.8. Cuadro de diálogo Variables del Universo de Salida.

Los congresores que podemos utilizar con este programa son:

- Primer máximo.
- Último máximo.
- Media de Máximos.
- Centro de gravedad.

2.4. Base de reglas.

Como en todo sistema de lógica difusa, esta aplicación necesita de una serie de reglas en las que almacena el conocimiento para poder funcionar.

Hay varias opciones para definir estas reglas:

- Asistente de Definición Rápida de Reglas.



Fig. A.9. Cuadro de diálogo Definición Rápida de la Base de Reglas.

Aunque existen varias formas de definir el conjunto de reglas, la más útil es la definición de un conjunto completo de reglas. Se crean tantas reglas como sea necesaria para cubrir el conjunto completo de posibilidades de combinación entre todos los conjuntos difusos de las variables lingüísticas de entrada.

Definir una nueva regla con este programa puede resultar algo tedioso, siendo más cómodo en muchas ocasiones crearlas todas, para ir modificando o eliminando posteriormente aquellas que sean necesarias. No obstante, la definición de reglas es el mayor inconveniente de la aplicación, por razones que se explicarán más adelante.

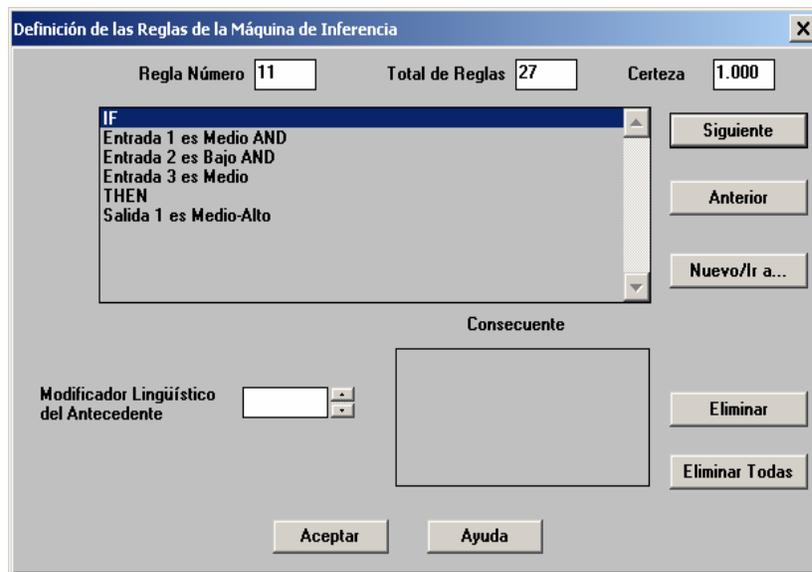


Fig. A.10 Cuadro de diálogo Definición de las Reglas de la Máquina de Inferencia.

En la parte superior se observan dos recuadros que indican el número total de reglas que tenemos definidas y cuál es la regla actual. Este cuadro de diálogo nos permite seleccionar el consecuente de una regla determinada.

La base de reglas de esta utilidad está muy limitada ya que sólo permite la definición de reglas AND y no de reglas OR. Esto motiva que haya que considerar todas las posibilidades que da una OR basándose en el operador AND con el consiguiente aumento de reglas. Este inconveniente se acentúa a medida que aumenta el número de variables y conjuntos de entrada.

2.5. Simulación.

Este programa permite dos tipos de simulaciones diferentes:

- Calcular Salidas.

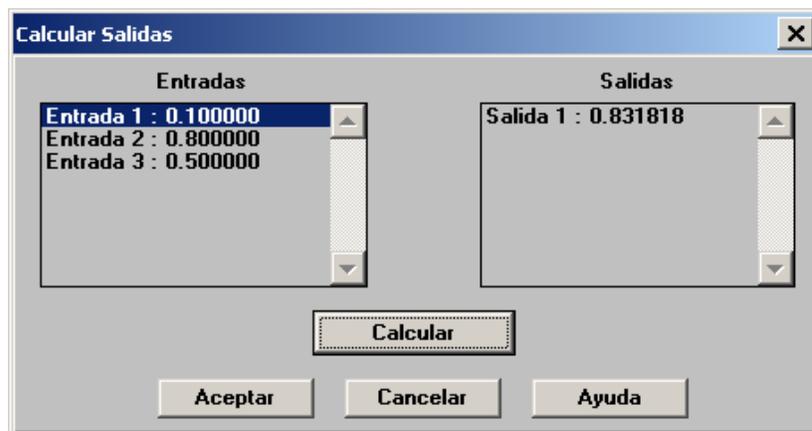


Fig. A.11 Cuadro de diálogo Calcular Salidas.

Desde este cuadro de diálogo podemos asignarle valores concretos a las entradas para observar como evolucionan las salidas. Esta opción resulta muy útil para poder comprobar si el funcionamiento de un sistema es el correcto.

- Análisis paso a paso.

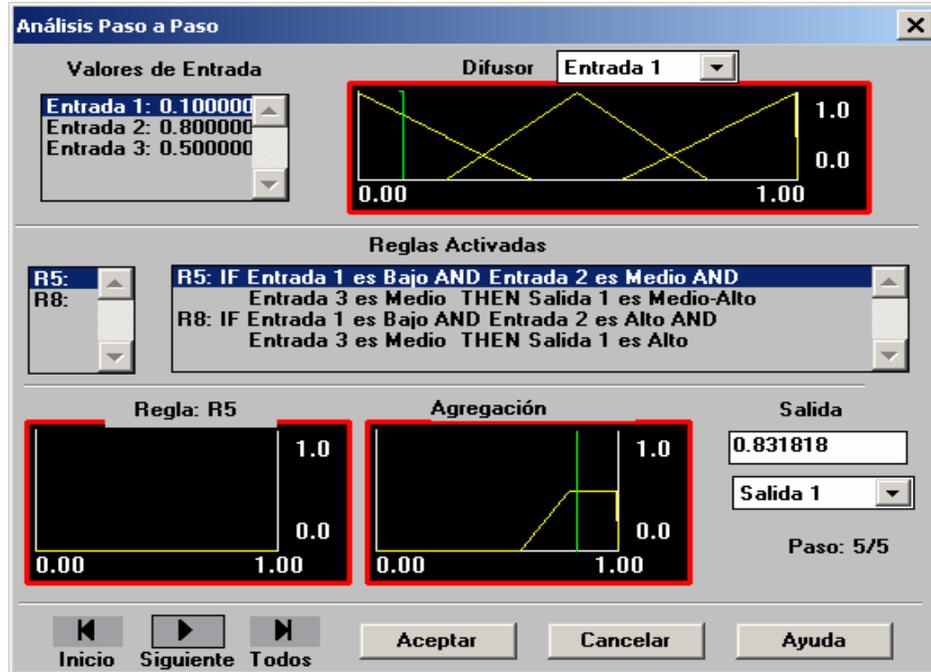


Fig. A.12. Cuadro de diálogo Análisis Paso a Paso.

Esta opción permite calcular cuáles son los valores concretos de las variables de salida para ciertos valores de las variables de entrada, mostrando los pasos intermedios que llevan a obtener tal resultado. Estos pasos intermedios nos muestran el difusor sobre la variable de entrada, las reglas que se activan, la función de pertenencia de esas reglas, la intersección de todas esas funciones de pertenencia y, por último, el resultado del conector.

2.6. Generación de código.

La principal ventaja de Un-fuzzy, y que nos hace decantarnos por la opción de usar este programa en lugar de por el potente toolbox de lógica borrosa, es la opción de generar automáticamente el código fuente del sistema de lógica borrosa en lenguaje C.

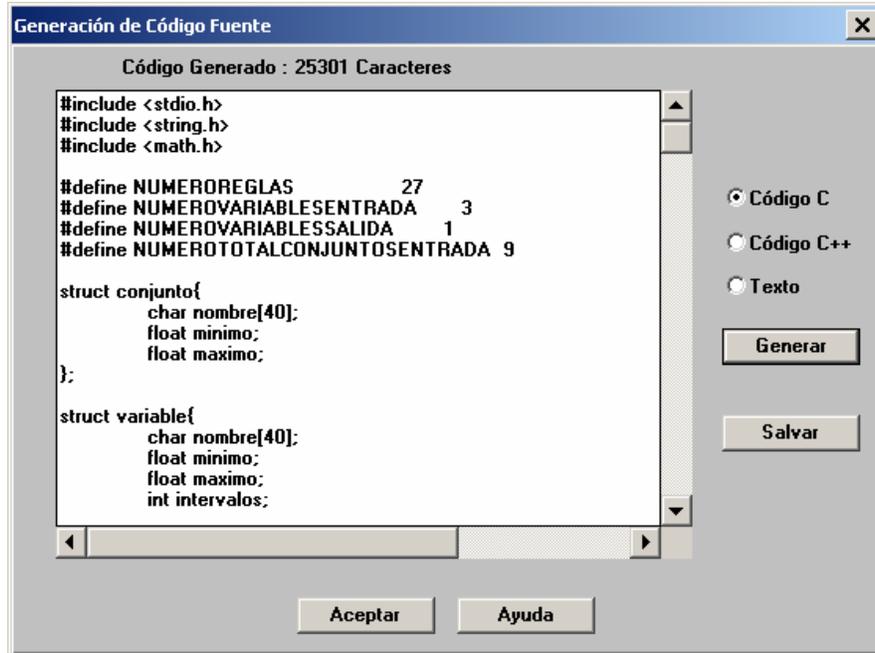


Fig. A.13 Cuadro de diálogo Generación de Código Fuente.

En esta ventana podemos escoger la generación de código fuente en C o en C++, teniendo la opción de guardarlo en el disco. El código generado es, además un código relativamente sencillo y de fácil comprensión y depuración.

APÉNDICE B. RESULTADOS DE LAS PRUEBAS

1. Relación de Preguntas-Tipo

Preguntas Tipo

111	¿Cómo puedo localizar información sobre un Centro de la Universidad de Sevilla?
112	Desearía corregir los datos de un Centro en la página WEB de la Universidad
113	¿Cómo puedo localizar información sobre un Departamento de la Universidad de Sevilla?
114	Desearía corregir los datos de un Departamento en la página WEB de la Universidad
115	¿Cómo puedo localizar información sobre un Servicio de la Universidad de Sevilla?
116	Corregir los datos mostrados en el portal de la Universidad sobre un Servicio
121	¿Cuáles son los Órganos colegiados de la Universidad de Sevilla?
122	¿Quiénes forman el equipo de Gobierno de la Universidad de Sevilla?
123	¿Está disponible en la página WEB de la Universidad la relación de puestos de trabajo de la Universidad de Sevilla?
131	Desearía conocer la Historia de la Universidad de Sevilla
132	¿Disponen de Imágenes on-line de los distintos campus de la Universidad?
133	¿Existen planos de la Universidad de Sevilla en la página WEB?
141	¿Cuáles son las leyes por las que se rige la Universidad de Sevilla?
142	¿Dónde puedo consultar el Estatuto de la Universidad de Sevilla?
143	¿Existe un Boletín de la Universidad de Sevilla?
144	¿Existe un manual que describa la imagen corporativa de la Universidad?
145	¿De dónde puedo descargar el nuevo logotipo de la Universidad de Sevilla para incluirlo en mis páginas WEB?
146	¿Dónde se pueden obtener los sellos para documentos oficiales?
211	¿Qué y cómo se estudia en las Universidades de Andalucía?
212	Desearía información sobre la prueba de acceso a la Universidad (Selectividad)
213	¿Cuáles son las notas de corte de las Titulaciones de primer y segundo ciclo?
214	¿Qué titulaciones se imparten en la Universidad de Sevilla?
215	¿Cómo realizar la Preinscripción (primer y segundo ciclo)?
216	¿Cómo realizar un traslado de expediente?
217	¿Cómo realizar las pruebas de acceso para mayores de 25 años?
218	Desearía información sobre Libre configuración
221	Necesito información sobre el programa de intercambio Erasmus
222	¿Qué cursos de postgrado ofrece la Universidad de Sevilla?
223	¿Existen cursos de español para extranjeros en la Universidad de Sevilla?
224	Necesito información y ayuda sobre alojamiento
225	Querría información sobre Sevilla
231	¿Cómo podría obtener información sobre las becas del Ministerio de Educación y Ciencia (M.E.C.)?

232	Querría saber sobre las Ayudas al estudio de la Universidad en colaboración con la Junta de Andalucía
233	¿Hay becas y ayudas propias de la Universidad de Sevilla?
234	Necesito información y ayuda sobre alojamiento
235	Necesito información sobre las becas Séneca
236	¿Cómo puedo obtener información sobre las becas ERASMUS?
241	¿Cómo buscar empleo a través de la Universidad?
242	¿Existe un Servicio de Prácticas de Empresa en la Universidad?
251	¿Cuáles son los programas de doctorado de la Universidad de Sevilla?
252	¿Qué cursos de postgrado hay en la universidad de Sevilla?
253	Querría información sobre becas y ayudas para movilidad
261	¿Qué cursos ofrece Extensión Universitaria?
262	¿Cuál es el calendario escolar de la Universidad de Sevilla?
263	¿Existe una Guía para el estudiante de la Universidad de Sevilla?
264	¿Qué es la Tarjeta Inteligente Universitaria?
265	¿Qué servicios financieros ofrece la Universidad de Sevilla a la Comunidad Universitaria?
266	¿Qué es el CADUS?
267	¿Qué es el Servicio de Asistencia a la Comunidad Universitaria (SACU)?
268	¿Cómo puedo localizar a un miembro de la Comunidad Universitaria?
269	¿Cómo puedo obtener información sobre Centros, Departamentos y Servicios de la Universidad de Sevilla?
2610	¿Existe alguna página informativa sobre Sevilla en el portal de la Universidad?
271	¿Cuáles son las normas para la matrícula?
272	¿Cuál es el régimen económico de la matrícula?
273	¿Qué normativa existe para los exámenes?
274	Querría conocer el calendario escolar
311	Querría conocer las últimas convocatorias de Acción Social
321	Querría conocer las convocatorias de oposiciones para el P.A.S.
331	¿Cómo puedo acceder al Servicio de Formación del P.A.S.?
332	Desearía obtener información sobre los Cursos de Extensión Universitaria
341	¿Cómo puedo acceder a la Vicegerencia de Recursos Humanos?
342	Necesito conocer la normativa de Recursos Humanos de la Universidad
343	Querría acceder a la Relación de Puestos de Trabajo del P.A.S.
411	¿Existen becas de apoyo a la investigación de la Junta de Andalucía?
412	¿Existen becas o ayudas para movilidad en la Universidad de Sevilla?
421	¿Cómo hacer la baremación en Concursos a plazas de Profesorado Contratado?
422	¿Qué salida tienen los profesores contratados ante la LOU?
431	Necesito información sobre el Servicio de Publicaciones de la Universidad de Sevilla
432	¿Cómo se accede al Plan de Organización Docente de la Universidad de Sevilla?
433	¿Como puedo acceder a las actas y las listas de clase? (sólo PDI de la US)
434	¿Cómo accedo a las Actas de Estudios Propios del Instituto de Idiomas?
435	¿Dónde puedo consultar la Convocatoria de Actividades para la oferta de créditos de libre configuración?

