

ESCUELTA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

Máster Oficial en Lógica, Computación e Inteligencia Artificial

Trabajo Fin de Máster

Conciliación de Conceptos Formales en Sistemas de Etiquetado mediante Sistemas MultiAgente

Autor:

Jesús Giráldez Crú

Tutores:

Joaquín Borrego Díaz

Gonzalo A. Aranda Corral

Sevilla, 30 de junio de 2011.

Curso académico 2010/11. Primera Convocatoria

Índice general

In	trodu	cción.	3
	Inte	rnet, Datos e Información	4
	Mot	ivación y Objetivos	6
	Solu	ción Propuesta	7
	Estr	uctura de la memoria	8
1.	Esta	do del Arte.	11
	1.1.	Clasificación del contenido en Internet	11
	1.2.	El Etiquetado y la Web Semántica.	13
	1.3.	Sistemas de Etiquetado	14
	1.4.	El Problema de la Heterogeneidad	14
	1.5.	El Etiquetado y la Conciliación de Conceptos	16
2.	Aná	lisis Formal de Conceptos.	19
	2.1.	Contextos Formales	20
	2.2.	Intención y Extensión	21
	2.3.	Conceptos Formales	21
	2.4.	Retículos de Conceptos	22

	2.5.	Bases Stem	24
3.	Extr	acción de Conocimiento en Sistemas de Etiquetado.	27
	3.1.	Estructura de los Sistemas de Etiquetado	27
	3.2.	Contextos y Conceptos Formales	29
	3.3.	Relaciones entre usuarios	30
Bi	bling	rafía.	33

Índice de figuras

2.1.	Ejemplo de Contexto Formal	•		•			•		•		20
2.2.	Retículo de Conceptos										24

Introducción

En ocasiones, la búsqueda de información sobre ciertos temas en Internet se convierte en una ardua tarea, debido a que los resultados que encontramos no son demasiado relevantes. Esto se debe a numerosas razones. En general, se puede decir que los parámetros con los que ajustamos la búsqueda no termina de ser los más apropiados para conseguir los resultados que esperamos; bien porque no hemos sabido ajustar correctamente las palabras claves en nuestra búsqueda, y por tanto, los resultados obtenidos versan sobre otros temas distintos; o bien porque, aunque estas palabras claves sí sean relativamente correctas, los resultados obtenidos son más específicos o más generales de lo que buscamos, y por tanto estos resultados tampoco son conveniententes. Desde el punto de vista de búsquedas en contextos *semantizados* ¹, podemos decir más concretamente que si los resultados encontrados no son apropiados, es porque el conocimiento con el que hemos ajustado la búsqueda es incompleto, o incluso incorrecto.

Una posible solución para solventar, o minimizar este problema, podría ser el uso del conocimiento de otros usuarios, además del conocimiento propio del usuario, que ya se dispone. De esta forma, una búsqueda podría ser más completa, ya que el usuario no es el único que aporta conocimiento a los parámetros de dicha búsqueda, sino que éstos se enriquecen con el conocimiento de otros usuarios. Por tanto, el problema se traslada a encontrar relaciones entre usuarios y elegir qué conocimiento será usado para complementar el conocimiento propio de cada usuario.

En este trabajo se propone una solución formal a este problema, en el que se intenta complementar los diferentes conocimientos de cada usuario mediante implicaciones lógicas, que permitan establecer una serie de

¹Se entiende por contextos *semantizados* a aquellos que están marcados con ciertos metadatos semánticos.

reglas entre ellos. El conocimiento común que genera un par de usuarios es lo que llamamos Conciliación de Conocimiento, como se cita en [2]. Con este conocimiento conciliado se dispone de una herramienta que puede enriquecer semánticamente cualquier búsqueda. Igualmente, es interesante concebir este conocimiento conciliado desde el punto de vista de la navegación entre usuarios, o entre la información de los mismos, que es, al fin y al cabo, otra forma de buscar información en Internet. Este conocimiento conciliado entre dos usuarios representa un conocimiento aceptado por ambos, ya que está basado en implicaciones lógicas, que aportan un trasfondo formal a este conocimiento. Por tanto, este conocimiento conciliado es equivalente a un contexto común entre ambos usuarios. La navegación semántica entre estos usuarios pasa por un punto intermedio, que es precisamente este conocimiento conciliado.

En este trabajo se propone un método de conciliación del conocimiento en sistemas multiusuario, centrando el caso de estudio en los sistemas de etiquetado. Este método se basa en la utilización de cálculos basados en Análisis Formal de Conceptos (se utilizarán técnicas descritas en [6]), y ha sido implementado en una estructura de Sistema MultiAgente (SMA) en una plataforma JADE (consutar [27]).

Internet, Datos e Información.

Hoy en día, Internet se ha convertido en una herramienta básica en nuestro día a día. La generación de contenido se produce de una forma masiva, y el volumen de contenido generado crece en forma exponencial; es decir, en cada momento, en Internet se genera una cantidad de contenido mucho mayor que el contenido que se ha generado previamente. Este crecimiento exponencial de Internet es una realidad, y existen algunos problemas que derivan de él, como la gestión de esa cantidad tan grande de contenido.

En los últimos tiempos, este crecimiento del contenido se ha potenciado debido al crecimiento de la Web 2.0, que aporta potentes tecnologías para compartir este contenido con el resto de usuarios (ver [2] y [3]). Ejemplo de ello, es el gran número de herramientas sociales (redes sociales, sistemas de etiquetado, sistemas de compartición de ficheros, etc.), en los que se establece cierta conexión entre los usuarios, y en los que, cada vez más, se integran contenidos multimedia junto con el formato de texto plano que siempre ha existido. El hecho de que estos contenidos multimedia se conviertan en uno de los pilares básicos de esta Web 2.0, trae consigo ciertas consecuencias:

- La generación de contenido se vuelve cada vez más masiva, debido al hecho de que el contenido multimedia es más amigable para el usuario. Los avances tecnológicos han propiciado este hecho, debido a que las fronteras de Internet llegan cada vez a más usuarios, y las velocidades de acceso a éste son cada vez más rápidas, y esto hace viable esta generación tan masiva de este contenido, así como su visualización posterior.
- El contenido multimedia es, si cabe, más difícil de gestionar y procesar que cualquier contenido en texto plano. Por una parte, el texto plano responde a unos formatos más restringidos, por lo que se podría considerar casi estándar. Sin embargo, el contenido multimedia se clasifica en diferentes tipos (audio, vídeo, imágenes, etc.), en los que cada tipo se clasifica en numerosos formatos. Por tanto, las tareas de gestión son bastante más complicadas, puesto que hay que enfrentarse a diferentes tareas en función del tipo de contenido y del formato del mismo. Por otra parte, el procesado de este contenido también es más complicado. Una búsqueda en un texto plano es, a grosso modo, una tarea de coincidencia de cierta secuencia de caracteres; aunque hoy en día, estas tareas son mucho más complejas que esa simple idea. Sin embargo, en cualquier caso, es una tarea con un coste computacional bajo, y por tanto, rápida. En cambio, en el ámbito del contenido multimedia esta tarea es mucho más complicada. Un símil equivalente podría ser buscar cierta forma en una imágen, cierto sonido en un audio o cierto fotograma en un vídeo. Estos ejemplos son casos de estudio actualmente, y la carga computacional de ellos es de un orden muy superior que el ejemplo de búsqueda en texto plano. En conclusión, el procesado de este contenido es más complicado.

