

Aplicación de Lógica Difusa en el Proceso de Expansión de Folksonomías

Carmen Martínez-Cruz¹ Ignacio J. Blanco² M. Amparo Vila²

¹ Dpto. Informática, Universidad de Jaén
Campus de Las Lagunillas, 035-A3, 23071, Jaén
cmcruz@ujaen.es

² Dpto. Ciencias de la Computación e I. A., Universidad de Granada
C/ Periodista Daniel Saucedo Aranda S/N, 18071, Granada
{iblanco,vila}@decsai.ugr.es

Resumen

Las folksonomías tienen una naturaleza, carente de semántica que hace muy difícil su utilización y explotación en la Web. Dichas estructuras pueden ser enriquecidas a través de su asociación con términos significativos identificados en la Web Semántica. Sin embargo dicho proceso implica una primera fase de desambiguación y una segunda de expansión del conjunto de etiquetas inicial. En esta propuesta se propone la utilización de la lógica difusa para extender el conjunto de etiquetas pertenecientes a la folksonomía y asociar un grado de pertenencia de cada etiqueta al conjunto en función del grado de similitud de las etiquetas añadidas con el conjunto inicial. Además se permite al usuario, a través del uso de etiquetas lingüísticas, seleccionar el subconjunto de etiquetas que mejor se ajuste a sus necesidades. Por último, también se comparan los dos principales repositorios de información disponibles en la actualidad, *WordNet* y *Wikipedia*, dada su influencia en ambos procesos.

1 Introducción

Las folksonomías[9] juegan un papel muy importante en la Web 2.0 puesto que permiten la descripción de los recursos que en ella se encuentran, p.e. imágenes, URL, noticias, etc. La naturaleza de estas estructuras es completamente opuesta a la de la Web Semántica y las Ontologías, dado que estos conjuntos de etiquetas los especifican los usuarios, normalmente no expertos y no existe ninguna convención de sintaxis o vocabulario. Con estas características, estas estructuras carentes de semántica, son difícilmente procesables y la asociación de semántica a las mismas pasa por establecer relaciones explícitas con ontologías [1, 7].

Una folksonomía puede contener términos que no están

relacionados, p.e. una imagen se puede describir como *amarillo, juguete, transporte, arte, Venezuela*, lo cual provoca que sea realmente complejo entender el verdadero significado de una etiqueta. Incluso un proceso de desambiguación puede conducir a error, a pesar de que los repositorios como *Wordnet* o *Wikipedia* juegan un papel crucial en el mismo dado que proporcionan un gran número de sentidos a cada etiqueta. Desafortunadamente, dichas definiciones son muy similares en muchos casos provocando el aumento de la complejidad del proceso. Por otro lado, las etiquetas usadas pueden ser muy diversas para describir el mismo recurso. El proceso de expansión de una folksonomía puede ayudar a la contextualización de la etiqueta y posteriormente del recurso. Sin embargo este proceso también puede conducir a errores dado que la relación entre los tags iniciales y los expandidos pueden tener ligeras diferencias semánticas, como una relación en hipernimia.

En esta propuesta se propone utilizar la lógica difusa para solventar los problemas descritos: En lugar de descartar los diferentes sentidos de una etiqueta, se conservan los mismos asignándoles un grado de confianza dependiendo de su similitud con el resto del conjunto de etiquetas. Tras esto, el proceso de expansión también asigna un grado de similitud a cada etiqueta añadida al conjunto. Dicho grado se ajusta en función de la relación semántica del término añadido con el conjunto inicial, p.e. sinonimia, hipernimia, meronimia, definiciones, etc. Como resultado, el conjunto inicial expandido puede filtrar los resultados utilizando etiquetas lingüísticas que facilitan al usuario final el proceso de selección de resultados. Además, los repositorios utilizados como fuentes de información (*WordNet* y *Wikipedia*), críticos en este proceso, son utilizados y analizados.