436	Querría acceder a la convocatoria de la Evaluación voluntaria de la Docencia
437	¿Cómo puedo acceder al Servicio de Personal Docente?
438	¿Cual es el procedimiento para la Evaluación de los Complementos Autonómicos para el Profesorado Universitario en Andalucía?
441	¿Cómo puedo contactar con el Instituto de Ciencias de la Educación (ICE)?
442	Querría saber cuales son los Proyectos de Formación del Profesorado aceptados por la U.C.U.A.
443	¿Existen cursos de formación organizados por la Universidad de Sevilla?
451	Querría contactar con el Vicerrectorado de Investigación
452	¿Cómo puedo acceder a la Oferta Tecnológica de la Universidad?
453	¿Cómo acceder a la información sobre Congresos, Actos y Jornadas de la Universidad de Sevilla?
454	¿Cómo puedo acceder al Sistema de Información Científica de Andalucía (S.I.C.A.)?
511	¿Cómo acceder al Catálogo de la Biblioteca de la Universidad de Sevilla?
512	Tengo problemas accediendo al Catálogo o Revistas desde fuera de la Universidad
513	¿Cómo acceder a la Fototeca de la Universidad de Sevilla?
521	¿Cómo puedo solicitar alojamiento para una página WEB institucional? (centros, departamentos, grupos de investigación, etc.)
522	Quiero modificar los datos del responsable o la persona de mantenimiento de un alojamiento institucional
523	¿Cómo puedo solicitar alojamiento para una página WEB personal? (PDI y PAS)
524	Soy alumno de la Universidad ¿Cómo puedo solicitar alojamiento para una página WEB personal?
525	¿Puedo cambiar la clave para acceder al alojamiento de mi página personal?
526	¿Puede solicitar alojamiento de página WEB personal un alumno o un profesor de secundaria?
527	Desearía que mi página WEB personal apareciera en el listado de páginas personales del portal de la Universidad.
528	He perdido u olvidado los datos de la cuenta de mi alojamiento personal en el portal de la Universidad.
529	Quiero aumentar el espacio disponible para mi página
531	¿Puedo consultar mi expediente por internet?
532	Tengo problemas con la clave de acceso para consultar mi expediente por internet
533	¿Que debo hacer si se me ha bloqueado la cuenta para consultar mi expediente por internet?
534	¿Cómo puedo corregir datos erróneos de mi expediente en internet?
535	¿Cómo puede un profesor acceder a las listas de clase o las actas a través de Internet?
541	Soy alumno de la Universidad ¿cómo puedo solicitar una cuenta de correo electrónico?
542	¿Cómo puede solicitar cuenta de correo electrónico el personal de la Universidad de Sevilla?
543	¿Cómo puedo solucionar un problema con mi clave de correo electrónico?
544	¿Puedo solicitar una lista de distribución de correo en la Universidad?
545	¿Cómo soluciono un problema con mi cuenta de correo o una lista de distribución?
551	¿Cómo solicitar un foro en el portal de la Universidad?
552	Error en la consulta

553	Quiero cambiar de moderador o el correo-e de moderación
554	Soy moderador de un foro de la Universidad ¿Puedo ocultar o dar de baja temporalmente un foro?
561	¿Cómo puedo dar un parte de incidencias al S.O.S.?
562	¿Cómo puedo dar una queja o realizar una sugerencia a la Universidad de Sevilla?
571	¿Cómo puedo localizar la dirección de correo electrónico de un alumno?
572	¿Cómo puedo localizar la dirección de correo electrónico de un PAS?
573	¿Cómo puedo localizar la dirección de correo electrónico de un profesor?
574	Querría localizar un Centro, Departamento o Servicio
575	¿Cómo puedo solicitar la corrección de datos del Servicio Directorio?
581	¿Cómo publicar Congresos, Actos y Jornadas de la Universidad en la Agenda del portal?
582	Información y reserva de Aulas de Informática
583	¿Cómo solicitar el servicio de escaneado de documentos?
584	¿Ofrece la Universidad de Sevilla un servicio de almacenamiento masivo de datos para sus usuarios?
585	Querría solicitar equipamiento: adquisición de programas y material informático
591	¿Qué significa PAUTA?
592	¿Cómo publicar en PAUTA?
593	¿Se puede usar PAUTA fuera de la red de la Universidad?
594	¿Cómo solicito la corrección de un error de datos (POD) en PAUTA?
595	¿Cómo solicito la corrección de un error en los datos de mis Asignaturas en PAUTA?
596	Necesito un aumento de mi espacio de alojamiento en PAUTA.
5101	¿Qué servicios ofrece la Tarjeta Universitaria?
5102	¿Cómo solicitar la Tarjeta Universitaria?
5103	¿Cómo puedo resolver un problema con la Tarjeta Universitaria?
5104	¿Cómo desbloqueo el PIN de mi Tarjeta Universitaria?
5105	¿Qué es el carné Universitario Mixto 4B Maestro?
5106	¿Cómo solicitar cambio de modalidad en mi carné Universitario?
5210	Quiero solicitar requisitos especiales para un alojamiento en el portal de la Universidad.

2. Resultados con motor de 3 entradas en Matlab.

1.1.1 Respuesta 1.1.1 (83.37%) Respuesta 2.2.5. (54.52%) Respuesta 2.6.9 (51.64%)
1.1.2 Respuesta 1.1.1 (83.37%) Respuesta 1.1.2 (83.37%) Respuesta 1.1.4 (83.37%)
 Respuesta 1.1.6 (83.37%) Respuesta 5.2.1 (59.99%) Respuesta 2.6.9. (51.64%) Respuesta
 5.2.3. (51.64%) Respuesta 5.2.4. (51.64%) Respuesta 5.2.6. (51.64%)
1.1.3 Respuesta 1.1.3 (83.37%) Respuesta 2.2.5. (54.52%) Respuesta 2.6.9 (51.64%)
1.1.4 Respuesta 1.1.2 (83.37%) Respuesta 1.1.3 (83.37%) Respuesta 1.1.4 (83.37%)
 Respuesta 1.1.6 (83.37%) Respuesta 5.2.1 (59.99%) Respuesta 2.6.9. (51.64%) Respuesta
 5.2.3. (51.64%) Respuesta 5.2.4. (51.64%) Respuesta 5.2.6. (51.64%)

1.1.5 Respuesta 1.1.5 (83.37 %) Respuesta 2.2.5. (54.52%) Respuesta 1.1.6 (51.64%)
Respuesta 2.6.9 (51.64%)

1.1.6 Respuesta 1.1.5 (83.37 %) **Respuesta 1.1.6 (83.37%)** Respuesta 1.1.2. (69.49 %)
Respuesta 1.1.4. (69.49%) Respuesta 2.6.9 (51.64 %)

1.2.1 Respuesta 1.2.1 (83.37 %) Respuesta 2.2.5. (54.52%)

1.2.2 Respuesta 1.2.2 (83.37 %) Respuesta 2.2.5. (54.52%)

1.2.3 Respuesta 3.4.3 (87.00%) O.K. Respuesta 1.2.3. (83.37%) Respuesta 1.1.2. (69.49
%) Respuesta 1.1.4. (69.49%) Respuesta 1.1.6 (69.49%) Respuesta 2.2.5. (54.52%)
Respuesta 5.2.1. (51.64%) Respuesta 5.2.3. (51.64%) Respuesta 5.2.4. (51.64%) Respuesta
5.2.6. (51.64%)

1.3.1 Respuesta 1.3.1 (83.37%) Respuesta 2.2.5. (54.52%)

1.3.2 Respuesta 1.3.2 (83.37 %)

1.3.3 Respuesta 1.3.3 (84.70%) Respuesta 1.1.2 (60.00%) Respuesta 1.1.4 (60.00 %)
Respuesta 1.1.6 (60.00%) Respuesta 5.2.1 (51.64%) Respuesta 5.2.3 (51.64%) Respuesta
5.2.4 (51.64 %) Respuesta 5.2.6. (51.64 %)

1.4.1 Respuesta 1.4.1 (83.37%) Respuesta 2.2.5. (54.52%)

1.4.2 Respuesta 1.4.2 (83.37%) Respuesta 2.2.5. (54.52%)

1.4.3 Respuesta 1.4.3 (83.37%) Respuesta 2.2.5. (54.52%)

1.4.4 Respuesta 1.4.4 (84.70 %)

1.4.5 Respuesta 1.4.5 (83.37 %) Respuesta 1.1.2 (51.64 %) Respuesta 1.1.4 (51.64 %)
Respuesta 1.1.6 (51.64 %)

1.4.6 Respuesta 1.4.6 (83.37 %)

2.1.1 Respuesta 2.1.1 (53.13 %)

2.1.2 Respuesta 2.1.2 (69.49%) Respuesta 2.1.7 (60.00%)

2.1.2 (bis) **Respuesta 2.1.2 (86.28%)**

2.1.3 Respuesta 2.1.5 (87.00%) Respuesta 2.1.3 (83.37 %)

2.1.4 Respuesta 2.2.5 (54.52%)

2.1.5 Respuesta 2.1.5 (84.70%) Respuesta 2.1.3 (53.13 %)

2.1.6 Respuesta 2.1.6 (84.70%) Respuesta 5.3.1 (84.70 %)

2.1.7 Respuesta 2.1.7 (84.70%) Respuesta 2.1.2 (65.85%)

2.1.8 Respuesta 2.1.8 (83.37%)

2.2.1 Respuesta 2.3.6 (86.28 %) OK Respuesta 2.2.1(84.70%)

2.2.2 Respuesta 2.5.2 (84.70 %) OK Respuesta 2.2.2 (83.37%) Respuesta 2.2.5. (54.52%)

2.2.3 Respuesta 2.2.3 (69.49%)

2.2.4 Respuesta 2.2.4 (86.28%) Respuesta 2.3.4 (86.28%)

2.2.5 NO

2.3.1 Respuesta 2.3.1 (69.49 %) Respuesta 4.1.1 (54.52%) Respuesta 4.1.2 (54.52%)

2.3.1 (bis) **Respuesta 2.3.1 (60.00 %)** Respuesta 4.1.1 (54.52%) Respuesta 4.1.2 (54.52%)

2.3.2 Respuesta 2.3.2 (85.83%) Respuesta 2.3.3 (60.00%) Respuesta 4.1.1 (85.83%)
Respuesta 4.1.2 (51.64%)

2.3.3 Respuesta 2.3.3 (85.83%) Respuesta 4.1.2 (69.49%) Respuesta 4.1.1 (65.85%)
Respuesta 2.2.5 (54.52%) Respuesta 2.3.2 (51.64%)

2.3.4 Respuesta 2.2.4 (86.28%) OK Respuesta 2.3.4 (86.28%)

2.3.5 Respuesta 2.3.5 (86.28%) Respuesta 4.1.1 (54.52%) Respuesta 4.1.2 (54.52%)

2.3.6 Respuesta 2.3.6 (86.28 %) Respuesta 2.2.1(84.70%) Respuesta 4.1.1 (54.52%)
Respuesta 4.1.2 (54.52%)
2.4.1 Respuesta 2.4.1 (54.52 %)
2.4.2 Respuesta 1.1.5 (83.37 %) **Respuesta 2.4.2 (69.49 %)** Respuesta 1.1.6 (51.64%)
Respuesta 2.6.9 (51.64%)
2.5.1 Respuesta 2.5.1 (83.37 %) Respuesta 2.2.5. (54.52%)
2.5.2 Respuesta 2.5.2 (84.70 %) Respuesta 2.2.2 (83.37%) Respuesta 2.2.5. (54.52%)
2.5.3 Respuesta 2.5.3 (83.37 %) Respuesta 4.1.2 (83.37%) Respuesta 4.1.1 (65.85%)
Respuesta 2.3.3 (60.00%)
2.6.1 Respuesta 2.6.1 (83.37 %) Respuesta 3.3.2 (83.37%)
2.6.2 Respuesta 2.6.2 (65.85 %) Respuesta 2.7.4 (65.85%) Respuesta 2.2.5. (54.52%)
2.6.3 Respuesta 2.6.3 (84.70 %) Respuesta 2.2.5. (54.52%)
2.6.4 Respuesta 2.6.4 (85.83 %) Respuesta 5.10.1 (65.85%) Respuesta 5.10.2 (65.85%)
Respuesta 5.10.3 (51.64%) Respuesta 5.10.4 (51.64%) Respuesta 5.10.5 (51.64%)
2.6.5 Respuesta 2.6.5 (83.37 %) Respuesta 5.10.1 (60.00%) Respuesta 2.6.7 (51.64%)
Respuesta 2.6.8 (51.64%) Respuesta 2.6.9 (51.64%)
2.6.6 Respuesta 2.6.2 (86.28 %)
2.6.7 Respuesta 2.6.7 (59.99 %) Respuesta 2.6.8 (51.64%)
2.6.7 (bis) **Respuesta 2.6.7 (86.28 %)**
2.6.8 Respuesta 2.6.8 (85.83 %)
2.6.9 Respuesta 2.6.9 (69.49 %) Respuesta 2.2.5. (54.52%) Respuesta 2.6.5 (51.64%)
2.7.1 Respuesta 2.7.1 (65.85 %) Respuesta 2.7.2 (54.52%)
2.7.2 Respuesta 2.7.2 (83.37 %) Respuesta 2.7.1 (54.52%)
2.7.3 Respuesta 2.7.3 (65.85 %)
2.7.4 Respuesta 2.6.2 (65.85%) OK Respuesta 2.7.4 (65.85 %)
3.1.1 Respuesta 3.1.1 (83.37 %)
3.2.1 Respuesta 3.2.1 (83.37 %) Respuesta 3.3.1 (54.52%)
3.3.1 Respuesta 3.2.1 (84.70%) Respuesta 3.3.1 (83.37 %) Respuesta 5.7.2 (54.52%)
3.3.2 Respuesta 2.6.1 (83.37 %) OK Respuesta 3.3.2 (83.37 %)
3.4.1 Respuesta 3.4.1 (84.70 %) Respuesta 3.4.2 (65.85%)
3.4.2 Respuesta 3.4.2 (83.37 %)
3.4.3 Respuesta 3.4.3 (87.00 %) Respuesta 3.2.1 (84.70%) Respuesta 1.2.3 (83.37%)
Respuesta 3.3.1 (54.52%)
3.4.3 (bis) **Respuesta 3.4.3 (86.28 %)** Respuesta 3.2.1 (84.70%) Respuesta 3.3.1 (54.52%)
4.1.1 Respuesta 4.5.1 (84.70%) Respuesta 4.1.1 (83.37 %) Respuesta 4.1.2 (57.51%)
Respuesta 2.3.2 (51.64%)
4.1.2 Respuesta 2.3.3 (85.83%) Respuesta 2.5.3 (83.37%) Respuesta 4.1.2 (83.37 %)
Respuesta 2.2.5 (54.52%) Respuesta 2.3.2 (51.64%)
4.2.1 Respuesta 4.2.1 (83.37 %) Respuesta 4.2.2 (65.85%)
4.2.2 Respuesta 4.2.2 (83.37 %) Respuesta 4.2.1 (65.85%)
4.3.1 Respuesta 1.1.5 (83.37 %) Respuesta 4.3.1 (83.37 %) Respuesta 2.2.5. (54.52%)
Respuesta 1.1.6 (51.64%) Respuesta 2.6.9 (51.64%)
4.3.2 Respuesta 4.3.2 (83.37 %) Respuesta 2.2.5. (54.52%)
4.3.3 Respuesta 5.3.5 (84.70%) OK Respuesta 4.3.3 (83.37 %)