Otra característica de estas herramientas 2.0 es la posibilidad de relación social que la propia herramienta ofrece al usuario para que se relacione con otros usuarios. Al igual que el crecimiento del contenido, se produce también un crecimiento en cuanto al número de relaciones entre usuarios, que cada vez más, van estableciendo relaciones con otros usuarios, lo que, dependiendo del sistema, permite acceder a su contenido, compartir contenido común, etc.

De una forma abstracta, esta idea puede entenderse como un concepto de grafo, en el que se represente todo este contenido, mediante nodos, y todas estas relaciones entre usuarios, mediante aristas. Por una parte, a medida que se produce un crecimiento en cuanto a la generación de contenido, encontraríamos un crecimiento en el número de nodos del grafo. Por otra, el crecimiento de las relaciones sociales en el ámbito de la Web 2.0, equivale a un crecimiento del número de aristas de este grafo abstracto que representa todo el conjunto de la web (o simplificadamente, representa a un sistema concreto: un sistema de etiquetado concreto, una red social concreta, etc.).

En conclusión, el crecimiento de Internet hoy en día lleva consigo:

- 1. Una generación masiva de contenido, en gran parte multimedia.
- 2. Un problema de procesado de este contenido, debido al gran volumen de los mismo. Y un problema asociado al contenido multimedia, debido a la propia naturaleza de este contenido.
- 3. Un conjunto de relaciones entre usuarios de gran volumen.

Es obvio que un trabajo de procesamiento o gestión sobre este compendio de contenido es inviable si se realiza a nivel global. En esta línea, hablaremos de **datos** cuando, en casos como el anteriormente descrito, se disponga de una gran cantidad de contenido pero que es difícil de procesar, analizar y tratar. Sin embargo, es posible añadir cierta metainformación sobre estos datos, para conseguir que estas tareas se puedan llevar a cabo con relativa viabilidad. Con esta metainformación, todo el contenido se puede clasificar más fácilmente, y esta clasificación posibilita un procesado posterior a nivel global que sea viable en cuanto a tiempos de ejecución o carga computacional. Esta metainformación será, en su mayoría, ciertos metadatos semánticos que se añadan a los datos. En el caso de contenido marcados semánticamente, hablaremos de **información**.

Motivación y Objetivos.

En la sección anterior, se introduce la posibilidad del marcado semántico (mediante metadatos añadidos al contenido), para convertir los *datos* en

información. Una vez hecho esto, las posibilidades de procesado a posteriori de esta información son innumerables.

En este trabajo se propone el uso de los Sistemas de Etiquetado, en los que encontramos información marcada semánticamente, con etiquetas, como se verá en el capítulo 3; para llevar a cabo un procesado semántico con el que obtener un conocimiento conciliado entre varios usuarios. Inicialmente se propone conseguir este conocimiento conciliado por pares de usuarios que formen parte del sistema, si bien, en un futuro se puede ampliar a un concepto superior de conocimiento conciliado global. La principal motivación de este trabajo es encontrar un método de conciliación que sea aplicable a sistemas de etiquetado de este tipo, aunque en un futuro se podría ampliar a otro tipo de sistemas. Estas conciliaciones representan un método de navegación entre usuarios, puesto que el conocimiento conciliado que se obtenga para cada par de usuarios, representará un conocimiento común que ambos usuarios aceptan, y por tanto, puede ser una herramienta útil de navegación entre ellos. En un futuro, este conocimiento conciliado podría ser usado en otras tareas como refinamientos de búsquedas, depuración semántica, etc.

Los principales objetivos de este trabajo son:

- Formalizar el conocimiento de los Sistemas de Etiquetado.
- Extraer los distintos elementos de AFC de estos sistemas.
- Implantar un método de conciliación a escala global para todo el conjunto de alguno de estos sistemas.
- Analizar los resultados y proponer soluciones futuras que aporten una funcionalidad clara y útil a estos sistemas.

Solución Propuesta.

La solución propuesta es un Sistema MultiAgente (SMA) implementado en JADE (ver [27]), en el que se extraiga el conocimiento de un caso práctico concreto, y calcule un conocimiento conciliado para este sistema. El caso práctico concreto que utilizaremos será un subconjunto de datos de Delicious (ver [23]). En el SMA, un conjunto de agentes, que representan a los usuarios del sistema, negociarán entre ellos para conciliar su conocimiento con otros usuarios, en función de su disponibilidad y de la prioridad que tengan en realizar otras conciliaciones con terceros. A medida que avance la ejecución del SMA, se irán produciendo múltiples conciliaciones de forma paralela, de forma que el conocimiento conciliado sea cada vez mayor; y por tanto, se enriquecerá este conocimiento común. Este conocimiento conciliado será almacenado, de forma que se puedan realizar cálculos o extraer conclusiones posteriormente.

En el capítulo ?? se comenta la estructura, componentes y funcionamientos de este SMA. En los capítulos anteriores se comentarán algunas técnicas y herramientas necesarias para llevarlo a cabo, o ciertas formalizaciones previas que justifican el trasfondo formal de este trabajo.

Estructura de la memoria.

Esta memoria se estructura en varios capítulos, con la siguiente distribución de los temas trabajados:

- Capítulo 1. Se presenta el Estado del arte en el que se detalla los diferentes modelos de clasificación de la información en Internet, y se plantean las diferentes problemáticas de cada uno de estos modelos.
- Capítulo 2. Se formalizan las nociones mátematicas de Análisis Formal de Conceptos (AFC) que sirven como soporte formal en este trabajo.
- Capítulo 3. Se introducen los Sistemas de Etiquetado, que serán los casos de estudio concretos de este trabajo, viendo su estructura y la extracción formal de los conceptos de AFC vistos en el capítulo anterior.
- Capítulo ??. Se explica el Algoritmo de Conciliación, que permite conciliar el conocimiento de un par de usuarios.
- Capítulo ??. Se describen las características del Conjunto de Datos con el que llevaremos a cabo los procesos de prueba empíricamente.
- Capítulo ??. Se explica el Sistema MultiAgente (SMA) que se plantea como solución en este trabajo: comportamiento, estructura y componenentes.

