

Sistema basado en reglas difusas para el mapeo de ontologías

Susel Fernández

Juan R. Velasco

Miguel A. López-Carmona

Departamento de Automática, Universidad de Alcalá
Edificio Politécnico, Ctra. N-II Km. 31,600
28871 Alcalá de Henares, Madrid, España
{susel,juanra,miguellop}@aut.uah.es

Resumen

Las ontologías son una herramienta crucial para especificar formalmente el vocabulario y la relación de conceptos en las plataformas de agentes, por lo que, para compartir información, agentes que utilizan diferentes vocabularios deben poder traducir los datos de un marco ontológico a otro. El tratamiento de la incertidumbre juega un papel fundamental en el mapeo de ontologías, puesto que el grado de solapamiento entre conceptos es difícil de representar lógicamente, y por lo general las aplicaciones no cuentan con la suficiente información al respecto. Este artículo está enfocado a ofrecer mecanismos de ayuda a los expertos en la primera fase del mapeo de ontologías, utilizando técnicas de lógica difusa para determinar el grado de similitud entre conceptos de ontologías diferentes. Para cada par de conceptos se calculan dos tipos de similitud: la primera utilizando el coeficiente de Jaccard sobre muestras de documentos relevantes tomados de la web, y la segunda basada en la relación lingüística de los conceptos. Finalmente se calcula la similitud total a través de un sistema basado en reglas difusas. Las ideas presentadas en este trabajo fueron validadas utilizando varias ontologías del mundo real.

Palabras Clave: Mapeo de ontologías, sistema basado en reglas difusas, similitud, conceptos.

1 INTRODUCCIÓN Y ANTECEDENTES

La investigación sobre los servicios de web semántica promete una mayor interoperabilidad entre agentes software y servicios web, utilizando ontologías compartidas publicadas en la web semántica. Sin embargo, los servicios producidos y descritos por diferentes desarrolladores pueden utilizar diferentes o quizás parcialmente solapados conjuntos de ontologías, por lo que, para compartir información, agentes que

utilizan diferentes vocabularios deben poder traducir los datos de un marco ontológico a otro.

El mapeo de ontologías no es más que el proceso de encontrar correspondencias entre los conceptos de dos ontologías a través de reglas. Este proceso es necesario para el intercambio de información y servicios dentro de una plataforma de agentes y puede ser un valioso paso en la solución al problema de la traducción de ontologías.

La incertidumbre está relacionada con todos los aspectos de la web semántica, puesto que una descripción de un concepto desconocido o un objeto puede ser incierta, y a menudo se da el caso de que un concepto definido en una ontología empareja solo parcialmente con uno o más conceptos en otra. En muchas aplicaciones el solapamiento entre conceptos y clases no puede ser representado lógicamente, lo que dificulta el proceso de mapeo.

Existen algunos trabajos anteriores encaminados al mapeo de ontologías, que han realizado aportes interesantes. Noy y Musen han desarrollado las herramientas SMART [15], PROMPT [16], y PROMPTDIFF [17], que usan emparejamiento de similitud lingüística entre los conceptos para iniciar un proceso de fusión o alineación y, entonces, utilizan las estructuras ontológicas para proporcionar un conjunto de heurísticos para identificar más coincidencias entre las ontologías.

McGuinness y colegas desarrollaron una herramienta interactiva similar, Chimaera [13], donde el ingeniero es el encargado de tomar las decisiones que afectarán el proceso de fusión. Chimaera analiza las ontologías que se van a fusionar, y si encuentra coincidencias lingüísticas, la fusión se realiza automáticamente. De lo contrario el usuario es avisado para que tome nuevas medidas. Es bastante similar a PROMPT en cuanto a que ambos están incrustados en los entornos de edición de la ontología, pero difieren en las sugerencias que hacen a sus usuarios con respecto a los pasos de fusión.

Otras investigaciones han seguido enfoques probabilísticos. Por ejemplo, Doan y colegas desarrollaron el sistema GLUE [6], que emplea técnicas de aprendizaje automático para encontrar mapas entre ontologías. Dadas dos ontologías, para cada concepto en una ontología, GLUE encuentra la mayoría de los conceptos similares en la otra ontología utilizando definiciones probabilísticas de varias medidas prácticas de

similitud. GLUE calcula la distribución conjunta de los conceptos, y permite a la aplicación la utilización de esta distribución para calcular cualquier medida adecuada de similitud. Sin embargo este enfoque no tiene en cuenta por completo la incertidumbre en el mapeo, ya que se ignoran conceptos que son similares pero en un menor grado.