4.3.4 Respuesta 4.3.4 (85.83%) Respuesta 5.3.5 (84.70%) Respuesta 4.3.3 (84.70%)
4.3.5 Respuesta 4.3.5 (85.83%) **Respuesta 2.1.8 (83.37%)**
4.3.6 Respuesta 4.3.6 (85.83%)
4.3.7 Respuesta 4.3.7 (85.83%)
4.3.8 Respuesta 4.3.8 (85.83%)
4.4.1 Respuesta 4.4.1 (85.83%)
4.4.1 (bis) **Respuesta 4.4.1 (86.78%)**
4.4.2 Respuesta 4.4.2 (83.37%) Respuesta 4.4.3 (54.54 %)
4.4.3 Respuesta 4.4.3 (83.37 %) Respuesta 2.2.5. (54.52%) Respuesta 4.4.2 (54.52%)
4.5.1 Respuesta 4.5.1 (84.70%)
4.5.2 Respuesta 4.5.2 (69.49%)
4.5.3 Respuesta 5.8.1 (85.83%) OK **Respuesta 4.5.3 (83.37%)** Respuesta 2.2.5. (54.52%)
4.5.4 Respuesta 4.5.4 (85.83%)
4.5.4 (bis) Respuesta 4.5.4 (86.78%)
5.1.1 Respuesta 5.1.2 (85.83%) **Respuesta 5.1.1 (83.37%)** Respuesta 2.2.5. (54.52%)
5.1.2 Respuesta 5.1.2 (85.83%) Respuesta 2.2.5. (54.52%)
5.1.3 Respuesta 5.1.3 (83.37%) Respuesta 2.2.5. (54.52%)
5.1.4 Respuesta 5.1.4 (69.49%) Respuesta 2.2.5. (54.52%)
5.2.1 Respuesta 2.2.4 (86.28%) Respuesta 2.3.4 (86.28%) **Respuesta 5.2.1 (85.83%)**
Respuesta 5.2.3 (69.49%) Respuesta 5.2.4 (69.49%) Respuesta 5.2.6 (69.49%) Respuesta
5.2.5 (51.64%)
5.2.2 Respuesta 2.2.4 (86.28%) **Respuesta 5.2.2 (85.83%)** Respuesta 5.2.1 (60.00%)
Respuesta 5.2.5 (51.64%)
5.2.3 Respuesta 2.2.4 (86.28%) Respuesta 2.3.4 (86.28%) Respuesta 5.2.1 (85.83%)
Respuesta 3.2.1 (84.70%) **Respuesta 5.2.3 (83.37%)** Respuesta 5.2.4 (69.49%) Respuesta
5.2.6 (69.49%) Respuesta 1.1.2 (60.00%) Respuesta 1.1.4 (60.00%) Respuesta 1.1.6
(60.00%) Respuesta 3.3.1 (54.52%) Respuesta 5.2.5 (51.64%)
5.2.4 Respuesta 2.2.4 (86.28%) Respuesta 2.3.4 (86.28%) Respuesta 5.2.3 (85.83%)
Respuesta 5.2.6 (85.83%) **Respuesta 5.2.4 (84.70%)** Respuesta 1.1.2 (69.49%) Respuesta
1.1.4 (69.49%) Respuesta 1.1.6 (69.49%) Respuesta 5.2.1 (69.49%) Respuesta 5.2.5
(69.49%) Respuesta 5.2.7 (60.00%)
5.2.5 Respuesta 2.2.4 (86.28%) Respuesta 2.3.4 (86.28%) Respuesta 5.2.3 (85.83%)
Respuesta 5.2.4 (85.83%) **Respuesta 5.2.5 (85.83%)** Respuesta 5.2.6 (69.49%) Respuesta
5.2.1 (60.00%) Respuesta 5.2.7 (60.00%) Respuesta 5.2.2 (51.64%)
5.2.6 Respuesta 2.2.4 (86.28%) Respuesta 2.3.4 (86.28%) Respuesta 5.2.3 (85.83%)
Respuesta 5.2.4 (84.70%) **Respuesta 5.2.6 (83.37%)** Respuesta 5.2.1 (69.49%) Respuesta
5.2.5 (69.49%) Respuesta 1.1.2 (60.00%) Respuesta 1.1.4 (60.00%) Respuesta 1.1.6
(60.00%) Respuesta 5.2.7 (60.00%) Respuesta 5.7.1 (54.52%) Respuesta 5.7.3 (54.52%)
5.2.7 Respuesta 1.1.2 (87.00%) Respuesta 1.1.4 (87.00%) Respuesta 1.1.6 (87.00%)
Respuesta 5.2.7 (85.83%) Respuesta 5.2.3 (69.49%) Respuesta 5.2.4 (69.49%) Respuesta
5.2.5 (60.00%) Respuesta 5.2.6 (59.99%) Respuesta 5.2.1 (51.64%)
5.2.8 Respuesta 2.2.4 (86.28%) Respuesta 2.3.4 (86.28%) **Respuesta 5.2.8 (85.83%)**
Respuesta 5.2.3 (60.00%) Respuesta 5.2.4 (60.00%) Respuesta 5.4.2 (60.00%) Respuesta

1.1.2 (59.99%) Respuesta 1.1.4 (59.99%) Respuesta 1.1.6 (59.99%) Respuesta 5.2.5 (51.64%) Respuesta 5.2.6 (51.64%)
5.2.9 Respuesta 5.2.9 (85.83%)
5.2.10 Respuesta 2.2.4 (86.28%) Respuesta 2.3.4 (86.28%) **Respuesta 5.2.10 (69.49%)**
5.3.1 Respuesta 5.3.2 (85.83%) Respuesta 5.3.3 (85.83%) **Respuesta 5.3.1 (83.37%)**
Respuesta 5.3.4 (60.00%) Respuesta 5.3.5 (54.52%)
5.3.2 Respuesta 5.3.2 (85.83%) Respuesta 5.3.3 (85.83%) Respuesta 5.3.1 (83.37%)
Respuesta 5.3.4 (60.00%) Respuesta 5.5.2 (60.00%) Respuesta 2.1.2 (54.52%) Respuesta 5.3.5 (54.52%)
5.3.3 Respuesta 5.3.2 (85.83%) **Respuesta 5.3.3 (83.37%)** Respuesta 5.3.1 (83.37%)
Respuesta 5.3.4 (60.00%) Respuesta 5.3.5 (54.52%)
5.3.4 Respuesta 5.3.4 (85.83%) Respuesta 5.3.1 (84.70 %) Respuesta 5.3.2 (60.00%)
Respuesta 5.3.3 (60.00%) Respuesta 1.1.2 (57.51%) Respuesta 1.1.4 (57.51%) Respuesta 1.1.6 (57.51%) Respuesta 5.7.5 (57.51%) Respuesta 5.3.5 (54.52%)
5.3.5 Respuesta 4.3.3 (83.37%) OK **Respuesta 5.3.5 (83.37%)**
5.4.1 Respuesta 5.4.1 (85.83%) Respuesta 5.2.4 (84.70%) Respuesta 5.4.2 (69.49%)
Respuesta 5.4.5 (69.49%) Respuesta 5.7.1 (69.49%) Respuesta 5.7.2 (51.64%) Respuesta 5.7.3 (51.64%)
5.4.2 Respuesta 5.4.2 (85.83%) Respuesta 5.4.1 (69.49%) Respuesta 5.4.5 (69.49%)
Respuesta 5.2.8 (60.00%) Respuesta 2.2.5 (54.52%) Respuesta 5.7.1 (51.64%) Respuesta 5.7.2 (51.64%) Respuesta 5.7.3 (51.64%)
5.4.3 Respuesta 5.4.5 (69.49%) **Respuesta 5.4.3 (60.00%)** Respuesta 5.4.4 (60.00%)
Respuesta 5.5.2 (60.00%) Respuesta 5.7.1 (51.64%) Respuesta 5.7.2 (51.64%) Respuesta 5.7.3 (51.64%)
5.4.4 Respuesta 5.4.4 (85.83%)
5.4.5 Respuesta 5.4.4 (85.83%) **Respuesta 5.4.5 (85.83%)** Respuesta 5.4.1 (60.00%)
Respuesta 5.4.2 (60.00%) Respuesta 5.4.3 (60.00%)
5.5.1 Respuesta 5.5.1 (65.85%)
5.5.2 Respuesta 5.5.2 (87.00%) Respuesta 5.3.2 (60.00%) Respuesta 5.5.1 (54.52%)
5.5.3 Respuesta 5.5.3 (85.83%)
5.5.4 Respuesta 5.5.4 (84.70%) Respuesta 5.5.1 (54.52%)
5.6.1 Respuesta 5.6.1 (85.83%)
5.6.2 Respuesta 5.6.2 (65.85%) Respuesta 2.2.5. (54.52%)
5.7.1 Respuesta 5.7.1 (83.37%) Respuesta 5.3.1 (69.49%) Respuesta 5.7.2 (69.49%)
Respuesta 5.7.3 (69.49%) Respuesta 5.4.2 (51.64%) Respuesta 5.4.5 (51.64%)
5.7.2 Respuesta 3.2.1 (84.70%) **Respuesta 5.7.2 (83.37%)** Respuesta 5.7.1 (69.49%)
Respuesta 5.7.3 (69.49%) Respuesta 3.3.1 (54.52%) Respuesta 5.4.1 (51.64%) Respuesta 5.4.2 (51.64%) Respuesta 5.4.5 (51.64%)
5.7.3 Respuesta 5.7.3 (83.37%) Respuesta 5.7.1 (69.49%) Respuesta 5.7.2 (69.49%)
Respuesta 5.4.1 (51.64%) Respuesta 5.4.2 (51.64%) Respuesta 5.4.5 (51.64%)
5.7.4 Respuesta 1.1.2 (86.28%) OK Respuesta 1.1.4 (86.28%) Respuesta 1.1.6 (86.28%)
Respuesta 2.6.9 (85.83%) **Respuesta 5.7.4 (84.70%)**

5.7.5 Respuesta 1.1.5 (86.28%) **Respuesta 5.7.5 (83.37%)** Respuesta 5.9.4 (60.00%)
Respuesta 5.9.5 (60.00%) Respuesta 1.1.6 (51.64%) Respuesta 5.8.3 (51.64%) Respuesta
5.8.4 (51.64%)
5.8.1 Respuesta 4.5.3 (83.37%) **OK Respuesta 5.8.1 (83.37%)**
5.8.2 Respuesta 5.8.2 (83.37%)
5.8.3 Respuesta 1.1.5 (86.28%) **Respuesta 5.8.3 (85.83%)**
5.8.4 Respuesta 1.1.5 (83.37 %) **Respuesta 5.8.4 (83.37%)** Respuesta 1.1.6 (59.99%)
Respuesta 2.2.5 (54.52%) Respuesta 2.6.9 (51.64%) Respuesta 5.7.5 (51.64%)
5.8.5 Respuesta 5.8.5 (85.83%)
5.9.1 Respuesta 5.9.1 (86.78 %)
5.9.2 Respuesta 5.9.1 (86.78 %) **Respuesta 5.9.2 (65.85 %)**
5.9.3 Respuesta 5.9.1 (86.78 %) **Respuesta 5.9.3 (69.49%)**
5.9.4 Respuesta 5.9.1 (86.78 %) **Respuesta 5.9.4 (85.83%)** Respuesta 5.9.5 (85.83%)
Respuesta 5.7.5 (57.51%)
5.9.4 (bis) Respuesta 5.9.1 (86.78 %) **Respuesta 5.9.4 (85.83%)** Respuesta 5.9.5 (60.00%)
5.9.5 Respuesta 5.9.1 (86.78 %) Respuesta 5.9.4 (85.83%) **Respuesta 5.9.5 (85.83%)**
Respuesta 5.7.5 (57.51%)
5.9.6 Respuesta 5.9.1 (86.78 %) Respuesta 2.2.4 (86.28%) Respuesta 2.3.4 (86.28%)
Respuesta 5.9.6 (85.83%) Respuesta 5.2.8 (60.00%)
5.10.1 Respuesta 5.10.1 (83.37 %) Respuesta 5.10.2 (65.85%) Respuesta 2.6.4 (60.00%)
Respuesta 2.6.5 (51.64%) Respuesta 5.10.3 (51.64%) Respuesta 5.10.4 (51.64%) Respuesta
5.10.5 (51.64%)
5.10.2 Respuesta 5.10.2 (83.37 %) Respuesta 5.10.1 (65.85%) Respuesta 2.6.4 (60.00%)
Respuesta 5.10.3 (51.64%) Respuesta 5.10.4 (51.64%) Respuesta 5.10.5 (51.64%)
Respuesta 5.10.6 (51.64%)
5.10.3 Respuesta 5.10.3 (69.49%) Respuesta 5.10.1 (65.85%) Respuesta 5.10.2 (65.85%)
Respuesta 2.6.4 (60.00%) Respuesta 5.10.4 (51.64%) Respuesta 5.10.5 (51.64%)
5.10.4 Respuesta 5.10.4 (85.83%) Respuesta 5.10.1 (65.85%) Respuesta 5.10.2 (65.85%)
Respuesta 2.6.4 (60.00%) Respuesta 5.10.3 (51.64%) Respuesta 5.10.5 (51.64%)
5.10.5 Respuesta 5.10.5 (84.70%) Respuesta 5.10.1 (65.85%) Respuesta 5.10.2 (65.85%)
Respuesta 5.10.3 (51.64%) Respuesta 5.10.4 (51.64%)
5.10.6 Respuesta 5.10.6 (85.83%) Respuesta 5.10.2 (83.37%) Respuesta 5.10.1 (65.85%)
Respuesta 5.10.3 (51.64%) Respuesta 5.10.4 (51.64%) Respuesta 5.10.5 (51.64%)