- Capítulo ??. Se detallan los resultados obtenidos en una prueba empírica llevada a cabo, y se analizan estos resultados.
- Capítulo ??. Finalmente, se comentan las conclusiones obtenidas, así como los puntos de trabajo futuro.
- Bibliografía.

Capítulo 1

Estado del Arte.

En este capítulo, se introducirá de forma general el estado del arte sobre la clasificación de los datos y la información en Internet, se introducirán los Sistemas de Etiquetado como herramienta de marcación semántica de ciertos datos, se tratarán los diferentes problemas asociados a esta marcación semántica, y finalmente se introducirá la Conciliación de Conceptos como método formal para encontrar conocimiento común entre usuarios.

1.1. Clasificación del contenido en Internet.

La organización del contenido en Internet se realiza a menudo a través de marcas en el contenido con términos descriptivos, también llamados palabras clave o etiquetas. Esta organización permite al usuario ciertas tareas futuras como navegación, filtro o búsquedas. Esta forma de organización no es, ni mucho menos, nueva; sin embargo, sus creadores en el ámbito de Internet, le acuñaron el nombre de **etiquetado**, con el que actualmente se conoce, y que está ganando mucha popularidad en Internet.

Históricamente, la asignación de estas palabras claves para la organización de colecciones o librerías ha sido tarea de una persona experta en la materia que, bajo su criterio, es quién elige el conjunto de palabras clave y la asignación de las mismas a los diferentes recursos, como se explica en [15]. En cambio, el etiquetado colaborativo permite que cualquier usuario participe en esta tarea, de forma que cada recurso pueda ser etiquetado

con palabras claves por cualquier usuario y por cualquier palabra clave que éstos elijan.

En cuanto a sistemas de clasificación de contenido, se van a distiguir dos tipos: taxonomías y folksonomías.

Las taxonomías son sistemas de clasificación de contenido basados en la categorización de los recursos (o de los elementos, en general). Esta definición se puede entontrar en [17], [11] o [7]. Estas categorías deben formar un conjunto completo¹ y disjunto; de forma que todos los recursos tengan una categoría con la que pueda ser categorizado, y que además, se elimine (o en todo caso, se minimice) las posibles ambigüedades entre varias categorías. La única excepción en este conjunto disjunto se produce debido a la posibilidad de establecer una jerarquía en las categorías, de forma que se establezca una especialización de las mismas. Estos sistemas son, por tanto, jerárquicos y exclusivos.

En la otra cara de la moneda, encontramos otros sistemas no jerárquicos y no exclusivos: las folksonomías, cuya definición está aproximadamente consensuada. Esta definición se puede consultar en [17], [11], [7], [8] o [12]. En este tipo de clasificación, cada recurso es marcado con una serie de etiquetas, que aportan a estos recursos una serie de metadatos semánticos. En [17], se comentan los distintos sistemas que incorporan folksonomías como herramientas; y se discute las diferentes restricciones de cada uno de ellos: conjunto de etiquetas restringido o abierto, etiquetación a recursos propios o globales del sistema, etc. En [7], se explica la diferencia entre estos sistemas y los anteriores, cuya principal diferencia es la opción de trabajar con interesecciones entre las diferentes categorías que aparecen, consiguiendo una mejor organización de los recursos y evitando posibles duplicidades de los mismos.

Esta diferencia hace que estos sistemas de etiquetado se hayan convertido en una herramienta muy útil en gran cantidad de sistemas que funcionan en Internet, gracias a la versatilidad que ofrecen al usuario, y además, al ser una herramienta colaborativa, se gestiona gracias a las aportaciones de todos los usuarios, sin necesidad de que una persona experta se encargue de esta tarea. Prueba del auge de estas herramientas, es el enorme número de trabajos que hay sobre ellos desde diversos puntos de vista, como el análisis (estadístico) de los modelos de etiquetación realizado en [7], la recuperación y navegación por la información realizado en [9] o [10], o

¹Entiéndase *completo* para un ámbito concreto.

el análisis de redes sociales realizado en [14] o [5].

1.2. El Etiquetado y la Web Semántica.

Como se ha comentado en la sección anterior, las folksonomías dotan a cualquier sistema en el que se implanten, de una herramienta colaborativa con la cual organizar el contenido. Además, debido a sus características no jerárquicas y no exclusivas, se evitan problemas derivados de otro tipo de clasificaciones, como las taxonomías. Por último, el carácter colaborativo de estas folksonomías han potenciado su popularidad en la Web 2.0, dónde el usuario es el protagonista y *creador* de la propia web (redes sociales, blogs, sistemas de etiquetado, sistemas para compartir ficheros multimedia, etc.).

Por otra parte, estas folksonomías ofrecen una buena alternativa a la web semántica o las ontologías, como se explica en [7]. Es cierto que estas opciones generan una mayor cantidad de información semántica referente a los recursos del sistema. En cambio, los sistemas de etiquetado, como las folksonomías generan una información semántica muy pobre, que además es ambigua, como se comenta en [2] y [7]; en comparación con un sistema basado en ontologías, por ejemplo. En éste último, el marcado semántico es completo, y el procesado semántico (que es la motivación de este trabajo), sería más fácil.

Sin embargo, el concepto de web semántica puede chocar en algunos aspectos con el carácter colaborativo propio de estas folksonomías de los sistemas de etiquetado. Precisamente, uno de los puntos más enriquecedores de estos sistemas es que la información semántica la generan los usuarios, para ser procesada y/o gestionada por ellos mismos. Si bien no existe una formalización en cuanto a la semantización de la información, el papel que juegan los usuarios es crucial para haber conseguido un desarrollo tran grande (frente a otros sistemas, basados en ontologías por ejemplo, que son propios de otro tipo de casos de estudio).

En cambio, como se introdujo en el capítulo anterior, es necesario cierto procesado de la información, para obtener un mecanismo que formalice la semántica de estos sistemas. Las etiquetas serán los elementos que proporcionen esta formalización.

1.3. Sistemas de Etiquetado.

En [17], se tratan diversos Sistemas de Etiquetado y las diferentes características de cada una de sus arquitecturas.