En la sección 2 se describe el proceso de enriquecimiento de folksonomías. A continuación, la definición de la propuesta y un ejemplo analizando los resultados obtenidos se encuentran en las secciones 3

y 4 respectivamente. La definición de las etiquetas lingüísticas se presentan en la sección 5 y finalmente las conclusiones en la sección 6.

2 Fundamentos

2.1 Proceso de Enriquecimiento de Etiquetas

El Enriquecimiento de Folksonomías es una práctica reciente en el campo de la Web Semántica dada la proliferación de este tipo de estructuras en la Web Colaborativa. Dicho proceso, consistente en proporcionar semántica a las etiquetas que componen la folksonomía, conecta la Web 2.0 con la Web Semántica. Existen varias propuestas que desarrollan este *proceso de descubrimiento semántico* en las folksonomías, como la de Specia et al. [7] que agrupa etiquetas relacionadas para relacionarlas con términos de ontologías o la propuesta de Angelou et al. [1] que enriquece una folksonomía limpiada y expandida con los recursos encontradas en diferentes ontologías. Esta última especifica la metodología seguida en este trabajo, resumida en la figura 1.

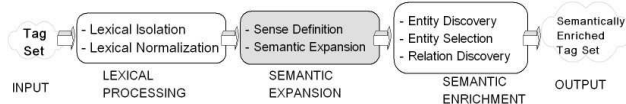


Figura 1: Secuencia del proceso de enriquecimiento de folksonomías en el proyecto *Flor* [1]

2.2 Fuentes de Información

2.2.1 WordNet

Wordnet es un diccionario en línea que modela el idioma inglés. Sus características principales son [6]:

1. Define cada sentido de una palabra, *synset*.
2. Cada *synset* tiene sus sinónimos asociados.
3. Cada *synset* tiene una definición.
4. La hiperonimia se representa para cada *synset* usando una estructura de árbol.
5. Define relaciones de holonimia (*es miembro de*) y meronimia (*tiene miembros*).

El éxito de *WordNet* se debe a la representación tan estructurada que realiza de la información, sin embargo, también presenta las siguientes desventajas [3]:

1. Carece de conexiones entre nombres y verbos y de las relaciones típicas entre las palabras.
2. Carece de relaciones morfológicas.
3. Carece de algunos sentidos y relaciones. También de relaciones temáticas.
4. No tiene una definición uniforme.

5. El primer sentido no es el más lógico, p.e. *tiger*.
6. Tiene una baja cobertura, sobretodo con nombres propios [4].
7. Define demasiados sentidos, y estos suelen tener una granularidad muy fina [4].

2.2.2 Wikipedia

Wikipedia es una enciclopedia en línea desarrollada por una comunidad colaborativa. Sus principales características son [4]:

1. Los artículos representan los sentidos,
2. Los enlaces proporcionan la fuente de sinónimos y otras variaciones de la palabra.
3. Las redirecciones son sinónimos y otras variaciones de la palabra
4. Las categorías representan la jerarquía de conceptos de *Wikipedia* en forma de red.
5. Las páginas de desambiguación se corresponden con los diferentes sentidos de una palabra.
6. La definición está en la primera frase del artículo.
7. Proporciona un repositorio de nombres propios y conceptos especializados [8].
8. Proporciona un gran repositorio de abreviaciones y acrónimos.

Sin embargo, los problemas que presenta son:

1. Las páginas de desambiguación son muy difíciles de procesar automáticamente [4, 5] y existen demasiados sentidos para un término.
2. Los títulos de página incluyen a veces información adicional.
3. Los enlaces incluyen polisemia.
4. Las categorías no representan sólo relaciones ES-UN[8].
5. Los artículos carecen de revisiones formales de expertos.