3. Resultados del motor de 3 entradas con umbrales fijos

Realizado con la Versión 050914 **ENTRADAS** con Umbrales 0.5/0.5/0.5 fijos
Resultados utilizando el **MOTOR DE 3**

En negrita el tema, apartado o pregunta correctos. Al final en negrita el mayor índice de

certeza. Si está en cursiva indica que no corresponde al correcto

Tema1: Tema1

Pregunta	Preguntas	CertezaPregunta
111	111 , 112 , 574	0,819 ; 0,517 , 0,831
112	111 ; 112 ; 114; 116 ; 521 ; 574 ; 575	0,600 ; 0,646 ; 0,517 ; 0,517 ; 0,517 ; 0,831 ; 0,564
113	113 ; 114 ; 571 ; 573 : 574	0,819 ; 0,517 ; 0,517 ; 0,517 ; 0,831
114	112 ; 113 ; 114 ; 116 ; 521 ; 574 ; 575	0,517 ; 0,600 ; 0,646 ; 0,517 ; 0,517 ; 0,831 ; 0,564
115	571 ; 573 ; 574 ; 575 ; 584	0,517 ; 0,517 ; 0,842 ; 0,599 ; 0,831
116	112 ; 114 ; 115 ; 116 ; 575 ; 584	0,517 ; 0,517 ; 0,600 ; 0,646 ; 0,680 ; 0,819
121	121 .	0,819 .
122	122 ; 225	0,819 ; 0,646
123		
131	131 ; 225	0,819 ; 0,646
132	132 .	0,819 .
133	112 ; 114 ; 116 ; 133 ; 521 ; 523 ; 524 ; 525 ; 526 ; 527 ; 529	0,680 ; 0,680 ; 0,680 ; 0,680 ; 0,517 ; 0,517 ; 0,680 ; 0,600 ; 0,680 ; 0,600 ; 0,600
141	141 ; 224 ; 225	0,819 ; 831 ; 0,531
142	142 .	0,819 .
143	143 ; 225	0,819 ; 0,531
144	144 .	0,831 .
145	112 ; 114 ; 116 ; 145	0,517 ; 0,517 ; 0,517 ; 0,819
146	146 .	0,819 .

Tema2: Tema2

Pregunta	Preguntas	CertezaPregunta
211	Ninguna	La certeza de la pregunta es inferior al umbral
212	212 ; 217	0,819 ; 0,600
213	213 .	0,646 .
214	225 .	0,531 .
215	215 .	0,831 .
216	216 ; 531	0,831 ; 0,831
217	212 ; 217	0,646 ; 0,831
218	218 ; 435	0,819 ; 0,600
221	221 ; 235 ; 236	0,831 ; 0,531 ; 0,846

222	222 ; 225 ; 252 ; 332	0,819 ; 0,531 ; 0,831 ; 0,531
223	223 .	0,599 .
224	224 ; 234 ;	0,846 ; 0,846
225		
231	231 ; 411 ; 412	0,517 ; 0,531 ; 0,531
232	232 ; 411	0,599 ; 0,842
233	233 ; 235 ; 411 ; 412	0,600 ; 0,842 ; 0,646 ; 0,680
234	224 ; 234 ; 411 ; 412 ;	0,846 ; 0,846 ; 0,646 ; 0,680
235	235 ; 411 ; 412	0,846 ; 0,646 ; 0,680
236	221 ; 236 ; 411 ; 412	0,831 ; 0,846 ; 0,646 ; 0,680
241	No hay ninguna pregunta	
242	242 .	0,646 .
251	225 ; 251 ; 582 ; 585	0,531 ; 0,831 ; 0,517 ; 0,600
252	222 ; 225 ; 252	0,819 ; 0,531 ; 0,819
253	233 ; 242 ; 253 ; 411 ; 412	0,600 ; 0,819 ; 0,819 ; 0,646 ; 0,819
261	261 ; 332	0,819 ; 0,819
262	225 ; 262 ; 265 ; 273 ; 274	0,531 ; 0,646 ; 0,531 ; 0,531 ; 0,646
263	225 ; 262 ; 265	0,531 ; 0,831 ; 0,531
264	264 .	0,599 .
265	265 ; 267	0,599 ; 0,599
266	266 .	0,846 .
267	267 .	0,680 .
268	267 ; 268	0,517 ; 0,680
269	133 ; 269	0,564 ; 0,600
2610	112 ; 114 ; 116 ; 525 ; 529	0,517 ; 0,517 ; 0,517 ; 0,599 ; 0,599
271	271 ; 272	0,646 ; 0,531
272	271 ; 272	0,531 ; 0,646
273	273 .	0,646 .
274	262 ; 274	0,646 ; 0,646

Tema3: Tema3

Pregunta	Preguntas	CertezaPregunta
311	311 ,	0,819 ,
321	321 ; 331	0,819 ; 0,531
331	321 ; 331 ; 442 ; 443	0,831 ; 0,531 ; 0,531 ; 0,531
332	261 ; 332	0,819 ; 0,819
341	342 ; 532	0,646 ; 0,819
342	342 .	0,680 .
343	123 ; 321 ; 331 ; 343	0,819 ; 0,831 ; 0,531 ; 0,850

Tema4: Tema4

Pregunta		
Pregunta	Preguntas	CertezaPregunta
411	411 ; 412 ; 451	0,819 ; 0,517 ; 0,831
412	232 ; 233 ; 253 ; 411 ; 412	0,517 ; 0,842 ; 0,646 ; 0,599 ; 0,646
421	421 .	0,599 .
422	421 ; 422	0,646 ; 0,819
431	431 ; 572	0,819 ; 0,646
432	432 .	0,680 .
433	433 ; 535	0,819 ; 0,819
434	433 ; 434 ; 535	0,831 ; 0,842 ; 0,831
435	218 ; 435	0,819 ; 0,517
436	436 .	0,842 .
437	437 ; 575	0,842 ; 0,600
438	436 ; 438	0,517 ; 0,599
441	441 .	0,842 .
442	442 ; 443	0,819 ; 0,531
443	225 ; 442 ; 443	0,531 ; 0,531 ; 0,819
451	451 .	0,831 .
452	525 .	0,600 .
453	453 ; 581	0,646 ; 0,842
454	451 ; 452 ; 453 ; 454	0,599 ; 0,599 ; 0,599 ; 0,819

Tema5: Tema5

Pregunta		
Pregunta	Preguntas	CertezaPregunta
511	511 ; 512	0,599 ; 0,600
512	512 .	0,599 .
513	513 .	0,819 .
521	224 ; 234 ; 269 ; 521 ; 523 ; 524 ; 526	0,846 ; 0,846 ; 0,600 ; 0,599 ; 0,517 ; 0,517 ; 0,517
522	224 ; 234 ; 522	0,846 ; 0,846 ; 0,600
523	224 ; 234 ; 321 ; 331 ; 521 ; 523 ; 524 ; 525 ; 526	0,846 ; 0,846 ; 0,831 ; 0,531 ; 0,599 ; 0,599 ; 0,599 ; 0,517 ; 0,599
524	224 ; 234 ; 521 ; 523 ; 524 ; 525 ; 526	0,846 ; 0,846 ; 0,680 ; 0,680 ; 0,680 ; 0,517 ; 0,680
525	224 ; 234 ; 522 ; 525	0,846 ; 0,846 ; 0,517 ; 0,599

526	224 ; 234 ; 521 ; 523 ; 524 ; 525 ; 526	0,846 ; 0,846 ; 0,599 ; 0,599 ; 0,599 ; 0,517 ; 0,599
527	112 ; 114 ; 116 ; 421 ; 521 ; 523 ; 524 ; 525 ; 526 ; 527	0,850 ; 0,850 ; 0,850 ; 0,680 ; 0,517 ; 0,599 ; 0,599 ; 0,600 ; 0,517 ; 0,600
528	224 ; 234 ; 421 ; 528	0,846 ; 0,846 ; 0,680 ; 0,599
529	529 ; 596	0,842 ; 0,600
531	531 ; 532 ; 533 ; 534 ; 535	0,819 ; 0,842 ; 0,842 ; 0,599 ; 0,531
532	212 ; 532 ; 552	0,531 ; 0,599 ; 0,599
533	531 ; 532 ; 533	
534	534 ; 575	0,599 ; 0,564
535	433 ; 524 ; 526 ; 5210 ; 535	0,819 ; 0,517 ; 0,600 ; 0,517 ; 0,517
541	541 ; 542 ; 545 ; 571 ; 572 ; 573	0,600 ; 0,600 ; 0,599 ; 0,517 ; 0,517 ; 0,517
542	528 ; 541 ; 542 ; 545 ; 571 ; 572 ; 573	0,599 ; 0,599 ; 0,599 ; 0,599 ; 0,517 ; 0,517 ; 0,517
543	543 ; 544 ; 545 ; 552 ; 571 ; 572 ; 573	0,599 ; 0,600 ; 0,599 ; 0,599 ; 0,517 ; 0,517 ; 0,517
544	544 ; 571 ; 572 ; 573 ; 574	0,842 ; 0,600 ; 0,600 ; 0,599 ; 0,600
545	541 ; 542 ; 543 ; 545	0,599 ; 0,599 ; 0,599 ; 0,599
551	551 .	0,531 .
552		
553	553 .	0,842 .
554	553 ; 554	0,517 ; 0,517
561	561 .	0,842 .
562	562 .	0,646 .
571	541 ; 542 ; 545 ; 571 ; 572 ; 573	0,517 ; 0,517 ; 0,517 ; 0,680 ; 0,680 ; 0,680
572	321 ; 331 ; 332 ; 541 ; 542 ; 544 ; 545 ; 571 ; 572 ; 573	0,831 ; 0,819 ; 0,599 ; 0,517 ; 0,517 ; 0,600 ; 0,517 ; 0,680 ; 0,680 ; 0,680
573	541 ; 542 ; 545 ; 571 ; 572 ; 573	0,517 ; 0,517 ; 0,517 ; 0,599 ; 0,599 ; 0,599
574	111 ; 113 ; 115 ; 269 ; 574	0,846 ; 0,846 ; 0,846 ; 0,600 ; 0,831
575	115 ; 116 ; 575 ; 594 ; 595	0,846 ; 0,517 ; 0,680 ; 0,600 ; 0,600
581	453 ; 581	0,819 ; 0,646
582	582 .	0,819 .
583	115 ; 571 ; 583	0,846 ; 0,600 ; 0,842
584	575 ; 584	0,517 ; 0,819
585	585 .	0,842 ;
591	591 .	0,849 .
592	591 ; 592	0,849 ; 0,646
593	591 ; 593	0,849 ; 0,680
594	575 ; 594 ; 595	0,564 ; 0,600 ; 0,600

595	575 ; 594 ; 595	0,564 ; 0,600 ; 0,600
596	224 ; 234 ; 529 ; 596	0,846 ; 0,842 ; 0,600 ; 0,842
5101	264 ; 265	0,599 ; 0,517
5102	264 .	0,599 .
5103	264 .	0,599 .
5104	264 .	0,599 .
5105	Ninguna supera el umbral	
5106	Ninguna supera el umbral	
5210	224 ; 223 ; 5210	0,846 ; 0,846 ; 0,680

4. Motor de 5 entradas con umbrales fijos

Realizado con la Versión 050923 y
050923b **MOTOR 5 ENTRADAS**

Umbrales 0.5/0.5/0.5 fijos

En negrita el tema, apartado o pregunta correctos. Al final en negrita el mayor índice de certeza. Si está en cursiva indica que no corresponde al correcto

Tema1: Tema1

Pregunta	Preguntas	CertezaPregunta
111	111 ; 225	0,517 ; 0,564
112	112 ; 114 ; 116	0,819 ; 0,680 ; 0,680
113	113 .	0,517 .
114	112 ; 114 ; 116	0,680 ; 0,819 ; 0,680
115	No identifica el tema correcto y sí otros	
116	112 ; 114 ; 115 ; 116 ; 575 ; 584	0,584 ; 0,584 ; 0,517 ; 0,680 ; 0,564 ; 0,680
121	121 .	0,819 .
122	No aparece pregunta	
123	112 ; 114 ; 116 ; 123 ; 343	0,584 ; 0,584 ; 0,584 ; 0,680 ; 0,600
131	131 .	0,517 .
132	132 .	0,680 .
133	112 ; 114 ; 116 ; 133 ; 523 ; 524 ; 526	0,517 ; 0,517 ; 0,517 ; 0,517 ; 0,680 ; 0,584 ; 0,584
141	141 .	0,517 .
142	142 .	0,517 .
143	143 .	0,517 .
144	144 .	0,819 .

145	145 ; 523	0,517 ; 0,517
146	146 .	0,680 .