De forma general, diremos que un Sistema de Etiquetado está compuesto por Recursos, Etiquetas y Usuarios. En nuestro caso de estudio, los recursos estarán disponibles para cualquier usuario, de forma que un recurso pueda estar etiquetado por múltiples usuarios. El conjunto de recursos es abierto; es decir, cualquier usuario puede añadir nuevos recursos al sistema, por lo que este conjunto está continuamente creciendo. Igualmente ocurre en el caso del conjunto de etiquetas, de forma que una etiqueta puede ser usada por múltiples usuarios, y cada usuario puede añadir las etiquetas que desee al sistema.

En el capítulo 3, se verá de forma detallada las características de los Sistemas de Etiquetado que se tratan en este trabajo. El sistema elegido ha sido Delicious (ver [23]), dónde los recursos son URLs que los usuarios guardan, y las etiquetas son campos de texto plano. De esta forma, una URL puede haber sido introducida en el sistema por dos usuarios distintos, por lo que este recurso es *compartido* para ambos (aunque tengan etiquetas distintas para cada uno de ellos). De igual forma, dos usuarios pueden tener etiquetas comunes (o no), que hayan introducido en el sistema.

1.4. El Problema de la Heterogeneidad.

El Etiquetado lleva consigo un problema implícito: la heterogeneidad semántica. En [7] se comentan algunos aspectos semánticos y cognitivos sobre el proceso de clasificación. Básicamente, se establecen tres problemas principales: polisemia, sinonimia y variaciones del nivel básico.

La polisemia es el problema que se produce debido a que a una palabra tiene varios significados. Al realizar una búsqueda sobre este término, un sistema no sería capaz de distinguir estos significados diferentes; y por tanto, se obtienen más resultados de los esperados, ya que muchos de ellos pertenecen a alguno de los otros significados del término. Esto deriva en un problema de incorrección, pues se obtienen resultados que no deberían obtenerse. Además, este problema es equivalente al producido por homonimia (en concreto, por homografía).

La sinonimia es el problema que se produce debido a que diferentes palabras tienen el mismo significado. Al realizar una búsqueda sobre alguno de estos términos, el sistema sólo será capaz de devolver aquellos resultados que coincidan con el término con el que se ha hecho la búsqueda, obviando todos los resultados de los otros términos sinónimos a éste. Esto deriva en un problema de incompletitud, pues no se obtienen todos los resultados esperados. Además de la definición de sinonimia en lingüística, este problema es equivalente a otros problemas derivados de la escritura propia del usuario. Por ejemplo, las etiquetas *gato* y *gatos* son sinónimos en la práctica, al igual que *para-leer* y *para-leer*.

Finalmente, existe un problema derivado de las variaciones del nivel básico. En [18] se explica que el nivel básico asocia un problema que surge al existir diferentes términos que describen un único objeto en el rango desde lo más general a lo más específico; el nivel básico es precisamente el que se asocia mayormente con las interacciones humanas. Por ejemplo, animal, perro y dálmata corresponde a la misma categoría, desde la más general a la más concreta. En el ejemplo anterior, para la mayoría de la gente, el nivel básico se establecería en el término perro. Sin embargo, existen algunas variaciones sistemáticas en lo que cada individuo considera el nivel básico. Y estas variaciones plantearían este problema, cuya principal consecuencia es incompletitud, aunque también incorrección.

No hay que olvidar que el proceso de etiquetado es básicamente un proceso de interpretación, en el que cada usuario categoriza el contenido en función de una serie de parámetros personales que, en algunos casos, no corresponden con los parámetros del resto.

En [2], se describen los problemas anteriores de una forma más formal, en el ámbito del campo de trabajo de los sistemas de etiquetado. Así pues, se tiene:

- Hetereogeneidad del Conocimiento dependiente del Contexto (CDKH):
 Es la limitación producida debido a que una misma etiqueta se refiera a diferentes términos.
- Ambigüedad Clásica (CA): Es la limitación producida por las ambigüedades heredadas del lenguaje natural y la elección del nivel básico ([18]) por parte de cada usuario.

Por una parte, CA no representa un problema crítico, puesto que el propio objeto (la URL) puede ayudar a desambiguar el significado del término. De hecho, una contextualización de las etiquetas en forma de grafo, puede ayudar a distinguir estos diferentes significados. Sin embargo, CDKH está asociado a estructuras de conceptos que los usuarios no representan en el sistema, aunque métodos de FCA pueden extraer, y por tanto, este problema debe ser tratado con más cautela, y corregido con métodos formales.

Finalmente, hay que comentar que el problema de CDKH es debido a la finalidad con la que se use cierta etiqueta. En [7], se definen diferentes tipos de etiquetas:

- Identificación del tema que trata: los temas que la URL trata. Por ejemplo *programación*.
- Identificación de lo que es: lo que es exactamente la URL. Por ejemplo libro o artículo.
- Identificación del propietario o el creador de la URL. Por ejemplo *Shakespeare*.
- Refinación de las categorías. En ocasiones, ciertas etiquetas se usan para complementar la etiquetación de las ya existente. Por ejemplo, números.
- Identificación de características: adjetivos, que son la opinión del usuario sobre la URL. Por ejemplo *gracioso* o *estúpido*.
- AutoReferencia: del propio usuario. Por ejemplo *mi-libro*.
- Etiquetas de organización: ciertas etiquetas se pueden usar para agrupar ciertas URL con un fin común. Por ejemplo para-leer o trabajopendiente.

1.5. El Etiquetado y la Conciliación de Conceptos.

Como se ha visto a lo largo del capítulo, en este trabajo se han elegido las folksonomías, y más concretamente los sistemas de etiquetado, como

sistemas en los que existen un entorno semantizado, debido a la semántica que las etiquetas aportan a cada objeto del sistema.

Se ha elegido esta forma de clasificar debido a las ventajas que presenta frente a otros métodos, como taxonomías u ontologías. Sin embargo, se ha explicado los problemas asociados de heterogeneidad y ambigüedad de los términos.

Debido a estos problemas, se han elegido métodos basados en Análisis Formal de Conceptos (AFC) para minimizar las consecuencias de estas ambigüedades semánticas, y establecer un marco de trabajo formal con el que realizar las diferentes operaciones.

La Conciliación de Conceptos es el proceso por el cual, aplicando técnicas de AFC, se van a extraer conceptos formales de los sistemas de etiquetado, tarea que se realizará para cada usuario. Estos conceptos formales nos permitirán calcular un conocimiento común entre los usuarios, que podrá ser usado posteriormente en diferentes tareas: navegación (semántica) entre usuario, refinamiento de búsquedas (semánticas), etc...

Capítulo 2

Análisis Formal de Conceptos.