Ron Pang y colegas desarrollaron un marco probabilístico de mapeo automático de ontologías basado en Redes Bayesianas (BNs) [18]. En este marco, las ontologías son primero traducidas a Redes Bayesianas, y a continuación el mapeo de conceptos se realiza como razonamiento evidencial entre las dos BNs. Las probabilidades necesarias tanto en la traducción como en el mapeo se obtienen mediante el uso de programas de clasificación de texto asociando distintos conceptos con documentos de texto relevantes recuperados de la web. Este enfoque solo tiene en cuenta la probabilidad de aparición de los conceptos en la web y no su similitud lingüística, por lo que falla en el caso en que dos conceptos muy similares no tengan el mismo índice de popularidad en la web.

Nuestra propuesta se centra en la primera etapa del mapeo de ontologías, hallando similitudes entre conceptos. Primeramente se calcula una similitud entre los conceptos utilizando el coeficiente de Jaccard, a partir de documentos relevantes recuperados de la web, específicamente de la Wikipedia [24]. Luego se calcula la similitud lingüística, y posteriormente se obtiene la similitud total utilizando un sistema basado en reglas difusas. La organización del trabajo es la siguiente: en la sección 2 se describe el uso del coeficiente de Jaccard y el método utilizado para obtener la similitud lingüística. En la sección 3 se muestra el sistema basado en reglas difusas, y en la sección 4 se reflejan los resultados de aplicar nuestro método en dos ontologías del mundo real. Finalmente en la sección 5 se presentan las conclusiones y las líneas de trabajo futuras.

2 CALCULO DE SIMILITUDES ENTRE CONCEPTOS

2.1. COEFICIENTE DE JACCARD

El coeficiente de Jaccard [21] es uno de los índices binarios de similitud más conocidos y utilizados. Se define como el tamaño de la intersección dividido entre el tamaño de la unión entre dos conjuntos de datos y su valor está en [0, 1].

Dada la siguiente tabla de contingencia con 2 variables binarias,

Tabla 1 Parámetros para calcular el coeficiente de Jaccard

i / j	1	0	Total
1	a	b	$a+b$
0	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$

Donde:

- a es el numero de veces que ambas observaciones tienen valor 1.
- b es el numero de veces que la observación i tiene valor 1 y la observación j tiene valor 0.
- c es el numero de veces que la observación i tiene valor 0 y la observación j tiene valor 1.
- d es el número de veces que ambas observaciones tienen valor 0.

El coeficiente de similitud de Jaccard se calcula a través de la siguiente fórmula:

$$s_{ij} = \frac{a}{a+b+c} \quad (1)$$

En nuestro trabajo hemos aplicado el coeficiente de Jaccard para calcular la similitud entre conceptos de dos ontologías diferentes. Los valores que intervienen en esta fórmula fueron obtenidos a través de búsquedas sucesivas de documentos relevantes a cada concepto en la web. De manera similar a la desarrollada en [18], para asegurar que la búsqueda de documentos de la web solo devolviera aquellos que fueran relevantes al concepto, la consulta de búsqueda se formó mediante la concatenación de todos los conceptos que se encontraban en el camino desde la raíz hasta el nodo actual en la taxonomía.

Los ejemplares que contenían el concepto A serían obtenidos mediante la búsqueda de A y todos sus ancestros en la taxonomía, mientras que los ejemplares que no contenían a A serían aquellos en los que estuvieran presentes todos sus ancestros en la taxonomía, y no estuviera presente el concepto A. Para cada par de conceptos A y B, pertenecientes a las ontologías origen y destino respectivamente se contaron (a) los documentos en los que aparecían los dos conceptos, (b) los documentos en los que aparecía A y no aparecía B, (c) los documentos en los que aparecía B y no aparecía A y (d) los documentos en los que no aparecía ninguno de los dos conceptos. Una vez obtenidos estos valores, para cada par de conceptos de las ontologías origen y destino respectivamente, fue calculada su similitud de Jaccard por la fórmula (1).