Tema2: Tema2

Pregunta	Preguntas	CertezaPregunta
211	No aparece pregunta	
212	212 .	0,819 .
213	213 .	0,680 .
214	No aparece pregunta	
215	215 .	0,831 .
216	No aparece pregunta	
217	217 .	0,831 .
218	435 .	0,831 .
221	221 .	0,599 .
222	222 .	0,517 .
223	223 .	0,517 .
224	No aparece pregunta	
225	No aparece pregunta	
231	231 .	0,584 .
232	232 ; 411	0,680 ; 0,599
233	233 ; 412	0,819 ; 0,564
234	No aparece pregunta	
235	235 .	0,600 .
236	236 .	0,600 .
241	No aparece pregunta	
242	242 .	0,680 .
251	No aparece pregunta	
252	252 .	0,517 .
253	253 ; 412	0,646 ; 0,646
261	261 ; 332	0,646 ; 0,819
262	No aparece pregunta	
263	No aparece pregunta	
264	No aparece pregunta	
265	265 .	0,819 .
266	No aparece pregunta	
267	267 .	0,680 .
268	No aparece pregunta	
269	269 .	0,680 .
2610	No aparece pregunta	
271	No aparece pregunta	

272	No aparece pregunta	
273	No aparece pregunta	
274	No aparece pregunta	

Tema3: Tema3

Pregunta	Preguntas	CertezaPregunta
311	311	0,680
321	No aparece pregunta	
331	No aparece pregunta	
332	261 ; 332	0,646 ; 0,819
341	No aparece pregunta	
342	342	0,564
343	123 ; 343	0,646 ; 0,600

Tema4: Tema4

Pregunta	Preguntas	CertezaPregunta
411	411	0,819
412	233 ; 253 ; 412	0,680 ; 0,646 ; 0,680
421	421	0,819
422	422	0,819
431	431	0,680
432	No aparece pregunta	
433	433 ; 535	0,819 ; 0,819
434	434 ; 535	0,600 ; 0,831
435	435	0,680
436	436	0,842
437	437 ; 535	0,599 ; 0,680
438	No aparece pregunta	
441	441	0,842
442	442	0,646
443	No aparece pregunta	
451	451	0,599
452	No aparece pregunta	
453	453 ; 581	0,680 ; 0,600
454	454	0,600

Tema5: Tema5

Pregunta	Preguntas	CertezaPregunta
511	511 ; 512	0,646 ; 0,600
512	512 .	0,831 .
513	513 .	0,517 .
521	Falla, la certeza sale menor de 0,5	
522	522 .	0,584 .
523	521 ; 523 ; 524 ; 525 ; 526	0,517 ; 0,819 ; 0,680 ; 0,517 ; 0,584
524	521 ; 523 ; 524 ; 525 ; 526	0,517 ; 0,680 ; 0,680 ; 0,517 ; 0,584
525	523 ; 524 ; 525 ; 526	0,599 ; 0,599 ; 0,819 ; 0,517
526	521 ; 523 ; 524 ; 525 ; 526	0,517 ; 0,680 ; 0,680 ; 0,517 ; 0,680
527	523 ; 524 ; 527	0,517 ; 0,517 ; 0,831
528	528 .	0,680 .
529	529 .	0,599 .
531	531 ; 532 ; 533	0,819 ; 0,599 ; 0,599
532	531 ; 532	0,646 ; 0,819
533	531 ; 532 ; 533	0,819 ; 0,599 ; 0,819
534	531 ; 534	0,599 ; 0,819
535	433 ; 535	0,819 ; 0,819
541	541 ; 542 ; 545	0,517 ; 0,517 ; 0,517
542	541 ; 542 ; 545	0,517 ; 0,680 ; 0,517
543	545 .	0,517 .
544	544 .	0,600 .
545	544 ; 545	0,600 ; 0,599
551	Falla, la certeza sale menor de 0,5	
552		
553	553 .	0,599 .
554	553 .	0,584 .
561	561 .	0,842 .
562	562 .	0,533 .
571	571 ; 572 ; 573	0,584 ; 0,584 ; 0,584
572	571 ; 572 ; 573	0,584 ; 0,584 ; 0,584
573	571 ; 572 ; 573	0,584 ; 0,584 ; 0,680
574	269 ; 574	0,600 ; 0,831
575	575 .	0,819 .
581	453 ; 581	0,680 ; 0,819
582	582 .	0,646 .
583	583 .	0,842 .
584	584 .	0,680 .
585	585 .	0,600 .

591	Falla, la certeza del tema sale menor de 0,5	
592	Falla, la certeza de la pregunta sale menor de 0,5	
593	593 .	0,517 .
594	594 ; 595	0,842 ; 0,842
595	594 ; 595	0,842 ; 0,842
596	596 .	0,600 .
5101	Identifica apartado, la pregunta tiene certeza menor de 0.5	
5102	Identifica apartado, la pregunta tiene certeza menor de 0.5	
5103	Identifica apartado, la pregunta tiene certeza menor de 0.5	
5104	Identifica apartado, la pregunta tiene certeza menor de 0.5	
5105	Identifica apartado, la pregunta tiene certeza menor de 0.5	
5106	Identifica apartado, la pregunta tiene certeza menor de 0.5	
5210	5210 .	0,517 .

5.- Motor de 5 entradas con umbrales variables

En color  los valores que cambian respecto al motor 5E fijo **MOTOR 5 ENTRADAS con UMBRAL VARIABLE**

Cuando no hay ningun resultado que supere el umbral, se baja en 0,5 dejando los demás igual

Pregunta	Acierta Preg	Acierta Partido 1	Acierta Partido 2	Acierta Partido 3	Falla Partido	Nuevos Umbrales	Preguntas Identificadas	Certeza	

111		Si							
112	Si								
113	Si								
114	Si								
115					Si				
116		Si							
121	Si								
122	Si					0,45/0,5/0,5	122 .	0,517 .	
123	Si								
131	Si								
132	Si								
133					Si				
141	Si								
142	Si								
143	Si								
144	Si								
145		Si							
146	Si								
211	Si					0,5/0,35/0,30	211 .	0,331 .	
212	Si								
213	Si								
214					Si				
215	Si								
216	Si					0,5/0,35/0,5	216 .	0,599 .	
217	Si								
218	Si					0,5/0,45/0,5	218 .	0,646 .	
221	Si								
222	Si								
223	Si								
224		Si				0,5/0,35/0,4	224 ; 234	0,400 ; 0,400	
225					Si				
231	Si								
232	Si								
233	Si								
234		Si				0,5/0,35/0,4	224 ; 234	0,400 ; 0,400	
235	Si								
236	Si								
241					Si				
242	Si								
251	Si					0,5/0,35/0,4	251 .	0,400 .	

252	Si							
253		Si						
261		Si						
262	Si				0,5/0,5/0,45	262 .	0,482 .	
263					Si	Parece cosa de índices		
264	Si				0,5/0,5/0,35	264 .	0,400 .	
265	Si							
266	Si				0,5/0,4/0,4	266 .	0,400 .	
267	Si							
268	Si				0,35/0,5/0,5	268 .	0,517 .	
269	Si							
2610					Si	No encuentra Tema2		
271	Si				0,5/0,45/0,45	271 .	0,482 .	
272	Si				0,5/0,40/0,45	272 .	0,482 .	
273	Si				0,5/0,45/0,45	273 .	0,482 .	
274		Si			0,45/0,45/0,45	262 ; 274	0,842 ;	
						5	0,842	
311	Si							
321	Si				0,5/0,45/0,45	321 .	0,482 .	
331					Si			
332		Si						
341					Si	Sale la 342		
342	Si							
343		Si						
411	Si							
412		Si						
421	Si							
422	Si							
431	Si							
432	Si				0,5/0,45/0,5	432 .	0,564 .	
433		Si						
434		Si						
435	Si							
436	Si							
437		Si						
438	Si				0,5/0,40/0,5	438 .	0,517 .	
441	Si							
442	Si							
443	Si				0,5/0,45/0,5	443 .	0,646 .	
451	Si							
452					Si	No identifica bien el tema		

453	Si								
454	Si								
511	Si								
512	Si								
513	Si								
521	Si					0,5/0,45/0,5	521 .	0,584 .	
522	Si								
523	Si								
524		Si							
525	Si								
526			Si						
527	Si								
528	Si								
529	Si								
5210	Si								
531	Si								
532	Si								
533		Si							
534	Si								
535		Si							
541			Si						
542	Si								
543						Si	Identifica 545. Parece problema de índices		
544	Si								
545	Si								
551	Si					0,5/0,5/0,25	551 .	0,292 .	
552	Si					0,5/0,5/0,5	552 .	0,842 .	
553	Si								
554	Si								
561	Si								
562	Si								
571			Si						
572			Si						
573	Si								
574	Si								
575	Si								
581	Si								
582	Si								
583	Si								
584	Si								
585	Si								

591	Si				0,4/0,4/0,4	591 .	0,400 .	
592	Si				0,5/0,5/0,45	592 .	0,482 .	
593	Si							
594		Si						
595		Si						
596	Si							
5101					Si	Sale la 264. Puede ser un problema de índices		
5102					Si	Sale la 264. Puede ser un problema de índices		
5103					Si	Sale la 264. Puede ser un problema de índices		
5104					Si	Sale la 264. Puede ser un problema de índices		
5105					Si	No sale nada ni poniendo el umbral de la pregunta a 0		
5106					Si	No sale nada ni poniendo el umbral de la pregunta a 0		

6. Ejemplos de informes de las pruebas realizadas.

6.1.- Informe de la consulta pregunta-tipo 122. Motor de 3 entradas con umbrales fijos.

La consulta realizada es "¿Quiénes forman el equipo de Gobierno de la Universidad de Sevilla?"

Las palabras identificadas como pertenecientes a contenido son:

Palabra1: gobierno; Concreción: 0,5

Palabra2: universidad; Concreción: 0,300000011920929

Palabra3: sevilla; Concreción: 0,300000011920929

Las palabras seleccionadas para evaluar los temas son:

Palabra1: gobierno; Concreción: 0,5

Palabra2: universidad; Concreción: 0,300000011920929

Palabra3: sevilla; Concreción: 0,300000011920929

Los índices de certeza de pertenencia a cada Tema son:

Indice de Certeza del Tema 1: 0,646807849407196
Indice de Certeza del Tema 2: 0,517307579517364
Indice de Certeza del Tema 3: 0,331999927759171
Indice de Certeza del Tema 4: 0,331999927759171
Indice de Certeza del Tema 5: 0,331999927759171

El umbral de Certeza para aceptar el tema como posible es:
0,5

Las palabras seleccionadas para evaluar el Apartado 1 del Tema 1 son:
gobierno, universidad, sevilla
Los índices de pertenencia al apdo son: 0, 0,400000005960464, 0,300000011920929
La certeza de pertenencia al Apartado 1 del Tema 1 es de 0,517307579517364
Las palabras seleccionadas para evaluar el Apartado 2 del Tema 1 son:
gobierno, universidad, sevilla
Los índices de pertenencia al apdo son: 0,800000011920929, 0,400000005960464,
0,300000011920929
La certeza de pertenencia al Apartado 2 del Tema 1 es de 0,819387793540955
Las palabras seleccionadas para evaluar el Apartado 3 del Tema 1 son:
gobierno, universidad, sevilla
Los índices de pertenencia al apdo son: 0, 0,400000005960464, 0,300000011920929
La certeza de pertenencia al Apartado 3 del Tema 1 es de 0,517307579517364
Las palabras seleccionadas para evaluar el Apartado 4 del Tema 1 son:
gobierno, universidad, sevilla
Los índices de pertenencia al apdo son: 0, 0,400000005960464, 0,300000011920929
La certeza de pertenencia al Apartado 4 del Tema 1 es de 0,517307579517364
Las palabras seleccionadas para evaluar el Apartado 1 del Tema 2 son:
universidad, sevilla, portal
Los índices de pertenencia al apdo son: 0,400000005960464, 0,200000002980232, 0
La certeza de pertenencia al Apartado 1 del Tema 2 es de 0,400000005960464
Las palabras seleccionadas para evaluar el Apartado 2 del Tema 2 son:
universidad, sevilla, portal
Los índices de pertenencia al apdo son: 0,200000002980232, 0,699999988079071,
0,200000002980232
La certeza de pertenencia al Apartado 2 del Tema 2 es de 0,531895160675049
Las palabras seleccionadas para evaluar el Apartado 3 del Tema 2 son:
universidad, sevilla, portal
Los índices de pertenencia al apdo son: 0,300000011920929, 0,200000002980232, 0
La certeza de pertenencia al Apartado 3 del Tema 2 es de 0,331999927759171
Las palabras seleccionadas para evaluar el Apartado 4 del Tema 2 son:
universidad, sevilla, portal
Los índices de pertenencia al apdo son: 0,300000011920929, 0, 0
La certeza de pertenencia al Apartado 4 del Tema 2 es de 0,331999927759171

Las palabras seleccionadas para evaluar el Apartado 5 del Tema 2 son:
universidad, sevilla, portal

Los índices de pertenencia al apdo son: 0,200000002980232, 0,200000002980232, 0,300000011920929

La certeza de pertenencia al Apartado 5 del Tema 2 es de 0,331999927759171

Las palabras seleccionadas para evaluar el Apartado 6 del Tema 2 son:
universidad, sevilla, portal

Los índices de pertenencia al apdo son: 0,300000011920929, 0,200000002980232, 0

La certeza de pertenencia al Apartado 6 del Tema 2 es de 0,331999927759171

Las palabras seleccionadas para evaluar el Apartado 7 del Tema 2 son:
universidad, sevilla, portal

Los índices de pertenencia al apdo son: 0, 0,100000001490116, 0

La certeza de pertenencia al Apartado 7 del Tema 2 es de 0,125757575035095

El umbral de consideración de un apartado se fija para una certeza superior a 0,5

Las palabras seleccionadas para determinar las preguntas probables son:

Del Tema 1:

Del Apartado 1:

Palabra 1: universidad

Palabra 2: sevilla

Palabra 3:

Del Apartado 2:

Palabra 1: gobierno

Palabra 2: universidad

Palabra 3: sevilla

Del Apartado 3:

Palabra 1: universidad

Palabra 2: sevilla

Palabra 3:

Del Apartado 4:

Palabra 1: universidad

Palabra 2: sevilla

Palabra 3:

Del Tema 2:

Del Apartado 2:

Palabra 1: universidad

Palabra 2: sevilla

Palabra 3: servicio

El umbral de certeza para mostrar la pregunta como posible es de: 0,5

Las preguntas posibles son:

Del Tema 1:

Del Apartado 1:

La consulta probablemente no corresponda a ninguna pregunta de este apartado

Del Apartado 2:

Pregunta 2 con un nivel de certeza del 0,819387793540955

Del Apartado 3:

La consulta probablemente no corresponda a ninguna pregunta de este apartado

Del Apartado 4:

La consulta probablemente no corresponda a ninguna pregunta de este apartado

Del Tema 2:

Del Apartado 2:

Pregunta 5 con un nivel de certeza del 0,646807849407196

6.2.- Informe de la consulta pregunta-tipo 122. Motor de 5 entradas con umbrales fijos.