En este capítulo se formalizan algunas nociones sobre **Análisis Formal de Conceptos** (AFC), que serán necesarias para comprender algunas de las técnicas que se usan en el trasfondo de este trabajo. Las definiciones básicas en AFC son las de Contexto Formal y Concepto Formal. El adjetivo *formal* se usa simplemente para indicar que existe un trasfondo matemático en ambas definiciones, y por tanto, su significado difiere de aquellos que tienen en el lenguaje natural.

A lo largo de todo el capítulo, se incluirá un ejemplo que ilustre todas las definiciones que se van incluyendo. Se ha escogido un ejemplo de pequeño tamaño, para facilitarle al lector la lectura y comprensión del proceso; así como la representación de los resultados que se van obteniendo. Sin embargo, estas técnicas que var a ser definidas en este capítulo, serán usadas en el desarrollo de este trabajo con muestras de un tamaño mucho mayor; y por tanto, los resultados serán igualmente de mayor tamaño, por lo que será mucho más difícil representarlos gráficamente.

Se recomienda consultar [6] como recurso bibliográfico, en el que se formaliza las definiciones de *conjunto*, *conjunto* ordenado o retículos completos, que utilizaremos a lo largo de todo el capítulo. Además, se incluyen las demostraciones de todos los teoremas aquí presentados.

2.1. Contextos Formales.

DEFINICIÓN 1 Un Contexto Formal $\mathcal{K}:=(G,M,I)$ está compuesto por dos conjuntos G y M y una relación I entre G y M ($I\subseteq \{G\times M\}$). Los elementos de G son los **objetos** (formales) y los elementos de M son los **atributos** (formales) del Contexto. La relación entre un objeto $g\in G$ y un atributo $m\in M$, se expresa como $(g,m)\in I$ (o gIm), y se lee como "el objeto g tiene el atributo m".

En el siguiente ejemplo, se ilustra un pequeño contexto sobre peces (objetos) y los hábitats en los que viven (atributos):

- Un conjunto de Objetos *G*, que son los peces que forman parte de este contexto: carpa, escatófagus, sargo, dorada y anguila.
- Un conjunto de Atributos M, que son las posibles características de los objetos anteriores, y que en este ejemplo expresan los posibles hábitats de los peces (que no tienen porqué ser exclusivos): fluvial, litoral y océano.
- Una relación *I*, entre *G* y *M*, expresada en la tabla 2.1.

	Océano	Litoral	Fluvial
Sargo	X	X	
Dorada	X	X	
Carpa			X
Anguila	X	X	X
Escatófagus		X	X

Figura 2.1: Ejemplo de Contexto Formal

Como se ve en el ejemplo anterior, un contexto puede ser expresado fácilmente en una tabla de doble entrada, dónde las filas se corresponden con los objetos, y las columnas con los atributos. La relación entre ambos se expresa marcando las casillas correspondientes, de forma que si la casilla de la fila g y columna m está marcada, significa que el objeto g tiene el atributo m; es decir que $(g,m) \in I$, (o gIm). En caso contrario, la casilla quedará sin marcar.

De esta forma, se ve como un objeto puede tener múltiples atributos, y un atributo puede pertenecer a varios objetos. Denotaremos $\{...\}_{\mathcal{X}}$ a un subconjunto de elementos del conjunto \mathcal{X} ; es decir que $\{...\}_{G}$ es un subconjunto de objetos, y $\{...\}_{M}$ es un subconjunto de atributos.

2.2. Intención y Extensión.

Se definen los conjuntos A' y B', de Intención y Extensión de un Contexto $\mathcal{K} := (G, M, I)$, como:

DEFINICIÓN **2** Para un contexto $\mathcal{K} := (G, M, I)$ y un conjunto $A \in G$ de objetos, se define la **Intención** como el conjunto A'; que es precisamente el conjunto de atributos comunes a los objetos de A:

$$A' := \{ m \in M | gIm \ \forall g \in A \}$$

DEFINICIÓN 3 Para un contexto $\mathcal{K} := (G, M, I)$ y un conjunto $B \in M$ de atributos, se define la **Extensión** como el conjunto B'; que es precisamente el conjunto de objetos comunes a los atributos de B:

$$B' := \{ g \in G | gIm \ \forall m \in B \}$$

Veamos algunos ejemplos de estas operaciones sobre nuestro contexto ejemplo (figura 2.1) para ilustrar estas operaciones:

- (1): { fluvial, oc'eano }'_M = { anguila }_G
- (2): $\{ carpa, sargo \}'_G = \emptyset_M$
- (3): $\{ escat\'ofagus, sargo \}'_G = \{ litoral \}_M$

El ejemplo (1) es un ejemplo de extensión de atributos, mientras que los ejemplos (2) y (3) son ejemplos de intención de objetos.

2.3. Conceptos Formales.

DEFINICIÓN 4 Un Concepto Formal de un Contexto $\mathcal{K} := (G, M, I)$ es un par (A, B) con $A \subseteq G$, $B \subseteq M$, tal que A' = B y B' = A, dónde A' y B' son respectivamente la extensión y la intención del concepto (A, B).

Veamos algunos ejemplos para identificar qué es y qué no es un concepto sobre el contexto de nuestro ejemplo (figura 2.1):

- (1): ({sargo, dorada, anguila}, {litoral, océano}) es un concepto; porque:
 - $\{sargo, dorada, anguila\}' = \{litoral, océano\}$
 - $\{litoral, océano\}' = \{sargo, dorada, anguila\}$
- (2): $(\emptyset_G, \emptyset_M)$ no es un concepto; porque:
 - $\bullet \ \emptyset'_G = M \neq \emptyset_M$
- (3): (G, \emptyset_M) es un concepto; porque:
 - No existe ningún atributo común a todos los objetos; es decir:
 - \bullet $G' = \emptyset_M$
 - $\bullet \ \emptyset_M' \ = \ G$
- (4): (\emptyset_G, M) no es un concepto pues:
 - Aunque $\emptyset'_G = M$,
 - se tiene $M' = \{anguila\} \neq \emptyset_G$

Los ejemplos (1) y (3) son conceptos porque la extensión del subconjunto de atributos es precisamente el subconjunto de objetos, y la intención del subconjunto de objetos es precisamente el subconjunto de atributos. Los ejemplos (2) y (4) no son conceptos porque no se cumplen alguna de estas dos condiciones.

DEFINICIÓN 5 Si (A_1, B_1) y (A_2, B_2) son conceptos de un contexto, (A_1, B_1) es un **subconcepto** de (A_2, B_2) si $A_1 \subseteq A_2$ (o equivalentemente si $B_2 \subseteq B_1$). En este caso, (A_2, B_2) es un **superconcepto** de (A_1, B_1) , y se escribe como $(A_1, B_1) \leq (A_2, B_2)$. La relación \leq se llama **orden de jerarquía** (o simplemente **orden**) de los conceptos.