Para la búsqueda de los documentos en la web hemos utilizado el API AJAX de Google [10]. En trabajos previos realizábamos la búsqueda en toda la web, pero se obtenían muchos documentos irrelevantes, por lo que en este trabajo restringimos el dominio de búsqueda al sitio de la Wikipedia [24].

2.2. SIMILITUD LINGÜÍSTICA

La similitud lingüística constituye el indicador más fuerte del parecido entre dos conceptos, debido a que por lo general los desarrolladores de ontologías dentro de un mismo dominio emplean términos relacionados lingüísticamente para expresar conceptos equivalentes.

Dados los conceptos A y B de dos ontologías diferentes, el primer paso de nuestra propuesta para calcular la similitud lingüística consiste en obtener las listas de sinónimos asociadas a cada uno de ellos, auxiliándonos de la herramienta WordNet [22]. Posteriormente, a cada una de las palabras de las listas de sinónimos obtenidas se les aplica el algoritmo de stemming de Porter [19], que no es más que el proceso de eliminar los finales morfológicos de las palabras en inglés, para obtener las raíces morfológicas de las mismas. Si llamamos L_A y L_B a las listas de raíces de los sinónimos de cada concepto obtenidas en esta etapa, podemos calcular la similitud lingüística entre los conceptos como sigue:

$$S = \min \left[\frac{c_A}{T_A}, \frac{c_B}{T_B} \right] \quad (2)$$

Donde c_A es la cantidad de palabras de la lista L_A que están en la lista L_B , T_A es la cantidad total de palabras de la lista L_A , c_B es la cantidad de palabras de la lista L_B que están en la lista L_A , y T_B es la cantidad total de palabras de la lista L_B .

Por ejemplo, si tenemos los conceptos “*Intelligent agents*” y “*agents*”. Después de ejecutar los primeros pasos del algoritmo obtendríamos como resultado la tabla 2.

Tabla 2 Valores obtenidos en los pasos iniciales del cálculo de la similitud lingüística

Conceptos	Lista de sinónimos del concepto	Aplicando el algoritmo de stemming
Intelligent Agents	intelligent, agile, alert, clever, smart, brilliant, agent, factor, broker, functionary	L_A =Intellig, agil, alert, clever, smart, brilliant, <u>agent</u> , <u>factor</u> , <u>broker</u> , <u>functionari</u>
Agents	agent, factor, broker, functionary	L_B = <u>agent</u> , <u>factor</u> , <u>broker</u> , <u>functionari</u>

Por tanto la similitud lingüística sería 0.4.

3 SISTEMA BASADO EN REGLAS DIFUSAS

La lógica difusa en la actualidad es un campo de investigación muy importante tanto por sus aplicaciones matemáticas y teóricas como por sus aplicaciones prácticas.

Los sistemas basados en reglas difusas constituyen una extensión de los sistemas basados en reglas clásicos debido a que se componen de reglas “*IF-THEN*” cuyos antecedentes y consecuentes están compuestos por instrucciones de lógica difusa en lugar de instrucciones de lógica clásica. Estos sistemas son usados como herramientas para representar diferentes formas de conocimiento de un problema, así como para modelar las interacciones y relaciones entre sus variables. Debido a esta propiedad, han sido exitosamente aplicados a un amplio rango de problemas en diferentes dominios con incertidumbre y conocimiento incompleto [4].

Los sistemas basados en reglas difusas se componen de una base de conocimientos, un motor de inferencias, una interfaz de fuzzificación en la entrada, y una interfaz de defuzzificación en la salida. Figura 1, tomada de [4].

La base de conocimientos está formada principalmente por una base de datos, que contiene todos los datos relativos al sistema, como variables y valores posibles, y una base de reglas, que contiene todas las proposiciones que van a regir el sistema. El motor de inferencias es el “supervisor”, y se encarga de extraer las conclusiones partiendo de los datos simbólicos que le han llegado, aplicando las reglas que rigen el sistema difuso en el cuál trabaja. Por último las interfaces de fuzzificación y defuzzificación tienen la función de convertir una entrada natural en un valor difuso (fuzzificación), y una salida difusa en un valor natural (defuzzificación).