3 Proceso de Expansión de Etiquetas

3.1 Descripción del Proceso

La *Fase de Expansión Semántica* consiste en expandir un conjunto de etiquetas con otras que tengan un significado similar, para lo cual se necesita un repositorio como *Wordnet* y *Wikipedia*. El proceso tiene dos subfases: i) Definición de los sentidos, la desambiguación

de cada etiqueta se realiza utilizando una función de similitud que compara cada par de etiquetas, ii) Expansión Semántica, en esta fase el conjunto de etiquetas se aumenta con hipéronimos y sinónimos (vase ejemplo en la figura 2).

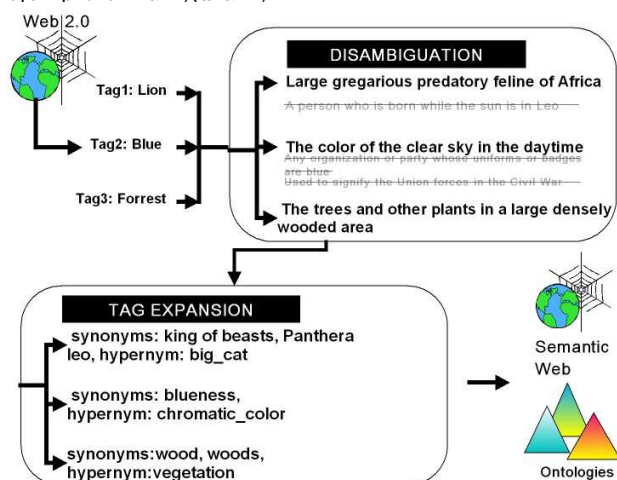


Figura 2: Ejemplo del proceso de expansión [1]

La propuesta descrita en este artículo mejora el proceso anteriormente descrito mediante la definición de un proceso de desambiguación capaz de elegir el número correcto de sentidos del conjunto de etiquetas y una posterior fase de expansión que incluye un grado de similitud en cada etiqueta del conjunto.

Además, el repositorio elegido para desarrollar estos procesos determinará el conjunto de etiquetas resultante, dada su estructura de datos y las relaciones representadas.

3.1.1 Fase 1. Proceso de Desambiguación

El conjunto de sentidos más relacionados con la folksonomía se obtiene en esta fase. Esta tarea presenta los siguientes problemas: i) Cada término tiene numerosos, a veces demasiados sentidos (ejemplo en tabla 3), ii) Las folksonomías se componen por términos que no están relacionados por naturaleza. Por tanto, es muy complicado seleccionar únicamente un único sentido de una palabra. La estructura de datos mostrada en la figura 3 permite gestionar todos los sentidos definidos en el conjunto de etiquetas:

1. Un *Grado de Confianza* que representa la similitud de la etiqueta con el resto del conjunto se define como un conjunto difuso, p.e. *Estrella* en la folksonomía *Estrella, Sol, Azul, Planeta* tiene asociado el conjunto difuso $[0.9, 0.3, 0.8]$ donde 0.9 representa la relación *Estrella-Sol*.
2. Cada sentido tiene asociado un *grado de confianza* que establece la relación de este con el resto del conjunto de etiquetas. Una etiqueta puede tener varios sentidos.