2.4. Retículos de Conceptos.

En la sección anterior se ha visto que los conceptos pueden ser ordenados jerárquicamente. Esto permite crear un conjunto ordenado de todos ellos. DEFINICIÓN 6 El conjunto ordenado de todos los conceptos de un contexto $\mathcal{K} := (G, M, I)$, se denota como $\mathcal{B}(G, M, I)$, y se llama **retículo de conceptos**.

En el ejemplo que estamos ilustrando, se obtienen los siguientes conceptos:

- $C_0 := (\{anguila, carpa, dorada, escat\'ofagus, sargo\}_G, \emptyset_M)$
- $C_1 := (\{anguila, carpa, escat\'ofagus\}_G, \{fluvial\}_M)$
- $C_2 := (\{anguila, dorada, escat\'ofagus, sargo\}_G, \{litoral\}_M)$
- $C_3 := (\{anguila, escat\'ofagus\}_G, \{litoral, fluvial\}_M)$
- $C_4 := (\{anguila, dorada, sargo\}_G, \{litoral, océano\}_M)$
- $C_5 := (\{anguila\}_G, \{litoral, fluvial, océano\}_M)$

Y se obtienen las siguientes relaciones de orden entre los conceptos:

- $C_0 > C_1$
- $C_2 > C_4$
- $C_0 \ge C_2$
 - $C_3 \ge C_5$
- $C_1 \geq C_3$
- $C_4 \ge C_5$
- $C_2 > C_3$

El retículo de conceptos se puede representar fácilmente en una estructura de grafo, de forma que cada nodo se corresponderá con cada uno de los conceptos que se obtienen del contexto que se representa. Por tanto, habrá tantos nodos como número de conceptos. La relación de orden se expresará en la dirección vertical del grafo, de forma que si un concepto es subconcepto de otro, el superconcepto estará posicionado más arriba que el subconcepto. Por último, y para evitar repeticiones innecesarias en la escritura de los objetos y los atributos, sólo se expresarán una única vez. Los objetos se expresarán en el nodo más bajo posible, de forma que todos los nodos superiores a éste, también contienen a dicho objeto. Los atributos se expresarán en el nodo más alto posible, de forma que todos los nodos inferiores a éste, también contienen a dicho atributo. En la figura 2.2, vemos el grafo obtenido de este retículo de conceptos.

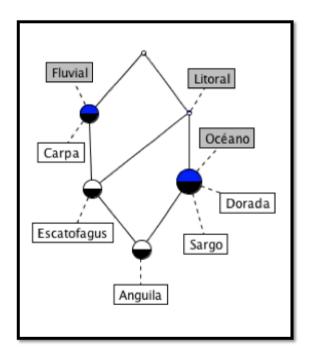


Figura 2.2: Retículo de Conceptos

2.5. Bases Stem.

DEFINICIÓN 7 Para un contexto $\mathcal{K} := (G, M, I)$, Una **implicación** entre atributos es una expresión de la forma $Y_1 \to Y_2$, dónde $Y_1, Y_2 \subseteq M$; es decir, son subconjuntos de atributos.

El objetivo es obtener un conjunto de relaciones entre los atributos, para que, usando operaciones lógicas, se pueda trabajar con el contexto dado.

A continuación se definen una serie de características sobre las implicaciones. Sea $\mathcal L$ un conjunto de implicaciones; se tiene:

- 1. $\mathcal{L} \models Y \to Z \ (Y \to Z \text{ es consecuencia} \text{ de } \mathcal{L}) \text{ si para todo } T \subseteq M, \text{ si } T \text{ respeta } \mathcal{L} \text{ entonces } T \text{ respeta } Y \to Z.$
- 2. \mathcal{L} es **cerrada** si contiene a toda implicación $Y \to Z$ que es consecuencia de \mathcal{L} .
- 3. $\mathcal L$ es **completo** si toda implicación válida en (G,M,I) es consecuencia de $\mathcal L$:

2.5. Bases Stem. 25

$$(G, M, I) \models Y \to Z \implies \mathcal{L} \models Y \to Z$$

4. \mathcal{L} es **no redundante** si ninguna implicación de \mathcal{L} es consecuencia del resto:

$$Y \to Z \in \mathcal{L} \implies \mathcal{L} \setminus \{Y \to Z\} \not\models Y \to Z$$

DEFINICIÓN 8 Se define como Base Stem al conjunto completo e irredundante de implicaciones de un contexto $\mathcal{K} := (G, M, I)$.

El problema se reduce, pues, a obtener una Base Stem para un contexto dado. Para ello, se hará uso del cálculo implicacional basado en las Reglas de Amstrong:

$$R_1: \frac{X \to Y}{X \cup Z \to Y}$$
 $R_3: \frac{X \to Y, Y \cup Z \to W}{X \cup Z \to W}$

Con este conjunto de reglas, se puede obtener el siguiente cálculo lógico:

 $\mathcal{L} \vdash L \iff L$ se prueba mediante las reglas de Amstrong a partir de \mathcal{L} .

TEOREMA 1 Si \mathcal{L} es un conjunto Stem para un contexto formal \mathcal{K} , entonces \mathcal{L} proporciona una teoría implicacionalmente completa para ese modelo; es decir:

$$\mathcal{L} \vdash L \Longrightarrow \mathcal{K} \models L$$

DEFINICIÓN 9 $P \subseteq M$ es una **pseudointención** de un contexto $\mathcal{K} := (G, M, I)$ si se verifica que:

- $P \neq P''$
- Para toda pseudointención $Q \subset P$, se verifica que $Q'' \subseteq P$.

TEOREMA **2** El conjunto $\{P \to P''|P \text{ es pseudointención}\}$ es una Base Stem del Contexto.

En la práctica, únicamente se toman las implicaciones $P \to (P''-P)$, ya que $P \to P$ siempre es válido.

A cada pseudointención podemos añadirle un atributo de **soporte**, que indica el número de objetos que cumplen esta implicación. Este soporte será un entero mayor o igual a cero.

Veamos en nuestro ejemplo, la traza del algoritmo seguido para calcular las distintas pseudointenciones del contexto. Dado que la definición de pseudointención es recursiva, iniciaremos con el conjunto vacío, e iremos recorriendo los subconjuntos de menor a mayor tamaño. De esta forma, las pseudointenciones calculadas previamente nos pueden ayudar a descartar otras pseudointenciones en el proceso. Por mejorar la visibilidad, únicamente se expresarán las iniciales de los atributos y de los objetos.