La consulta realizada es "¿Quiénes forman el equipo de Gobierno de la Universidad de Sevilla?"

Las palabras identificadas como pertenecientes a contenido son:

Palabra1: gobierno; Concreción: 0,5

Palabra2: universidad; Concreción: 0,300000011920929

Palabra3: sevilla; Concreción: 0,300000011920929

Las palabras seleccionadas para evaluar los temas son:

Palabra1: gobierno; Concreción: 0,5

Palabra2: universidad; Concreción: 0,300000011920929

Palabra3: sevilla; Concreción: 0,300000011920929

Palabra4: ; Concreción: 0

Palabra5: ; Concreción: 0

Los índices de certeza de pertenencia a cada Tema son:

Índice de Certeza del Tema 1: 0,482691496610641

Índice de Certeza del Tema 2: 0,331999927759171

Indice de Certeza del Tema 3: 0,155263155698776
Indice de Certeza del Tema 4: 0,155263155698776
Indice de Certeza del Tema 5: 0,155263155698776

El umbral de Certeza para aceptar el tema como posible es:
0,5

El umbral de consideración de un apartado se fija para una certeza superior a 0,5

Las palabras seleccionadas para determinar las preguntas probables son:

El umbral de certeza para mostrar la pregunta como posible es de: 0,5

Las preguntas posibles son:

6.3. Informe de la consulta pregunta-tipo 112. Motor de 5 entradas con umbrales variables.

La consulta realizada es "¿Quiénes forman el equipo de Gobierno de la Universidad de Sevilla?"

Las palabras identificadas como pertenecientes a contenido son:

Palabra1: gobierno; Concreción: 0,5
Palabra2: universidad; Concreción: 0,300000011920929
Palabra3: sevilla; Concreción: 0,300000011920929

Las palabras seleccionadas para evaluar los temas son:

Palabra1: gobierno; Concreción: 0,5
Palabra2: universidad; Concreción: 0,300000011920929
Palabra3: sevilla; Concreción: 0,300000011920929
Palabra4: ; Concreción: 0
Palabra5: ; Concreción: 0

Los índices de certeza de pertenencia a cada Tema son:

Indice de Certeza del Tema 1: 0,482691496610641
Indice de Certeza del Tema 2: 0,331999927759171
Indice de Certeza del Tema 3: 0,155263155698776
Indice de Certeza del Tema 4: 0,155263155698776
Indice de Certeza del Tema 5: 0,155263155698776

El umbral de Certeza para aceptar el tema como posible es:
0,449999988079071

Las palabras seleccionadas para evaluar el Apartado 1 del Tema 1 son:
gobierno, universidad, sevilla

,
Los índices de pertenencia al apdo son: 0, 0,400000005960464, 0,300000011920929
0,400000005960464, 0,300000011920929

La certeza de pertenencia al Apartado 1 del Tema 1 es de 0,331999927759171

Las palabras seleccionadas para evaluar el Apartado 2 del Tema 1 son:
gobierno, universidad, sevilla

,
Los índices de pertenencia al apdo son: 0,800000011920929, 0,400000005960464,
0,300000011920929

0,400000005960464, 0,300000011920929

La certeza de pertenencia al Apartado 2 del Tema 1 es de 0,680232465267181

Las palabras seleccionadas para evaluar el Apartado 3 del Tema 1 son:
gobierno, universidad, sevilla

,
Los índices de pertenencia al apdo son: 0, 0,400000005960464, 0,300000011920929
0,400000005960464, 0,300000011920929

La certeza de pertenencia al Apartado 3 del Tema 1 es de 0,331999927759171

Las palabras seleccionadas para evaluar el Apartado 4 del Tema 1 son:
gobierno, universidad, sevilla

,
Los índices de pertenencia al apdo son: 0, 0,400000005960464, 0,300000011920929
0,400000005960464, 0,300000011920929

La certeza de pertenencia al Apartado 4 del Tema 1 es de 0,331999927759171

El umbral de consideración de un apartado se fija para una certeza superior a 0,5

Las palabras seleccionadas para determinar las preguntas probables son:

Del Tema 1:

Del Apartado 2:

Palabra 1: gobierno
Palabra 2: universidad
Palabra 3: sevilla
Palabra 4:
Palabra 5:

El umbral de certeza para mostrar la pregunta como posible es de: 0,5

Las preguntas posibles son:

Del Tema 1:
Del Apartado 2:
Pregunta 2 con un nivel de certeza del 0,517307579517364

6.4. Informe de la consulta pregunta-tipo 521. Motor de 3 entradas con umbrales fijos.

La consulta realizada es "¿Cómo puedo solicitar alojamiento para una página WEB institucional? (centros, departamentos, grupos de investigación, etc.)"

Las palabras identificadas como pertenecientes a contenido son:

Palabra1: solicitar; Concreción: 0,400000005960464
Palabra2: alojamiento; Concreción: 0,699999988079071
Palabra3: web; Concreción: 0,5
Palabra4: institucional; Concreción: 0,400000005960464
Palabra5: centros; Concreción: 0,699999988079071
Palabra6: departamentos; Concreción: 0,899999976158142
Palabra7: grupos; Concreción: 0,400000005960464
Palabra8: investigacion; Concreción: 0,699999988079071

Las palabras seleccionadas para evaluar los temas son:

Palabra1: centros; Concreción: 0,699999988079071
Palabra2: alojamiento; Concreción: 0,699999988079071
Palabra3: departamentos; Concreción: 0,899999976158142

Los índices de certeza de pertenencia a cada Tema son:

Indice de Certeza del Tema 1: 0,831817984580994
Indice de Certeza del Tema 2: 0,831817984580994
Indice de Certeza del Tema 3: 0,116666659712791

Indice de Certeza del Tema 4: 0,125757575035095

Indice de Certeza del Tema 5: 0,819387972354889

El umbral de Certeza para aceptar el tema como posible es:

0,5

Las palabras seleccionadas para evaluar el Apartado 1 del Tema 1 son:

web, ,

Los índices de pertenencia al apdo son: 0,5, 0, 0

La certeza de pertenencia al Apartado 1 del Tema 1 es de 0,400000035762787

Las palabras seleccionadas para evaluar el Apartado 2 del Tema 1 son:

web, ,

Los índices de pertenencia al apdo son: 0, 0, 0

La certeza de pertenencia al Apartado 2 del Tema 1 es de 0,11666659712791

Las palabras seleccionadas para evaluar el Apartado 3 del Tema 1 son:

web, ,

Los índices de pertenencia al apdo son: 0,400000005960464, 0, 0

La certeza de pertenencia al Apartado 3 del Tema 1 es de 0,400000005960464

Las palabras seleccionadas para evaluar el Apartado 4 del Tema 1 son:

web, ,

Los índices de pertenencia al apdo son: 0,300000011920929, 0, 0

La certeza de pertenencia al Apartado 4 del Tema 1 es de 0,331999927759171

Las palabras seleccionadas para evaluar el Apartado 1 del Tema 2 son:

alojamiento, centros, departamentos

Los índices de pertenencia al apdo son: 0, 0, 0

La certeza de pertenencia al Apartado 1 del Tema 2 es de 0,11666659712791

Las palabras seleccionadas para evaluar el Apartado 2 del Tema 2 son:

alojamiento, centros, departamentos

Los índices de pertenencia al apdo son: 0,800000011920929, 0, 0

La certeza de pertenencia al Apartado 2 del Tema 2 es de 0,831817984580994

Las palabras seleccionadas para evaluar el Apartado 3 del Tema 2 son:

alojamiento, centros, departamentos

Los índices de pertenencia al apdo son: 0,800000011920929, 0, 0

La certeza de pertenencia al Apartado 3 del Tema 2 es de 0,831817984580994

Las palabras seleccionadas para evaluar el Apartado 4 del Tema 2 son:

alojamiento, centros, departamentos

Los índices de pertenencia al apdo son: 0, 0, 0

La certeza de pertenencia al Apartado 4 del Tema 2 es de 0,11666659712791

Las palabras seleccionadas para evaluar el Apartado 5 del Tema 2 son:

alojamiento, centros, departamentos

Los índices de pertenencia al apdo son: 0, 0, 0

La certeza de pertenencia al Apartado 5 del Tema 2 es de 0,11666659712791

Las palabras seleccionadas para evaluar el Apartado 6 del Tema 2 son:

alojamiento, centros, departamentos
Los índices de pertenencia al apdo son: 0, 0,5, 0,400000005960464
La certeza de pertenencia al Apartado 6 del Tema 2 es de 0,599999964237213
Las palabras seleccionadas para evaluar el Apartado 7 del Tema 2 son:
alojamiento, centros, departamentos
Los índices de pertenencia al apdo son: 0, 0, 0
La certeza de pertenencia al Apartado 7 del Tema 2 es de 0,116666659712791
Las palabras seleccionadas para evaluar el Apartado 1 del Tema 5 son:
alojamiento, web, grupos
Los índices de pertenencia al apdo son: 0, 0, 0
La certeza de pertenencia al Apartado 1 del Tema 5 es de 0,116666659712791
Las palabras seleccionadas para evaluar el Apartado 2 del Tema 5 son:
alojamiento, web, grupos
Los índices de pertenencia al apdo son: 0,699999988079071, 0,5, 0,200000002980232
La certeza de pertenencia al Apartado 2 del Tema 5 es de 0,599999904632568
Las palabras seleccionadas para evaluar el Apartado 3 del Tema 5 son:
alojamiento, web, grupos
Los índices de pertenencia al apdo son: 0, 0, 0
La certeza de pertenencia al Apartado 3 del Tema 5 es de 0,116666659712791
Las palabras seleccionadas para evaluar el Apartado 4 del Tema 5 son:
alojamiento, web, grupos
Los índices de pertenencia al apdo son: 0, 0, 0
La certeza de pertenencia al Apartado 4 del Tema 5 es de 0,116666659712791
Las palabras seleccionadas para evaluar el Apartado 5 del Tema 5 son:
alojamiento, web, grupos
Los índices de pertenencia al apdo son: 0, 0, 0
La certeza de pertenencia al Apartado 5 del Tema 5 es de 0,116666659712791
Las palabras seleccionadas para evaluar el Apartado 6 del Tema 5 son:
alojamiento, web, grupos
Los índices de pertenencia al apdo son: 0, 0, 0
La certeza de pertenencia al Apartado 6 del Tema 5 es de 0,116666659712791
Las palabras seleccionadas para evaluar el Apartado 7 del Tema 5 son:
alojamiento, web, grupos
Los índices de pertenencia al apdo son: 0, 0, 0
La certeza de pertenencia al Apartado 7 del Tema 5 es de 0,116666659712791
Las palabras seleccionadas para evaluar el Apartado 8 del Tema 5 son:
alojamiento, web, grupos
Los índices de pertenencia al apdo son: 0, 0, 0
La certeza de pertenencia al Apartado 8 del Tema 5 es de 0,116666659712791
Las palabras seleccionadas para evaluar el Apartado 9 del Tema 5 son:
alojamiento, web, grupos
Los índices de pertenencia al apdo son: 0,400000005960464, 0, 0
La certeza de pertenencia al Apartado 9 del Tema 5 es de 0,400000005960464
Las palabras seleccionadas para evaluar el Apartado 10 del Tema 5 son:

alojamiento, web, grupos

Los índices de pertenencia al apdo son: 0, 0, 0

La certeza de pertenencia al Apartado 10 del Tema 5 es de 0,116666659712791

El umbral de consideración de un apartado se fija para una certeza superior a 0,5

Las palabras seleccionadas para determinar las preguntas probables son:

Del Tema 1:

Del Tema 2:

Del Apartado 2:

Palabra 1: alojamiento

Palabra 2:

Palabra 3:

Del Apartado 3:

Palabra 1: alojamiento

Palabra 2:

Palabra 3:

Del Apartado 6:

Palabra 1: centros

Palabra 2: departamentos

Palabra 3:

Del Tema 5:

Del Apartado 2:

Palabra 1: alojamiento

Palabra 2: web

Palabra 3: institucional

El umbral de certeza para mostrar la pregunta como posible es de: 0,5

Las preguntas posibles son:

Del Tema 1:

Del Tema 2:

Del Apartado 2:

Pregunta 4 con un nivel de certeza del 0,846296191215515

Del Apartado 3:

Pregunta 4 con un nivel de certeza del 0,846296191215515
Del Apartado 6:
Pregunta 9 con un nivel de certeza del 0,600000023841858

Del Tema 5:
Del Apartado 2:
Pregunta 1 con un nivel de certeza del 0,599999904632568
Pregunta 3 con un nivel de certeza del 0,517307579517364
Pregunta 4 con un nivel de certeza del 0,517307579517364
Pregunta 6 con un nivel de certeza del 0,517307579517364

6.5. Informe de la consulta pregunta-tipo 521. Motor de 5 entradas con umbrales variables.

La consulta realizada es "¿Cómo puedo solicitar alojamiento para una página WEB institucional? (centros, departamentos, grupos de investigación, etc.)"