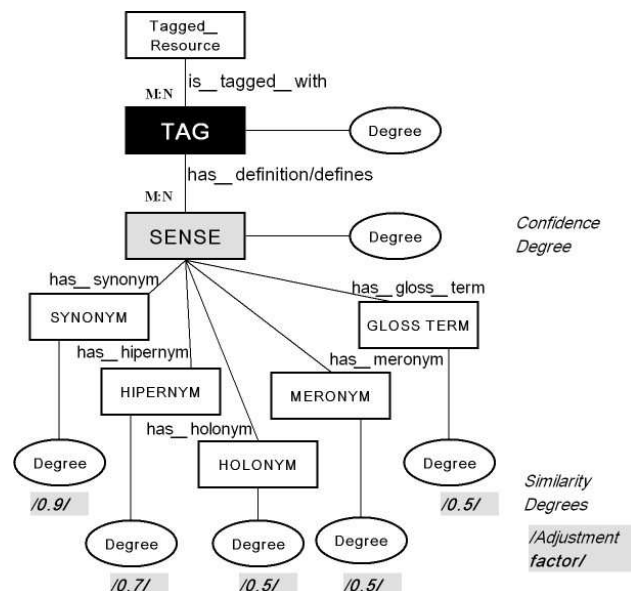


Figura 3: Estructura para extender una folksonomía usando grados difusos

El cálculo del *grado de confianza* consiste en aplicar un proceso de desambiguación donde se valora la adecuación del sentido con la etiqueta. Para ello se calcula la distancia entre cada dos términos dividiendo el camino más corto (hipéronimos) entre su distancia máxima. El *grado de confianza* final de la etiqueta será la *s-norma* (Dubois and Prade [2]) obtenida a partir de los conjuntos difusos de los sentidos que lo describen.

3.1.2 Fase 2: Proceso de Expansión

La fase de expansión incrementa el conjunto de etiquetas con otros términos relacionados. El *grado de similitud* que describe la similitud de las nuevas etiquetas con las originales se asigna de forma siguiente (véase figura 3):

1. Los sinónimos se añaden directamente al conjunto. La *Wikipedia* utiliza redirecciones y enlaces para definirlos.
2. Los hipéronimos presentan dos problemas: i) el número puede ser demasiado grande, ii) el significado puede ser demasiado genérico. En esta propuesta los hipéronimos se ponderan para mantener los mejor valorados.
3. Las definiciones pueden contextualizar una palabra, especialmente en una folksonomía. Sólo se conservan los nombres con una relación mínima con el sentido original. Esta medida se calcula utilizando la distancia Wu and Palmer [10] en *WordNet*.
4. Otros términos como meronimos y homónimos se incluyen. En esta categoría *Wikipedia* incluye en-

laces a otras páginas, sin embargo sólo aquellas con un gran número de referencias se añaden al conjunto.

El *grado de similitud* es el producto del *grado de confianza* del sentido inicial y el correspondiente *Factor de Ajuste*, mostrado en la figura 3. El *Factor de Ajuste* se asigna a cada estructura del sistema de acuerdo con la relación semántica con el sentido inicial.

4 Ejemplo de Aplicación.

4.1 Cálculo de Sentidos y Grados de Confianza

Para describir esta propuesta se muestra el proceso de expansión de una folksonomía de ejemplo compuesta por los términos ya filtrados y descritos en la tabla 1. Una vez el sistema ha obtenido los sentidos para cada etiqueta utilizando un repositorio particular, se calculan sus grados de similitud. Los resultados se muestran en las tablas 2 y 3 para *WordNet* y *Wikipedia* respectivamente.

Tag	WordNet	Wikipedia
Car	5	48
Auto	1	7
Compact	3	14
Environment	2	15

Tabla 1: Número de sentidos de la folksonomía. Para obtener un sentido se compara con todos los demás sentidos de la folksonomía. Este sentido puede guardarse si su grado de similitud es mayor que un umbral de 0.5 y si su relación con alguna de las otras etiquetas del conjunto es mayor que 0.7. Además, el conjunto difuso se obtiene aplicando la s-norma de Dubois and Prade s-norm con el factor de 0.3. Por último, el *grado de confianza final* es la media de los conjuntos difusos respectivos normalizada.

CAR				
sense	Auto	Compact	Environment	Sim.Deg.
1:RoadCar	0.6	0.6	0	0.4/0.8
Snorm	0.6	0.91	0	0.506/1
AUTO				
Sense	Car	Compact	Environment	Sim.Deg.
1:Car, Automobile	0.6	0.9	0	0.5/0.98
Snorm	0.6	0.9	0	0.5/0.98
COMPACT				
Sense	Car	Auto	Environment	Sim.Deg.
1:Economical car	0.5	0.9	0	0.46/0.92
Snorm	0.6	0.9	0	0.5/0.98
ENVIRONMENT				
sense	Car	Auto	Compact	Sim.Deg.
1:Surrounding Cond.	0	0	0	0/0
Snorm	0	0	0	0/0

Tabla 2: Sentidos evaluados usando *WordNet*.