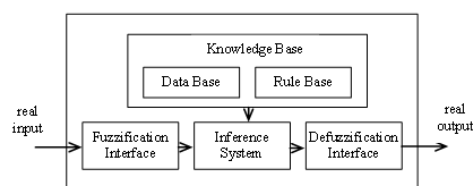


Figura 1. Estructura básica de un sistema basado en reglas difusas.

En este trabajo hemos utilizado el entorno de desarrollo XFuzzy 3.0 [23], que integra varias herramientas para cubrir las diferentes etapas del diseño de sistemas difusos.

3.1. VARIABLES DEL SISTEMA

En el sistema se han definido dos variables de entrada y una de salida. A cada una de ellas hemos asociado un conjunto de términos lingüísticos cuya semántica ha sido representada por medio de funciones triangulares de pertenencia. Estas variables son las siguientes:

a) **Sim_Jaccard**: Es una variable de entrada y representa el valor de la similitud calculado a través de coeficiente de Jaccard. Tiene asociado el siguiente conjunto de términos lingüísticos: $D_{jacc} = \{Baja, Regular, Media, Alta, MuyAlta\}$.

Para definir las funciones de pertenencia primeramente fue necesario dividir los datos de la similitud de Jaccard en grupos, por lo que decidimos utilizar los cuartiles del conjunto de similitudes para acotar los triángulos de pertenencia de la siguiente manera:

Baja: (-0.00324149, 0, 0.00324149)
Regular: (0, 0.00324149, 0.0831758)
Media: (0.00324149, 0.0831758, 0.20338982)
Alta: (0.0831758, 0.20338982, 1)
MuyAlta: (0.20338982, 1, 0.20338982)

Las funciones de pertenencia de la variable *Sim_Jaccard* se muestran gráficamente en la figura 2.

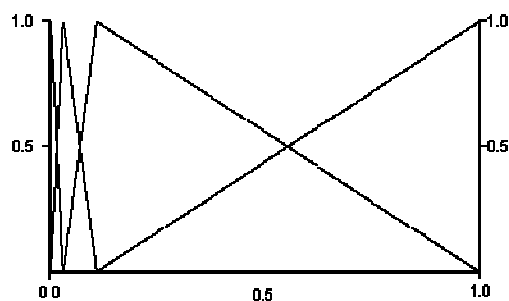


Figura 2. Funciones triangulares de pertenencia de la variable *Sim_Jaccard*.

b) **Sim_Ling**: Es una variable de entrada que representa el valor de la similitud lingüística. Tiene asociado el siguiente conjunto de términos lingüísticos: $D_{ling} = \{Baja, Regular, Media, Alta, MuyAlta\}$. Sus funciones de pertenencia se muestran en la figura 3.

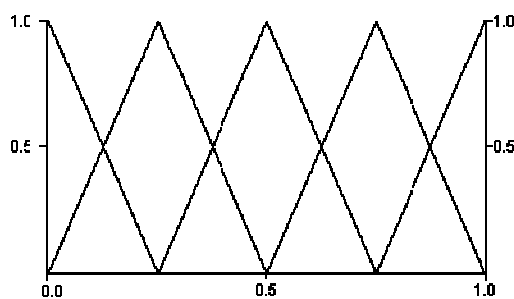


Figura 3. Funciones triangulares de pertenencia de la variable *Sim_Ling*.

c) **Similitud**: Es la variable de salida y representa el valor de la similitud total, obtenida por el sistema difuso. Tiene asociado el siguiente conjunto de términos lingüísticos: $D_{Similitud} = \{MuyBaja, Baja, MedioBaja, Regular, MedioAlta, Alta, MuyAlta\}$. Sus funciones de pertenencia se muestran en la figura 4.

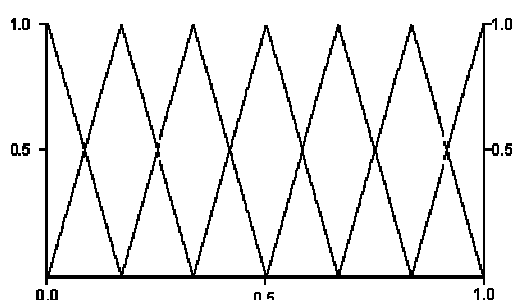


Figura 4. Funciones triangulares de pertenencia de la variable *Similitud*.