Las palabras identificadas como pertenecientes a contenido son:

Palabra1: solicitar; Concreción: 0,400000005960464
Palabra2: alojamiento; Concreción: 0,699999988079071
Palabra3: web; Concreción: 0,5
Palabra4: institucional; Concreción: 0,400000005960464
Palabra5: centros; Concreción: 0,699999988079071
Palabra6: departamentos; Concreción: 0,899999976158142
Palabra7: grupos; Concreción: 0,400000005960464
Palabra8: investigacion; Concreción: 0,699999988079071

Las palabras seleccionadas para evaluar los temas son:

Palabra1: departamentos; Concreción: 0,899999976158142
Palabra2: alojamiento; Concreción: 0,699999988079071
Palabra3: web; Concreción: 0,5
Palabra4: investigacion; Concreción: 0,699999988079071
Palabra5: centros; Concreción: 0,699999988079071

Los índices de certeza de pertenencia a cada Tema son:

Indice de Certeza del Tema 1: 0,831817984580994
Indice de Certeza del Tema 2: 0,831817984580994
Indice de Certeza del Tema 3: 0,116666659712791
Indice de Certeza del Tema 4: 0,400000035762787
Indice de Certeza del Tema 5: 0,819387972354889

El umbral de Certeza para aceptar el tema como posible es:
0,5

Las palabras seleccionadas para evaluar el Apartado 1 del Tema 1 son:
web, ,

,
Los índices de pertenencia al apdo son: 0,5, 0, 0
0, 0

La certeza de pertenencia al Apartado 1 del Tema 1 es de 0,116666659712791

Las palabras seleccionadas para evaluar el Apartado 2 del Tema 1 son:
web, ,

,
Los índices de pertenencia al apdo son: 0, 0, 0
0, 0

La certeza de pertenencia al Apartado 2 del Tema 1 es de 0,116666659712791

Las palabras seleccionadas para evaluar el Apartado 3 del Tema 1 son:
web, ,

,
Los índices de pertenencia al apdo son: 0,400000005960464, 0, 0
0, 0

La certeza de pertenencia al Apartado 3 del Tema 1 es de 0,130645170807838

Las palabras seleccionadas para evaluar el Apartado 4 del Tema 1 son:
web, ,

,
Los índices de pertenencia al apdo son: 0,300000011920929, 0, 0
0, 0

La certeza de pertenencia al Apartado 4 del Tema 1 es de 0,155263155698776

Las palabras seleccionadas para evaluar el Apartado 1 del Tema 2 son:
alojamiento, centros, departamentos
sevilla,

Los índices de pertenencia al apdo son: 0, 0, 0
0, 0

La certeza de pertenencia al Apartado 1 del Tema 2 es de 0,125757575035095

Las palabras seleccionadas para evaluar el Apartado 2 del Tema 2 son:
alojamiento, centros, departamentos
sevilla,

Los índices de pertenencia al apdo son: 0,800000011920929, 0, 0
0, 0

La certeza de pertenencia al Apartado 2 del Tema 2 es de 0,400000035762787

Las palabras seleccionadas para evaluar el Apartado 3 del Tema 2 son:
alojamiento, centros, departamentos
sevilla,

Los índices de pertenencia al apdo son: 0,800000011920929, 0, 0

0, 0

La certeza de pertenencia al Apartado 3 del Tema 2 es de 0,400000035762787

Las palabras seleccionadas para evaluar el Apartado 4 del Tema 2 son:

alojamiento, centros, departamentos

sevilla,

Los índices de pertenencia al apdo son: 0, 0, 0

0, 0

La certeza de pertenencia al Apartado 4 del Tema 2 es de 0,125757575035095

Las palabras seleccionadas para evaluar el Apartado 5 del Tema 2 son:

alojamiento, centros, departamentos

sevilla,

Los índices de pertenencia al apdo son: 0, 0, 0

0, 0

La certeza de pertenencia al Apartado 5 del Tema 2 es de 0,125757575035095

Las palabras seleccionadas para evaluar el Apartado 6 del Tema 2 son:

alojamiento, centros, departamentos

sevilla,

Los índices de pertenencia al apdo son: 0, 0,5, 0,400000005960464

0,5, 0,400000005960464

La certeza de pertenencia al Apartado 6 del Tema 2 es de 0,517307579517364

Las palabras seleccionadas para evaluar el Apartado 7 del Tema 2 son:

alojamiento, centros, departamentos

sevilla,

Los índices de pertenencia al apdo son: 0, 0, 0

0, 0

La certeza de pertenencia al Apartado 7 del Tema 2 es de 0,11666659712791

Las palabras seleccionadas para evaluar el Apartado 1 del Tema 5 son:

alojamiento, web, grupos

investigacion, departamentos

Los índices de pertenencia al apdo son: 0, 0, 0

0, 0

La certeza de pertenencia al Apartado 1 del Tema 5 es de 0,11666659712791

Las palabras seleccionadas para evaluar el Apartado 2 del Tema 5 son:

alojamiento, web, grupos

investigacion, departamentos

Los índices de pertenencia al apdo son: 0,699999988079071, 0,5, 0,200000002980232

0,5, 0,200000002980232

La certeza de pertenencia al Apartado 2 del Tema 5 es de 0,482691496610641

Las palabras seleccionadas para evaluar el Apartado 3 del Tema 5 son:

alojamiento, web, grupos

investigacion, departamentos

Los índices de pertenencia al apdo son: 0, 0, 0

0, 0

La certeza de pertenencia al Apartado 3 del Tema 5 es de 0,11666659712791

Las palabras seleccionadas para evaluar el Apartado 4 del Tema 5 son:

alojamiento, web, grupos

investigacion, departamentos

Los índices de pertenencia al apdo son: 0, 0, 0

0, 0

La certeza de pertenencia al Apartado 4 del Tema 5 es de 0,116666659712791

Las palabras seleccionadas para evaluar el Apartado 5 del Tema 5 son:

alojamiento, web, grupos

investigacion, departamentos

Los índices de pertenencia al apdo son: 0, 0, 0

0, 0

La certeza de pertenencia al Apartado 5 del Tema 5 es de 0,116666659712791

Las palabras seleccionadas para evaluar el Apartado 6 del Tema 5 son:

alojamiento, web, grupos

investigacion, departamentos

Los índices de pertenencia al apdo son: 0, 0, 0

0, 0

La certeza de pertenencia al Apartado 6 del Tema 5 es de 0,116666659712791

Las palabras seleccionadas para evaluar el Apartado 7 del Tema 5 son:

alojamiento, web, grupos

investigacion, departamentos

Los índices de pertenencia al apdo son: 0, 0, 0

0, 0

La certeza de pertenencia al Apartado 7 del Tema 5 es de 0,116666659712791

Las palabras seleccionadas para evaluar el Apartado 8 del Tema 5 son:

alojamiento, web, grupos

investigacion, departamentos

Los índices de pertenencia al apdo son: 0, 0, 0

0, 0

La certeza de pertenencia al Apartado 8 del Tema 5 es de 0,116666659712791

Las palabras seleccionadas para evaluar el Apartado 9 del Tema 5 son:

alojamiento, web, grupos

investigacion, departamentos

Los índices de pertenencia al apdo son: 0,400000005960464, 0, 0

0, 0

La certeza de pertenencia al Apartado 9 del Tema 5 es de 0,130645170807838

Las palabras seleccionadas para evaluar el Apartado 10 del Tema 5 son:

alojamiento, web, grupos

investigacion, departamentos

Los índices de pertenencia al apdo son: 0, 0, 0

0, 0

La certeza de pertenencia al Apartado 10 del Tema 5 es de 0,116666659712791

El umbral de consideración de un apartado se fija para una certeza superior a 0,449999988079071

Las palabras seleccionadas para determinar las preguntas probables son:

Del Tema 1:

Del Tema 2:

Del Apartado 6:

Palabra 1: centros

Palabra 2: departamentos

Palabra 3: universidad

Palabra 4: sevilla

Palabra 5:

Del Tema 5:

Del Apartado 2:

Palabra 1: alojamiento

Palabra 2: web

Palabra 3: institucional

Palabra 4: centros

Palabra 5: departamentos

El umbral de certeza para mostrar la pregunta como posible es de: 0,5

Las preguntas posibles son:

Del Tema 1:

Del Tema 2:

Del Apartado 6:

La consulta probablemente no corresponda a ninguna pregunta de este apartado

Del Tema 5:

Del Apartado 2:

Pregunta 1 con un nivel de certeza del 0,584545314311981

APÉNDICE C. ARTÍCULO PRESENTADO A MELECON '06

Uno de las consecuencias primordiales de la investigación debe ser la presentación de artículos en congresos y revistas internacionales. Hasta el momento, y a pesar de que el trabajo aún no se encuentra en su madurez completa, se ha publicado ya un artículo y se sigue trabajando para la inclusión en otros congresos y revistas.

El artículo que se presenta a continuación se titula *A fuzzy logic system for classifying the contents of a database and searching consultations in natural language* (Un sistema de lógica borrosa para la clasificación de los contenidos de una base de datos y la búsqueda de consultas) y se presentará en la 13ª Conferencia Electrotécnica Mediterránea, organizada por el IEEE en Benalmádena (Málaga) y que tendrá lugar en Mayo de 2006.

Así mismo, se ha presentado otro artículo para la Conferencia IDEAL '06. Esta conferencia tendrá lugar en Septiembre de 2006 e incluye la publicación del artículo en la revista LNCS (Lecture Notes in Computer Science).

A fuzzy logic system for classifying the contents of a database and searching consultations in natural language

Ariel Gómez¹, Jorge Ropero¹ and Carlos León¹, Member, IEEE.

¹Department of Electronic Technology, University of Seville, Spain.

ariel@us.es, jropero@dte.us.es, cleon@us.es

Abstract—This paper presents a method for the classification of the contents in a database in order to answer to user consultations using natural language. Artificial Intelligence (AI) is used to relate these consultations to the database contents. The system is based on a fuzzy logic engine to take advantage of its so suitable properties for this application and is ideal for sets of accumulated knowledge that can be built in hierarchic levels by a tree structure. The eventual aim of this system is the implementation of a virtual web assistant for an internet portal.

Index Terms— Artificial Intelligence, Fuzzy Logic, database, searching.

I. MOTIVATIONS

The access to the contents of an extensive set of accumulated knowledge – a database, a summary of documents, web contents, etc – is becoming an important concern in the last decades. In some occasions, a user tries to get to them knowing that what he needs is out there but ignoring the exact denomination for what he is looking for and/or the suitable method to make the extraction of the desired knowledge. All these disadvantages meet increased when the user in question is not a habitual of the matter or there are ambiguous contents, bad organization or, simply, complex topics or a great amount of information difficult to manage.

In these cases the unsuccessful attempts can turn out to be frustrating for not using the exact term to make the consultation - a machine only will answer adequately if it is asked in an exact way -, and one can eventually end in a paradox: the less one knows the more difficult it is to find the answers.

In many cases the solution is to look for another person who is an expert on the topic. Actually, the demanded help is an interpreter who is able to generate a syntactically and semantically correct search that leads us to the obtaining of the desired answers. Consequently, the need of an assistant who

interprets the vague information we have arises, giving us concrete answers related to existing contents in its database. This should be based on an estimation of the certainty of the relation between what we have expressed in natural language and the contents stored its database.

To solve this, we develop a method of classification of the contents by means of the creation of a few indexes based on key words, and a method of consultation based on a fuzzy logic application provided with an interface that one may interact with in natural language. We propose then an application of the artificial intelligence (AI) based on the use of fuzzy logic.

II. FUZZY LOGIC

Fuzzy logic arises as response to the inflexibility of the classic binary logic. In this one, the middle term does not exist: either it is hot or it is cold. Fuzzy logic gives us the possibility of having intermediate grades: cool, fresh, warm... and, besides, by means of a set of functions, a degree of flexibility may be given to these epithets: what may be cool for a Sevillian might be mild for a Berliner.

Fuzzy logic turns out to be useful for:

- a) Dealing with uncertainty.
- b) Dealing with precise information held together with uncertainty.

Therefore, we have simultaneously precise information and uncertainty. Using fuzzy logic certain quantity of precision is sacrificed in favour of uncertainty with the hope of obtaining conclusions that, though more vague, are more robust. [1-6]

III. MODE OF OPERATION OF THE METHOD OF CLASSIFICATION AND INFERENCE

A. Classification of accumulated knowledge

We propose to group the knowledge in sets of different levels: the final content, it is to say, every element, belongs to an N level set. Several level N sets with some common features gather in groups forming another N-1 level set. Analogously, several N-1 level sets with some common features gather in other groups forming an N-2 level set, and so on up to coming to level 0 set (N-N), which represents the totality of the accumulated knowledge. The proposed classification is seen in figure one, clearly forming a tree structure.

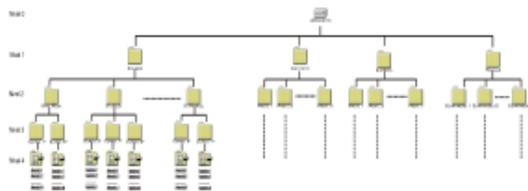


Figure 1 - Classification in a tree structure.

The first step of the processing consists of distinguishing the words of the consultation. These words must be searched in a database which must contain the words that are related somehow to the content of the matter we have to deal with. Another database with the possible answers is necessary.

Key words are assigned to every possible answer in order to identify it. These words are chosen among those that could appear in a possible consultation. The belonging of these words to every level is determined by means of a few numerical coefficients indicating how significant the considered word inside the level in question is. It is important to notice that the same word can belong to several different sets.

B. Example

To illustrate the process, the following example is used: a user of the University of Seville internet portal asks 'How can I find a professor e-mail address?' From here we would extract: find, professor, e-mail, and address. We could also add some related words: get - related -, professors - plural -, electronic - related - and mail - similar.

The following step is the assignment of the coefficients to every word. Sure that there are many questions relative to find something so this word does not have a very high coefficient;

there must be enough questions relating to professors, too; with regard to e-mail or address it is sure that there are fewer questions so that these words must have higher coefficients than others.

This way, coefficients for N level - level 3 in the example - might be:

- E-mail = mail = 0.7.
- Address = 0.6
- Professor = professors = 0.5.
- Find = get = 0.3.

Due to the use of fuzzy logic it is not necessary to be very precise with indexes, the system tends to be convergent and to identify the content. When identification tests for every anecdote are made all key words are introduced and it is decided whether it is necessary to change these coefficients or not.

Then we must group all the questions according to the section they belong to form level N-1 groups. The word indexes will not be the same but they will reflect how significant is this word in the following level. Likewise, the top level tables would be built up to coming to a level 0 table in which all the words would appear. This way the whole content of the database would be grouped in hierarchic levels, identified by a set of key words with an assigned index of importance which means how much this word identifies the content of the knowledge database as belonging to one of the levels.

As mentioned above, the aim of the system is to allow the user to formulate consultations by means of questions in natural language, to relate them to the contents using the index generated with fuzzy logic and to obtain the answers corresponding to the contents stored in the set of knowledge.

C. Fuzzy logic system

The inference method for finding out the knowledge demanded by the consultation is described in this section.

All the key words are extracted for comparison with the ones contained in our word database.

The belonging to every level 1 set is analyzed bearing in mind the value returned by the fuzzy engine. If the level of certainty is lower than a predefined value, the belonging of the content to the corresponding set is rejected. The facts of starting from level one and using a tree structure make possible the rejection of a great amount of content, which will not be considered in future searches.