- $\emptyset_M = G = \emptyset_M$; por tanto, NO.
- $\{f\}'' = \{c, e, a\}' = \{f\}$; por tanto, NO.
- $\{l\}'' = \{e, s, d, a\}' = \{l\}; \text{ por tanto, NO.}$
- $\{o\}'' = \{s, d, a\}' = \{l, o\}$; por tanto, SÍ.
- $\{f, l\}'' = \{e, a\}' = \{f, l\}$; por tanto, NO.
- $\{f,o\}'' = \{a\}' = \{f,l,o\}$; sin embargo se tiene que $\{o\} \subset \{f,o\}$, y $\{o\}'' = \{l,o\} \not\subseteq \{f,o\}$; por tanto NO.
- $\{l, o\}'' = \{s, d, a\}' = \{l, o\}$; por tanto NO.
- $\{f, l, o\}'' = \{a\}' = \{f, l, o\}$; por tanto, NO.

Por tanto, la Base Stem obtenida (simplificando las implicaciones) en el ejemplo que hemos seguido a lo largo de todo el capítulo, es:

• $(\{oceanico\} \rightarrow \{litoral\})$, con soporte 3 (puesto que 3 objetos la cumplen).

Capítulo 3

Extracción de Conocimiento en Sistemas de Etiquetado.

En este capítulo vamos a ver algunas características de los Sistemas de Etiquetado (SE), se enlazarán estas características con los elementos formales que se han visto en el capítulo anterior; y finalmente se comentarán las relaciones entre usuarios desde el punto de vista del Análisis Formal de Conceptos (AFC).

3.1. Estructura de los Sistemas de Etiquetado.

Existen multitud de Sistemas de Etiquetado. En [17] se citan un gran número de ejemplos, y se explica las diferentes estructuras de cada uno de ellos.

En este trabajo, vamos a centrarnos en Delicious ([23]) como Sistema de Etiquetado, y en la folksonomía que el sistema ofrece a partir de la actuación colaborativa de los usuarios. Tomaremos la definición de folksonomía dada en [10]:

DEFINICIÓN **10** Una **folksonomía** es una tupla $\mathcal{F} := (U, E, R, Y)$, dónde:

■ *U*, *E* y *R* son conjuntos finitos, cuyos elementos son, respectivamente, los usuarios, etiquetas y recursos del sistema. En sistemas como Delicious, los recursos son URLs y las etiquetas son campos de texto.

• Y es una relación ternaria entre ellos; es decir, $Y \subseteq \{U \times E \times R\}$, cuyos elementos se conocen como etiquetaciones.

Existe un consenso en cuanto a la estructura de este tipo de sistemas, como se puede ver en diferentes definiciones (como en [7] o [21]).

Como se ha comentado en capítulos anteriores, el sistema se va creando gracias a las aportaciones colaborativas de todos los usuarios, que, unido a las características del sistema (no jerárquico y no exclusivo), potencian su popularidad, lo que a su vez genera nuevos usuarios y nuevas aportaciones; el conjunto de datos está en continuo crecimiento. Sin embargo, el problema de heterogeneidad semántica en cuanto a la etiquetación que cada usuario realiza hace que no se produzca un consenso por parte de todos ellos. Por ello, emergen ciertos métodos de resolución de estos conflictos para un posterior procesado de todo el volumen de datos que el sistema posee, que puede llegar a ser muy útil en algunos casos.

En [7], se presentan pruebas empíricas sobre la actividad de los usuarios, expresadas en el número de etiquetas que cada usuario usa, la tasa de crecimiento de cada etiqueta, así como el uso de ciertas URLs y la etiquetación de las mismas.

En cuanto al número de etiquetas, existen comportamientos diferentes. Inicialmente, el número de etiquetas por usuario crece en todos los casos. Sin embargo, llegado un punto, existen usuarios que siguen añadiendo nuevas etiquetas, por lo que este número sigue creciendo; mientras otros estabilizan el número de etiquetas que usan, y todas las etiquetaciones que realizan, las hacen con algunas de las etiquetas ya utilizadas.

Por norma general, el uso de una etiqueta va creciendo a lo largo del tiempo, ya que siempre aparecen nuevos usuarios que comienzan a usar esta etiqueta. En el caso de los enlaces es diferente, ya que normalmente emerge un comportamiento en el cual se produce un momento puntual en el cual se producen muchas etiquetaciones de ese enlace de forma masiva, y luego vuelve a descender para seguir unas tasas más pequeñas. Este crecimiento puntual se suele producir o bien al inicio del ciclo de vida de dicho enlace en el sistema, o bien una vez que ha transcurrido un largo período.

3.2. Contextos y Conceptos Formales.

El estudios de los Sistemas de Etiquetado desde el punto de vista formal ha sido un campo de trabajo muy concurrido en los últimos tiempos, y podemos encontrar una gran variedad de propuestas en diversas publicaciones: en [1] se propone un método de extensión de atributos; en [21] se ofrece un método para minimizar los efectos de hetereogenidad semántica; en [19] se propone un método para convertir en ontologías estas folkonomías; etc...

En el capítulo 2 se han introducido las características de los Contextos formales. Estos contextos están formado por un conjunto G de objetos, un conjunto M de atributos, y una relación binaria I entre ambos.

Existe una analogía clara entre estos Contextos formales y los Sistemas de Etiquetado. De hecho, todo Sistema de Etiquetado puede transformarse en un conjunto de Contextos formales. Formalicémoslo:

TEOREMA 3 Toda folksonomía $\mathcal{F}:=(U,E,R,Y)$ es transformable en un conjunto de Contextos formales $\mathcal{K}_i:=(G_i,M_i,I_i)$ mediante una función α , de forma que:

$$\alpha(\mathcal{F}) = \bigcup_{k=1}^{size(U)} \mathcal{K}_k$$

La demostración es trivial estableciendo que $G_i \subseteq R$, $M_i \subseteq E$ e $I_i \subseteq (R \times E)$. Concretamente está relación binaria I_i se consigue fijando el valor de cada usuario en la relación ternaria Y; ésta es la razón por la que se obtienen size(U) contextos diferentes. Los conjuntos G_i y M_i están formados por los objetos y atributos que participan en la relación binaria obtenida anteriormente.

El teorema anterior nos permite transformar un Sistema de Etiquetado (o un subconjunto completo de éstos) en un conjunto de Contextos Formales. En total se obtienen tantos contextos como usuarios forman parte del Sistema de Etiquetado.