4.1.1 Resultados

WordNet identifica el dominio *Automobiles* y detecta que *Environment* no tiene relación con el conjunto. Sin embargo, *Wikipedia* identifica el dominio de *computer variables*, ya que *Wikipedia* define numerosas categorías relacionadas con cada sentido.

Ambos repositorios identifican correctamente el dominio de *Automobiles* sin embargo *WordNet* ha ex-

CAR				
Sentido	Auto	Compact	Envir.	Sim.Deg.
1:Automobile	0.7	0.7	0.39	0.59/0.8
2:Car and CDR	0.7	0.3	0.7	0.56/0.76
3:CCR and CAR	0.3	0.8	0.6	0.56/0.76
4:Central Apparatus Room	0.39	0.6	0.7	0.56/0.76
Snorm	0.7	0.8	0.7	0.73/1
AUTO				
Sentido	Car	Compact	Envir.	Sim.Deg.
1:Auto Riskshaw	0.7	0.5	0.39	0.53/0.72
2:Automatic Variable	0.7	0.39	0.8	0.63/0.86
Snorm	0.7	0.7	0.8	0.73/1
COMPACT				
Sentido	Car	Auto	Envir.	Sim.Deg.
1:Measure	0.8	0.3	0.39	0.5/0.68
Snorm	0.8	0.5	0.5	0.59/0.80
ENVIRONMENT				
Sentido	Car	Auto	Compact	Sim.Deg.
1:Environment Variable	0.7	0.8	0.3	0.59/0.8
Snorm	0.88	0.8	0.39	0.63/0.86

Tabla 3: Sentidos evaluados usando *Wikipedia*.

cluido el término *Environment* que está actualmente relacionado con el campo *Environmental cars*. Sin embargo, *Wikipedia* ha identificado otro dominio, no correcto pero lógico.

Resumiendo, las categorías de *WordNet* están más especializadas y por tanto, las relaciones definidas entre los términos son mas distantes. Además, en este repositorio el número de definiciones de cada término es menor y por tanto más precisas. *Wikipedia*, al contrario, define un gran número de sentidos que facilita contextualizar los diferentes pares de palabras en el conjunto. Además cada sentido tiene numerosas categorías que provocan la obtención de distancias más cercanas.

4.2 Fase de Expansión

4.2.1 Ejemplo. Etiquetas Resultantes y Grados de Similitud

Una vez evaluadas las etiquetas y sus sentidos, se procede a extender y ponderar el conjunto de etiquetas. Siguiendo con el ejemplo anterior, el conjunto de etiquetas se expande usando *WordNet*, tal y como se muestra en la tabla 4 y con *Wikipedia*, tabla 5.

El cálculo de los grados de similitud se realiza, por ejemplo, para el sinónimo *Motorcar* = Grado de confianza del sentido * factor del sinónimo = $0.98 * 0.9 = 0.88$. Para el hiperónimo es ligeramente distinto, para *Motor-vehicle* el proceso sería: Grado de confianza del sentido * Grado del hiperónimo * Factor del hiperónimo = $0.98 * 0.93 * 0.7 = 0.63$.

4.2.2 Resultados

Wikipedia contiene más datos que *WordNet*. Además, *Wikipedia* no define explícitamente las relaciones que existen entre las palabras, así, los enlaces desde y hacia otros sentidos no tienen un claro significado en el conjunto resultante. Además, la cantidad de sinónimos en la *Wikipedia* es grande dado que este repositorio expresa el mismo sentido de diferente forma, como plurales, acrónimos, abreviaturas, etc.