For every set that has overcome the minimum certainty

threshold, the process is repeated and the coefficients of belonging corresponding to every level 2 set are evaluated. Again, the sets in which the returned by the fuzzy engine certainty does not overcome a certain minimal threshold are rejected. If they overcome the threshold, the method for determining the belonging to the following level is applied to them. This process is repeated up to coming to the last level. The answers correspond to those last level elements which certainly overcomes the definite threshold. There can be more than one answer. The vaguer are the questions, the more answers we will obtain.

The heart of the fuzzy logic system is the fuzzy engine. This engine is the responsible of determining the probability for the key words contained in the consultation of belonging to a certain fuzzy set in a concrete level. The engine must evaluate the belonging to every set for the corresponding level. For that reason, the engine takes the coefficients of the key words for that set as inputs. The fuzzy engine output will be determined by the defined rules. These rules are of the IF ... THEN type. An example of rule might be the following one:

IF word_index1 is HIGH AND word_index2 is MEDIUM AND word_index3 is LOW, THEN output is HIGH.

The definition of these rules corresponds to the system administrator.

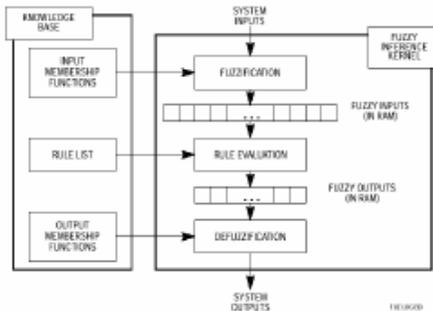


Figure 2 - Fuzzy logic system

D. System administrator

In previous sections the need of a system administrator has been mentioned. The functions of this administrator must be basically three.

- a) Defining and modifying the coefficients of belonging to a topic, paragraph or answer for a certain word.
- b) Adding new words to the database when necessary.
- c) Making a system feedback asking the eventual users their opinion about the answers given by the assistant in order to

take the necessary steps in every case.

IV. TESTS AND RESULTS

A. System simulation

To make the first functioning tests for the proposed method, we used a knowledge database consisting of a set of the more frequent questions belonging to the administrator system. Concretely, the analyzed case considers 133 questions. This allows to study the functioning of the method on a population big enough to get results which are reliable and comfortable to handle at the same time.

Knowledge is structured in three levels: Level 1 is correspondent to the Topic to which the question belongs; level 2 corresponds to a Section inside every topic - the number of Sections for each Topic varies between 3 and 10 - ; finally, level 3 corresponds to every concrete Question inside every Section in every Topic - the number of Questions for each Section also varies between 1 and 10).

For every Question, between 3 and 6 key words are defined. Nevertheless, the user will include in his consultation from 1 to 6 of these words. Defining an only engine with a few inputs causes the rapid saturation of the system, whereas to define an only six input engine causes values with a very low degree of certainty. The solution provided is to implement a flexible system with variable inputs. If the user consultation three key words as much, a three input fuzzy engine is used, whereas if the user consultation includes four or more key words, the system will use a five input fuzzy engine.

As said above, rule definition corresponds to the administrator. Logically, the more inputs the engine has, the more rules there are. For example, for a three inputs engine, the inputs can take three values: LOW, MEDIUM and HIGH. The outputs can take the values LOW, MEDIUM-LOW, MEDIUM-HIGH and HIGH. The inference rules defined are:

If all inputs are LOW, output is LOW.

If one input is MEDIUM and the others are LOW, output is MEDIUM-LOW.

If two inputs are MEDIUM and the other are LOW, output is MEDIUM-HIGH.

If all the input are MEDIUM or one input is HIGH, output is HIGH.

The possible combinations generate 27 rules for the fuzzy engine. A five input engine generates 243 rules.

All the obtained results are shown in the following figures.

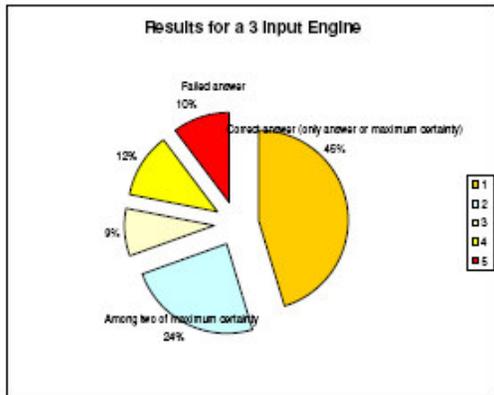


Figure 3 - Results for a 3 input engine

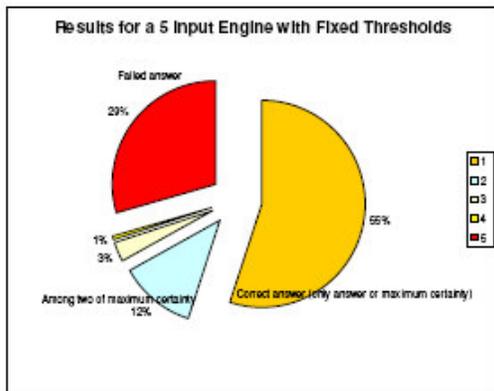


Figure 4 - Results for a 5 input engine with fixed thresholds.

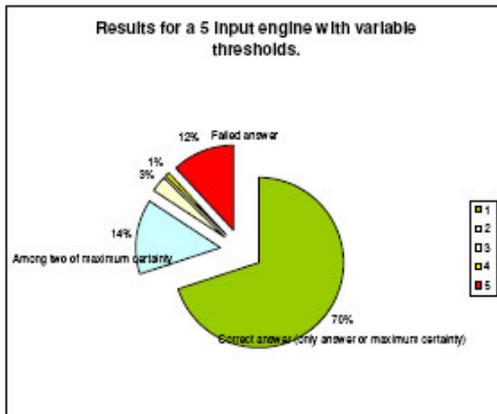


Figure 5 - Results for a 5 input engine with variable thresholds.

V. CONCLUSIONS AND FUTURE LINES OF WORK

Up to now, the obtained results using fuzzy logic are good enough, as the number of correctly detected consultations is high.

Future lines of work must begin by a possible improvement of the system based on tests using different parameters for the fuzzy logic system and the execution of new tests where the consultations have a different structure than those found in the database. Likewise, other words must be added to the database which contains the related words. This way, we will be able to consider synonymous or expressions that are similar to the standard one.

Another line of investigation is the creation of the user interface. Up to now we have only developed a version for the administrator, which can be seen in the following figure.



Figure 6 - Administrator version interface

REFERENCES

- [1] B. Martín del Brío, A. Sanz Molina, *Redes neuronales y sistemas borrosos* Ra-Ma, 2001.
- [2] *Fuzzy Logic Toolbox. User's guide*. The Mathworks Inc., 2002.
- [3] T. Bouazziz, A. Wolski, *Applying Fuzzy Events to Approximate Reasoning in Active Databases*. Proc. Sixth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE97). July 1-5, Barcelona, Spain.
- [4] S. A. Moriello. *Inselecos ultrarracionales*. ISSN: 1597-0223, 2004
- [5] D. Xie. *Fuzzy Association Rules Discovered on Effective Reduced Database Algorithm*. FUZZ-IEEE 2005, (The IEEE International Conference on Fuzzy Systems), Reno, USA, May, 2005.
- [6] P. Bedi, H. Kaur, A. Malhotra, *Fuzzy dimension to databases*. 37th National Convention of Computer Society of India, Bangalore, India, November 2002.
- [7] D.Olsen. *Fuzzy Logic Control in Autonomous Robotic*. University of Minnesota. Research Project. October, 2002

Bibliografía

- [1]. K.L. Kwok. *A neural network for probabilistic information retrieval*. Proceedings of the 12th annual international ACM SIGIR conference on Research and development in information retrieval. Cambridge, Massachusetts, United States. 1989.
- [2]. Alexandros Ntoulas, Gerald Chao, Junghoo Cho. *The Infocious web search engine: improving web searching through linguistic analysis*. In Proceedings of the World-Wide Web Conference (WWW), May 2005.
- [3]. G. Petasis, V. Karkaletsis, D. Farmakiotou, I. Androutsopoulos_, and C. D. Spyropoulo. *A Greek Morphological Lexicon and Its Exploitation by Natural Language Processing Applications*. PCI 2001, LNCS 2563, pp. 401–419, 2003._c Springer-Verlag Berlin Heidelberg 2003.
- [4]. J.J. Paijmans. *Applications of Natural Language Processing to Information Retrieval*. <http://pi0959.uvt.nl/Paai/Onderw/V-I/Content/nlir2002.html>
- [5]. K. Sparck-Jones. *What Is The Role for NLP in Text Retrieval*. T. Strzalkowski (ed.). Natural Language Information Retrieval. Kluwer. pp. 1-25. 1999.
- [6]. A.R. Aronson, T.C. Rindflesch. *Query Expansion Using the UMLS Metathesaurus*. Journal of the American Medical Informatics Association, 1997: Supplement. Proceedings of the 1997 AMIA Annual Fall Symposium, p. 485-489.
- [7]. *Wikipedia, the Free Encyclopedia*. <http://en.wikipedia.org/wiki>
- [8]. A.R. Aronson, T.C. Rindflesch, A.C. Browne. *Exploiting a large thesaurus for information retrieval*. Proceedings of RIAO, 197-216. 1994.
- [9]. THOR Center for Neuroinformatics. *Vector Space Model*. <http://isp.imm.dtu.dk/thor/projects/multimedia/textmining/node5.html>, 1998.
- [10]. G. Salton. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.
- [11]. C.J. van Rijsbergen, . *Information retrieval*. Butterworths, 1979.
- [12]. S. Chakrabarti, B. Dom, R. Agrawal, P. Raghavan. *Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies*. VLDB Journal Vol.7 3, p. 163-178, 1998.
- [13]. A. Bookstein, S.T. Klein, T. Raita. *Detecting content bearing words by serial clustering*. SIGIR Forum (ACM Special Interest Group on Information Retrieval), p. 319-327, 1995.
- [14]. G. Salton, C. Buckley. *Term Weighting Approaches in Automatic Text Retrieval*. Technical Report TR87-881, Department of Computer Science, Cornell University, 1987. Information Processing and Management Vol.32 (4), p. 431-443, 1996.
- [15]. G. Salton. *Automatic Text Processing*. Addison-Wesley Publishing Company, 1988.
- [16]. *TREC: Text REtrieval Conference*. <http://trec.nist.gov/overview.html>.
- [17]. D. Harman. *How effective is suffixing?* Journal of the American Society for Information Science, 42(1): pp 7-15, 1991.
- [18]. M. Popovic, P. Willett. *The effectiveness of stemming for natural language access to Slovene textual data*. Journal of the American Society for Information Science, 43(5): pp. 384-390, 1992.

- [19]. R. Krovetz. *Viewing morphology as an inference process*. Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1993.
- [20]. D. W. Oard. *Adaptive Vector Space Text Filtering for Monolingual and Cross-Language Applications*. PhD thesis, University of Maryland, College Park, August 1996. http://www.ee.umd.edu/medlab/filter/filter_project.html
- [21]. T.E. Doszkocs. *A large semantic network for associative searching*. Proceedings of The 5th International Conference on New Information Technology HongKong University of Science & Technology Kowloon, Hongkong. Nov. 30-Dec. 2, 1992.
- [22]. R.K. Belew. *Adaptive Information Retrieval: Using a Connectionist Representation to retrieve and learn about documents*. Proceedings of 12 th ACM-SIGIR Conference, pp. 11-20, Cambridge, Mass., USA. 1989
- [23]. M.E. Ruiz, P. Srinivasan. *Automatic Text Categorization Using Neural Networks*. Advances in Classification Research vol. 8: Proceedings of the 8th ASIS SIG/CR Classification Research Workshop. Ed. Efthimis Efthimiadis. Information Today, Medford:New Jersey. 1998. pp 59-72.
- [24]. H. Chen. and K. Lynch. *Automatic Construction of Networks of Concepts Characterizing Document Databases*. IEEE Transactions on Systems, Man, and Cybernetics, 22(5), 885-902.
- [25]. Chen, H. and Ng T. (1993) *An Algorithmic Approach to Concept Exploration in a Large Knowledge Network (Automatic Thesaurus Consultation): Symbolic Branch-and-Bound Search vs. Connectionist Hopfield Net Activation*. Journal of the American Society for Information Science. 46(5), 348-369.
- [26]. X. Lin, D. Soergel, and G. Marchionini. *A self organizing semantic map for information retrieval*. Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, pages 262-269, Chicago, IL, October 13-16 1991.
- [27]. A. Mercier, M. Beigbeder. *Fuzzy Proximity Ranking wit Boolean Queries*. Fourteenth Text REtrieval Conference (TREC 2005) Proceedings, 2005.
- [28]. O. Cordón, F. Moya, C. Zarco. *Fuzzy Logic and Multiobjective Evolutionary Algorithms as Soft Computing Tools for Persistent Query Learning in Text Retrieval Environments*. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2004), Budapest (Hungary), 571-576, 2004.
- [29]. P. Subasic, A. Huettner. *Affect Analysis of Text Using Fuzzy Semantic Typing*, IEEE Transactions on Fuzzy Systems, Special Issue, 2001.
- [30]. B. Martín del Brío, A. Sanz Molina. *Redes neuronales y sistemas borrosos*. Ra-Ma, 2001.
- [31]. M. Hellmann. *Fuzzy Logic Introduction*. Marzo, 2001.
- [32]. J. Jantzen. *Tutorial on fuzzy logic*. Technical University of Denmark, Department of Automation, Tech. report no 98-E 868, 19 Aug 1998 (logic).
- [33]. *Fuzzy Logic Toolbox. User`s guide*. The Mathworks Inc., 2002.
- [34]. W. Siler. *Building Fuzzy Expert Systems*. <http://members.aol.com/wsiler/>

- [35]. C.S. Ho. *Introduction to Information Technology*.
<http://ailab2.et.ntust.edu.tw/~csho/IntroAI/Lecture07.ppt>
- [36]. R.I. John. *Fuzzy Inferencing Systems – Problems and Some Solutions*. Computing Science Research. School of Computing Sciences. De Montfort University. The Gateway, Leicester, Working Paper N. 62, Diciembre, 1995.
- [37]. <http://www.cajamadrid.es/CajaMadrid/Home/puente?pagina=3447>
- [38]. D. Larios. *Un-Fuzzy y la implementación de sistemas de lógica difusa: Comparación con otras alternativas*. 2004.