3.2 BASE DE REGLAS

La base de reglas que hemos definido en el sistema difuso ha sido formada mediante heurísticos de los expertos y se muestra en la Tabla 3.

Tabla 3. Base de reglas

Ling/Jacc	Baja	Regular	Media	Alta	Muy Alta
Baja	Muy Baja	Baja	Medio Baja	Regular	Medio Alta
Regular	Baja	Medio Baja	Regular	Medio Alta	Alta
Media	Baja	Medio Baja	Regular	Medio Alta	Alta
Alta	Medio Baja	Regular	Medio Alta	Alta	Muy Alta
Muy Alta	Regular	Medio Alta	Alta	Alta	Muy Alta

4 EXPERIMENTOS

Hemos realizado varios experimentos, utilizando diversas ontologías, con resultados satisfactorios. A modo de ejemplo, describimos un experimento en el que se utilizan dos ontologías del mundo real extraídas de la Web. Estas son, la ontología ACM [1], y la ontología DMOZ [5]. Nuestro objetivo ha sido demostrar la eficacia de esta propuesta para la obtención de un índice de similitud válido entre conceptos, que ayude al proceso de mapeo de ontologías. De manera similar a [18], hemos seleccionado el tópico de Inteligencia Artificial, haciendo una poda de conceptos en ambas ontologías de forma tal que exista cierto grado de solapamiento entre ambas, y se pueda comprobar la efectividad del método. La figura 5 muestra los conceptos seleccionados de ambas ontologías, así como las relaciones evidentes entre ellos. Para cada par de conceptos A y B, pertenecientes a las ontologías ACM y DMOZ respectivamente, hemos calculado primero la similitud de Jaccard, utilizando los conjuntos de documentos relevantes recuperados de la Web. Como hemos mencionado anteriormente esta búsqueda está restringida a la Wikipedia. Posteriormente se calcula la similitud lingüística, como se describe en secciones anteriores.

Si consideramos como n la cantidad de conceptos de la primera ontología, y m la cantidad de conceptos de la segunda, el proceso se completaría en $n \times m$ iteraciones. La tabla 4 muestra los conceptos para los cuáles hemos obtenido índices de similitud más elevados. La tabla 5 muestra los conceptos para los cuales la similitud ha resultado ser más alta de lo debido.

En el trabajo de [18], utilizando fragmentos similares de las ontologías ACM y DMOZ, plantean que según su método se obtiene un índice de similitud de 0.61 para los conceptos “*connectionism and neural nets*” y “*Neural Networks*”, debido a que solo tienen en cuenta la probabilidad de aparición de ambos conceptos en la web, y el término “*connectionism*” es muy poco popular. Como se puede observar en la figura 4, aplicando nuestro método, hemos obtenido un grado de similitud más alto para estos dos conceptos, debido fundamentalmente a su relación lingüística.

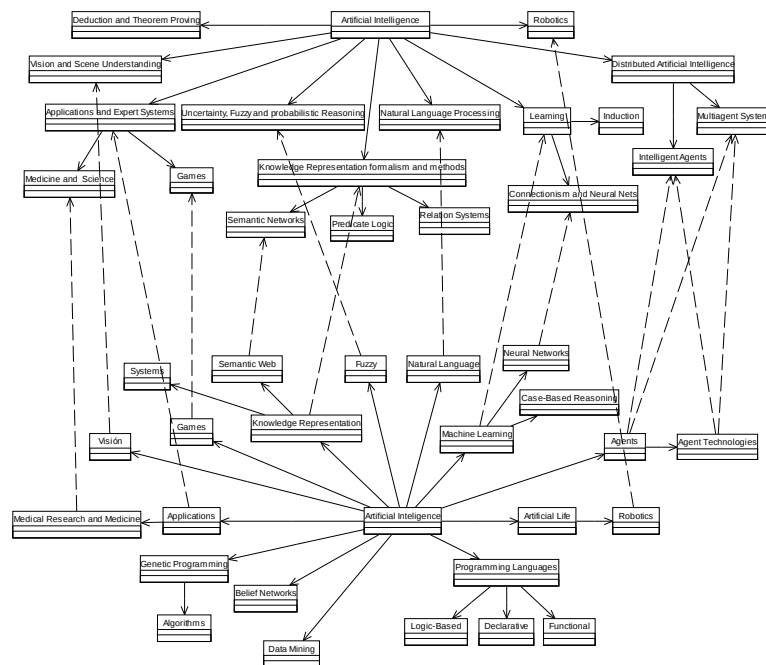


Figura 5. Conceptos de las ontologías ACM y DMOZ.