Por otro lado, los hiperónimos en *Wikipedia* son más

Seq.	Sinónimos	Hiperónimos	Def.	Otros
CAR				
1	[railcar, railway_car, railroad_car]/0.72	wheeled_vehicle/0.50, vehicle/0.44, conveyance/0.39, transport/0.33, instrumentality/0.28]	[wheel]/0.4	[suspension, suspension_system]/0.4 ,[train, railroad_train]/0.4
AUTO				
1	[car, automobile, machine, motorcar]/0.88	motor_vehicle/0.63, self-propelled_vehicle/0.55, wheeled_vehicle/0.40, vehicle/0.35, conveyance/0.30, transport/0.25, instrumentality/0.20	[wheel, engine]/0.4	[accelerator, accelerator_pedal, gas_pedal, gas, throttle, gun, air_bag, auto_accessory, automobile_engine, automobile_horn, car_horn, motor_horn, horn, hooter, buffer, fender, bumper, car_door, car_mirror, car_seat, etc.]/0.49
COMPACT				
1	compact_car/0.82	car/0.61, auto/0.57, automobile/0.59, machine/0.56, motorcar/0.54, motor_vehicle/0.51, automotive_vehicle/0.49, self-propelled_vehicle/0.46, wheeled_vehicle/0.44, vehicle/0.41	[small, economical, car]/0.46	-
ENVIRONMENT				
1		situation/0	condition/0	

Tabla 4: Ejemplo de folksonomía semánticamente expandida y sus etiquetas de similitud usando *WordNet*

Seq.	Sinónimos	Hiperónimos	Def.	Otros
CAR				
1	[Cars, Motorcar, Automobiles, Motor Car, Ottomobile, Automobiles, Motorization, Motorisation, Cardoor, Autos, Automobile, etc.]/ 0.72	Industries/0.18, Travel/0.18, Technology by type/0.18, Transportation/0.34, Marketing/0.18, Manufacturing/0.18, Goods/0.18, Manufactured goods/0.34, Mechanical engineering/0.18, etc.	[wheel, motor, vehicle, engine, motor, transport, goods, university, press, vehicle]/0.4	[wheeled motor vehicle, transporting passengers, engine, transport, vehicle, Russia, Internal combustion, Gustave Trouve, Karl Benz]/0.4
2	[Car (function), Car and cdr] / 0.68	Programming language families/0.17, Lisp programming language family/0.35, Lisp programming language/0.53, etc.	-	[Lisp, Programming, Language]/0.38
3	CCR algebra, CAR algebra, Canonical anticommutation relation, CCR/CAR algebra/0.68	Mathematics/0.17, Mathematical analysis/0.35, Functions and mappings/0.35, Functional analysis/0.53, Physics/0.17, etc.	-	-
4	Central Machine Room, Central machine room, Central apparatus Room, Machine Room/0.68	Electronics/0.17, Communication/0.17, Networks/0.17, Electronic engineering/0.17, Technology by type/0.17, Telecommunications/0.35, etc.	[apparatus, or, or, machine, room, or, facility, gear, house, router, server, etc.]/0.38	[broadcast, television, audio, video routers, compressors, multiplexer]/0.38
AUTO				
1	Auto Rickshaw, Tuk-Tuk, Mototaxi, Auto-wallah, Auto Rickshaws in India, Auto Rickshaws in Pakistan, etc./0.64	South Asian countries, Transport in India, Transport in Pakistan, Transport in Bangkok, Capitals in Asia, Vehicles for hire/0.41, automotive_vehicle/0.5 -Only one level included-	vehicle, hire, chief, mode, transport, transport, south, asia, motorize, etc./0.36	South East Asia, Riskshaw, Bangladesh, motor vehicle, traffic congestion, seabelt, TVS motors, etc./0.36
2	Automatic memory allocation/0.77	Computer programming/0.2, Programming language design/0.2, Programming constructs/0.4, Algebra/0.2, Variables/0.4, Variable (computer programming)/0.6	-	lexally scope, stack grame, processor, registers, Java, C, C++, Perl/0.43
COMPACT				
1		Mathematics/0.15, Mathematical analysis/0.31, Functions and mappings/0.31, Functional analysis/0.6		[volume area, shape, circle]
ENVIRONMENT				
1	SystemRoot, SystemDrive, Environment variables, Env var, LC ALL, Environment (comp. science), etc./0.72	Computing/0.18, System software/0.18, Computer Science/0.18, Operating systems/0.36, Operating system technology/0.56, Computer programming/0.18, etc.	-	vales processes, Unix, Dos, Microsoft Windows, file system, user, command line, bash, computer terminal, etc./0.4

Tabla 5: Ejemplo de folksonomía semánticamente expandida y sus etiquetas de similitud usando *Wikipedia*

específicos que en *WordNet* dado que en *Wikipedia* se define un gran número de categorías. De esta forma alcanzar una categoría "genérica" requeriría obtener un gran número de hiperónimos. Consecuentemente, el número y grado de los hiperónimos en *Wikipedia* será finito y establecido a priori. *Wikipedia*, concretamente, define más categorías para cada sentido y todas ellas son más específicas que las de *WordNet* que son más genéricas y concretas.

Por ltimo, los términos de la definición obtenidos mediante el uso de las categorías, no devuelven buenos resultados dado que los términos de la definición no tienen relación de hiperonimia con los sentidos iniciales.

5 Etiquetas Lingüísticas

El establecimiento de umbrales correctos y flexibles permiten adecuar el número de etiquetas resultantes y sus grados de similitud en el conjunto extendido final

es una tarea compleja. La definición de diferentes etiquetas lingüísticas puede ayudar en este proceso, y así permitir obtener el resultado de manera más intuitiva al usuario. De esta forma se han definido tres tipos de etiquetas: i) *Poco Expandido* dónde los grados de similitud son mayores o iguales que 0.7, ii) *Medianaemente Expandido* donde el umbral está en 0.5, iii) *Muy Expandido* con un umbral de 0.3.

En la 6 se muestra el resultado de haber aplicado la etiqueta *Poco Expandido* sobre el conjunto de etiquetas resultantes mostrado en los ejemplos anteriores.

Thesaurus	Etiquetas Poco Expandidas
<i>WordNet</i>	Car/1, [railcar, railway_car, railroad_car]/0.72, Auto/1, [car, automobile, machine, motorcar]/0.88, automotive_vehicle/0.6, compact/1, compact_car/0.82, environment/1
<i>Wikipedia</i>	Car/1, [Cars, Motor car, Motorcar, Automobiles, Motor Car, Ottomobile, Automobiles, Motorization, Motorisation, Cardoor, Autos, Automobile, Passenger vehicle, A car, Passenger Cars, Car Automobile, Automobilism, Passenger vehicles, Motor-car]/ 0.72, Auto/1, [Automatic memory allocation]/0.77, Compact/1, environment/1, etc./0.72

Tabla 6: Ejemplo de conjunto *Poco Expandido*

6 Conclusiones

El principal objetivo de este trabajo ha sido mejorar el proceso de identificación semántica. Para ello se propone un novedoso proceso que incluye el tratamiento de datos difusos, de tal forma que el conjunto de etiquetas se incremente y cada elemento sea evaluado de acuerdo con su similitud con el conjunto de etiquetas inicial. De esta forma se proporciona al sistema la capacidad de seleccionar aquellas etiquetas que satisfagan las necesidades del usuario sin excluir ningún sentido significativo. Con el uso de esta propuesta, los sentidos ahora no son excluidos sino ponderados y pueden utilizarse en el proceso de identificación de estructuras semánticas en ontologías en las fases posteriores.

La lógica difusa ha proporcionado las herramientas necesarias para representar y ponderar la información relacionada con un conjunto de etiquetas concreto. Cada etiqueta se representa como un conjunto difuso donde se evalúa con precisión la pertenencia de la misma al conjunto de etiquetas inicial.