Tabla 4. Parejas de conceptos con similitud más alta.

Conceptos ACM	Conceptos DMOZ	Similitud
Robotics	Robotics	0.99
Games	Games	0.99
Predicate Logic	Logic-Based	0.91
Intelligent Agents	Agents	0.91
Natural Language Processing	Natural Language	0.91
Multiagent Systems	Systems	0.87
Learning	Machine learning	0.87
Multiagent Systems	Agent Technologies	0.85
Intelligent Agents	Agent Technologies	0.74
Multiagent Systems	Agents	0.72
Knowledge Representation Formalisms and Methods	Knowledge Representation	0.71
Connectionism and neural nets	Neural Networks	0.71
Applications and Expert Systems	Applications	0.71
Semantic Networks	Semantic Web	0.71
Relation Systems	Systems	0.58
Medicine and Science	Medical Research and Medicine	0.58
Natural Language Processing	Programming Language	0.55
Vision and Scene Understanding	Vision	0.55
Uncertainty, fuzzy, and probabilistic reasoning	Fuzzy	0.55
Semantic Networks	Neural Networks	0.55

Tabla 5. Parejas de conceptos con similitud más alta de lo esperado.

Conceptos ACM	Conceptos DMOZ	Similitud
Predicate Logic	Logic-Based	0.91
Multiagent Systems	Systems	0.87
Semantic Networks	Neural Networks	0.55
Natural Language Processing	Programming Language	0.55

Finalmente, hemos realizado una prueba estadística χ^2 para comprobar si los resultados obtenidos se corresponden con los esperados para estas dos ontologías. Para ello planteamos la hipótesis H_0 que consiste en suponer que no hay diferencia entre estos valores. Calculando con 1 grado de libertad, hemos comprobado que se cumple la hipótesis H_0 con un 95% de confianza, por lo que se acepta la suposición y se demuestra la efectividad del método propuesto. La tabla 6 muestra los valores de contingencia para el experimento.

Tabla 6. Tabla de contingencia para la prueba χ^2

	Baja	Alta	Total
Esperados	416	16	432
Observados	412	20	432

5 CONCLUSIONES Y TRABAJO FUTURO

En este artículo se describe nuestro trabajo encaminado a ofrecer un método que ayude a los expertos en la fase inicial del mapeo de ontologías. Nuestra propuesta consiste en un sistema difuso para obtener similitudes

entre conceptos de diferentes ontologías. Este sistema utiliza como entrada valores obtenidos aplicando el coeficiente de similitud de Jaccard a documentos recuperados de la Wikipedia, y un índice de similitud basado en la relación lingüística de los conceptos, con mucha influencia en el cálculo de la similitud total. Actualmente este proceso se realiza para cada pareja de conceptos de las dos ontologías, y estamos trabajando para reducir el número de iteraciones, haciendo que el sistema tenga en cuenta otros factores como las similitudes entre conceptos hermanos, hijos y padres en la taxonomía de cada ontología. Otro de nuestros objetivos es refinar el cálculo de la similitud lingüística incluyendo otro tipo de relaciones lingüísticas entre los conceptos, y la aplicación de reglas difusas para calcular la similitud lingüística global. También como trabajo futuro nos hemos propuesto incluir las propiedades de los conceptos en el proceso de mapeo, así como la escalabilidad de la aplicación.

Agradecimientos

Este trabajo es parte del proyecto T2C2, financiado por el Ministerio de Educación y Ciencia, TIN2008-06739-C04-04, y el proyecto europeo ITEA-2 No 2008005, "Do-it-Yourself Smart Experiences", financiado por el Ministerio de Industria, Turismo y Comercio de España.

Referencias

- [1] ACM Topic, <http://www.acm.org/about/class/1998/>
- [2] Barwise, J., Seligman, J.: *Information Flow: The Logic of Distributed Systems*. Cambridge University Press (1997).
- [3] Calvanese, D., Giacomo, G., Lenzerini, M.: *Ontology of integration and integration of ontologies*. In: *Description Logic Workshop (DL 2001)*, pp 10–19 (2001).
- [4] Córdón, O., Herrera, F., Hoffman, F., Magdalena, L.: *Genetic Fuzzy Systems. Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*. World Scientific, Singapore (2001).
- [5] DMOZ hierarchie, <http://www.dmoz.org>
- [6] Doan A., Madhavan, J., Domingos, P., Halevy, A.: *Ontology Matching: A Machine Learning Approach*. Handbook on Ontologies in Information Systems. In: S. Staab and R. Studer (eds.), *Invited paper*. Pp. 397–416. Springer-Verlag, (2004).
- [7] Fernández-Breis, J., Martínez-Béjar, R.: *A cooperative framework for integrating ontologies*. In: *International Journal of Human-Computer Studies*, 56: pp. 665–720, (2002).
- [8] Gruber, T., Olsen, G.: *An ontology for engineering mathematics*. In: J. Doyle, P. Torasso, E. Sandewall, (eds). *Fourth International Conference on Principles of Knowledge Representation and Reasoning*, pp. 258–269, San Mateo, CA, USA (1994).
- [9] Grüninger, M.: *Ontologies for translation: Notes for refugees from Babel*. EIL Technical Report, Enterprise Integration Laboratory (EIL), University of Toronto, Canada (November 1997).
- [10] Google AJAX API, <http://code.google.com/intl/es/apis/ajaxsearch>
- [11] Jannink, J., Pichai, S., Verheijen, D., Wiederhold, G.: *Encapsulation and Composition of Ontologies*. In: *AAAI'98 Workshop on Information Integration*, Madison, WI, USA, pp. 43–51 (July 1998).
- [12] Kalfoglou, Y., Schorlemmer, M.: *Ontology mapping: the state of the art*. *The Knowledge Engineering Review*, 18(1) pp. 1–31, (2003).
- [13] McGuinness, D., Fikes, R., Rice, J., Wilder, S.: *An Environment for Merging and Testing Large Ontologies*. In: *17th International Conference on Principles of Knowledge Representation and Reasoning (KR-2000)*, Colorado, USA, pp. 483–493, (April 2000).
- [14] Mitra, P., Noy, N. F., Jaiswal, A. R. 2004. OMEN: A Probabilistic Ontology Mapping Tool. In: *Workshop on Meaning Coordination and Negotiation at the Third International Conference on the Semantic Web (ISWC-2004)*. Hiroshima, Japan, pp. 71–83, (2004).
- [15] Noy, N. F., Musen, M. A.: *SMART: Automated Support for Ontology Merging and Alignment*. In: *12th Workshop on Knowledge Acquisition, Modelling and Management (KAW'99)*, Banff, Canada, pp. 1–20, (October 1999).
- [16] Noy, N. F., Musen, M. A.: *The PROMPT suite: Interactive tools for ontology merging and mapping*. *International Journal of Human-Computer Studies*, 59(6), pp. 983–1024 (2003).
- [17] Noy, N. F., Musen, M. A.: *PROMPTDIFF: A Fixed-Point Algorithm for Comparing Ontology Versions*. In: *18th National Conference on Artificial Intelligence (AAAI'02)*, pp. 744–750, Edmonton, Alberta, Canada, (August 2002).
- [18] Pan, R., Ding, Z., Yu, Y., Peng, Y.: *A Bayesian Network Approach to Ontology Mapping*. *The Semantic Web – ISWC 2005*, Vol. 3729/2005, pp. 563–577. Springer Berlin / Heidelberg (October 2005).
- [19] Porter Stemming algorithm, <http://tartarus.org/~martin/PorterStemmer>
- [20] Quesada, V., Isidoro, A., López, L.: *Curso y Ejercicios de Estadística*. Alhambra Longman, Madrid (1994).
- [21] Rijsbergen, V., C. J.: *Information Retrieval*. Butterworths. Second Edition, London (1979).
- [22] Wordnet, <http://wordnet.princeton.edu>
- [23] XFuzzy 3.0 framework, http://www.imse.cnm.es/Xfuzzy/Xfuzzy_3.0/tools/xfuzzy_sp.html
- [24] Wikipedia, <http://www.wikipedia.org>