Por otro lado, *WordNet* proporciona una representación de información muy eficiente, donde se devuelve el tipo de las palabras, las relaciones entre ellas, una finita y bien definida jerarquía de categorías, etc. Sin embargo esta estructura jerárquica de categorías no funciona bien para buscar la relación semántica entre los elementos de la folksonomía, dado que estas estructuras contienen, a menudo, términos que no están relacionados. Como resultado, las folksonomías incluyen términos pertenecientes a diferentes contextos o tienen relaciones de similitud muy bajas.

El catálogo de la *Wikipedia* incluye un gran número de relaciones entre términos, y por lo tanto, grados de similitud elevados. La consecuencia es la opuesta al caso anterior, *Wikipedia* permite al sistema encontrar numerosos contextos donde un conjunto de etiquetas pueden estar relacionadas. Sin embargo, estos resultados son desventajosos dado que el amplio número de sentidos que *Wikipedia* proporciona, hacen el proceso de desambiguación muy difícil.

Por otro lado, el proceso de desambiguación utilizando únicamente hiperónimos no funciona demasiado bien en el caso de las folksonomías dada su naturaleza. Un mecanismo que tenga en cuenta otros términos relacionados como sinónimos, términos de la definición, términos relacionados, etc. pueden mejorar la identificación del contexto y reducir el número de sentidos obtenidos.

El resultante conjunto de etiquetas con sus grados de similitud asociados proporcionan al sistema la posibilidad de elegir aquellas que se ajusten a los requerimientos exigidos por el usuario. Estos gra-

dos también se utilizarán en las fases siguientes donde los grados de similitud se utilizarán para encontrar términos enriquecidos semánticamente. Finalmente, las etiquetas lingüísticas ayudan a escoger el conjunto de etiquetas final dado que dichos conjuntos se representan utilizando expresiones del lenguaje natural.

Agradecimientos

Esta investigación ha sido financiada por los proyectos de la consejería Andaluza de Ciencia y Tecnología TIC1570, TIC1433, TIC03175 y por el proyecto del ministerio con referencia TIN2008-02066. Además se agradece especialmente al profesor Enrico Motta, Sofia Angelotou y al KMI (Knowledge Media Center) en UK, lugar donde se ha realizado esta investigación, por su colaboración en este trabajo.

Referencias

- [1] Sofia Angeletou. Semantic enrichment of folksonomy tagspaces. In *ISWC '08*.
- [2] H. Dubois and H. Prade. Ranking fuzzy numbers in the setting of possibility theory. *Information Sciences*, 30:183–224, 1983.
- [3] Sanda M. Harabagiu, George A. Miller, and Dan I. Moldovan. WordNet 2 – a morphologically and semantically enhanced resource. In *Proc. SIGLEX 1999*, 1999.
- [4] Olena Medelyan, Catherine Legg, David Milne, and Ian H. Witten. Mining meaning from wikipedia, 2008.
- [5] R. Mihalcea. Using wikipedia for automatic word sense disambiguation. In *Proceedings of (NAACL)*, Rochester.
- [6] Ray Richardson, Alan F. Smeaton, and J. Murphy. Using wordnet as a knowledge base for measuring semantic similarity between words. Technical Report CA-1294, Dublin, Ireland, 1994.
- [7] Lucia Specia and Enrico Motta. Integrating folksonomies with the semantic web. pages 624–639. 2007.
- [8] M. Strube and S. Paolo-Ponzetto. Wikirelate! computing semantic relatedness using wikipedia. In *Proceedings of the 21st NCAI*, Boston, USA.
- [9] Thomas Vanderwal. Folksonomy coinage and definition, February 2007.
- [10] Zhibiao Wu and Martha Palmer. Verb semantics and lexical selection. In *32nd. Annual Meeting of the Association for Computational Linguistics*, pages 133–138, New Mexico State University, Las Cruces, New Mexico, 1994.