En el caso concreto de este trabajo, cada contexto estará formado por un conjunto de enlaces (objetos) y un conjunto de atributos (atributos), que es precisamente lo que cada usuario aporta de forma colaborativa al sistema de Delicious. Para cada uno de los contextos anteriores, se puede calcular fácilmente los Retículos de Conceptos que representan el contexto, y las Bases Stem asociadas a éstos.

3.3. Relaciones entre usuarios.

En la sección anterior se ha visto cómo extraer un conjunto de Contextos (formales) de un Sistema de Etiquetado. Cada uno de estos contextos representa el conocimiento de cada usuario dentro del sistema. Es posible establecer una serie de relaciones entre los usuarios, o más bien, entre los contextos de cada uno de ellos. Las más importantes, que son las que utilizaremos en este trabajo, son: Lenguaje común y Conjunto de Objetos comunes.

■ Lenguaje común: Dados dos usuarios y los contextos formales de cada uno de ellos, el lenguaje común es el conjunto de atributos que ambos usuarios tienen. Sean M_1 y M_2 los conjuntos de atributos de cada usuario respectivamente; el lenguaje común \mathcal{L} es precisamente la intesección de estos dos conjuntos:

$$\mathcal{L} = M_1 \cap M_2$$

■ Conjunto de Objetos comunes: Dados dos usuarios y los contextos formales de cada uno de ellos, el conjunto de Objetos comunes es el conjunto de objetos que ambos usuarios tienen. Sean G_1 y G_2 los conjuntos de objetos de cada usuario respectivamente; el conjunto de objetos comunes \mathcal{O} es precisamente la intersección de estos dos conjuntos:

$$\mathcal{O} = G_1 \cap G_2$$

Además de las dos relaciones anteriores, se pueden establecer otro tipo de relaciones entre usuarios. Se pueden establecer restricciones en las
relaciones anteriores para crear nuevas relaciones. Por ejemplo, se puede
establecer un lenguaje común reducido entre dos usuarios escogiendo sólo
aquellos atributos que, además de pertenecer al lenguaje común, son parte
de alguna etiquetación en el mismo enlace para ambos usuarios. Es obvio
que este nuevo lenguaje común reducido es bastante más pequeño que

31

aquel en el que no se impone ninguna restricción. Igualmente, se puede crear un conjunto de objetos comunes reducido escogiendo únicamente los enlaces que, además de pertenecer al conjunto inicial de objetos comunes, están etiquetados con una misma etiqueta por ambos usuarios. Igualmente, este último conjunto es mucho más pequeño que el inicial.

Por último, hay que indicar que se pueden establecer tantas relaciones como criterios se quieran inventar. Sin embargo, en este trabajo sólo hemos utilizado las dos anteriores: el Lenguaje común servirá para crear contextos reducidos en los que sólo existan objetos etiquetados con etiquetas pertenecientes a este lenguaje, previo a la conciliación del conocimiento de ambos usuarios; el conjunto de objetos comunes servirá para decidir que pares de usuarios deben realizar conciliación y cuáles no, esto se decidirá en función del tamaño de este conjunto. En este trabajo, únicamente se producirá conciliación entre dos usuarios cuando compartan 3 objetos o más; es decir, cuando esté conjunto de objetos comunes esté formado por 3 o más objetos.

Bibliografía

- [1] Alonso-Jiménez, J.A., Aranda-Corral, G.A., Borrego-Díaz, J., Fernández-Lebrón, M.M., Hidalgo-Doblado, M.J. (2008). Extending Attribute Exploration by Means of Boolean Derivatives. In: Proc. 6th Int. Conf. on Concept Lattices and Their Applications. CEUR Workshops Proc., p. 433.
- [2] Aranda-Corral, G.A., Borrego-Díaz, J. (2010). Reconciling Knowledge in Social Tagging Web Services. Lecture Notes in Computer Science, 6077(1), 101-121.
- [3] Aranda-Corral, G.A., Borrego-Díaz, J., Gómez-Marín, F. (2009). Toward Semantic Mobile Web 2.0 through Multiagent Systems. In: Håkansson, A., Nguyen, N.T., Hartung, R.L., Howlett, R.J., Jain, L.C. (eds.) KES-AMSTA 2009. LNCS, vol. 5559, pp. 400–409. Springer, Heidelberg.
- [4] Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, Vol. 2008, No. 10.
- [5] Brooks, C.H., Montanez, N. (2006). Improved annotation of the blogosphere via autotagging and hierarchical clustering. In WWW 06. Proceedings of the 15th international conference on World Wide Web, (pp.625-632.). New York: ACM Press.
- [6] Ganter, B., Wille, R. (1999). Formal Concept Analysis Mathematical Foundations. Springer, Heidelberg.
- [7] Golder, S., Huberman, B.A. (2006). The structure of collaborative tagging systems. Journal of Information Science 32(2), 98-208.
- [8] Gruber, T. (2007). Ontology of Folksonomy: A Mash-up of Apples and Oranges. Int'l. Journal on Semantic Web & Information Systems 3(2).

34 Bibliografía

[9] Halpin, H., Valentin R., and Hana S. (2006). The dynamics and semantics of collaborative tagging. Proceedings of the 1st Semantic Authoring and Annotation Workshop (SAAW06).

- [10] Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., Stumme, G. (2008). Discovering shared conceptualizations in folksonomies. Journal of Web Semantics 6(1), 38–53.
- [11] Kim, H.-L., Scerri, S., Brslin, J., Decker, S., Kim, H.-G. (2008). The state of the art in tag ontologies: A semantic model for tagging and folk-sonomies. In: International Conference on Dublin Core and Metadata Applications, Berlin, Germany.
- [12] Knerr, T. (2006). Tagging ontology-towards a common ontology for folksonomies. http://tagont.googlecode.com/files/TagOntPaper.pdf.
- [13] Mathes, A. (2004). Folksonomies Cooperative Classification and Communication Through Shared Metadata. http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html
- [14] Mika, P. (2005). Ontologies Are Us: A unified model of social networks and semantics. Proceedings of the 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, (pp. 522-536). Berlin, Heidelberg: Springer.
- [15] Rowley, J. (1995). Organizing Knowledge. 2nd Ed. Brookfield, VT: Gower.
- [16] Shirky, C. (2005). Ontology is Overrated: Categories, Links and Tags. http://www.shirky.com/writtings/ontology_overrated.html
- [17] Smith, G. (2007). Tagging: People-Powered Medatada for the Social Web. First. New Riders Publishing, Indianapolis.
- [18] Tanaka, J.W., Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? Cognitive Psychology 23(3), 457–482.
- [19] Van Damme, C., Hepp, M., Siorpaes, K. (2007). FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies. In: ESWC 2007 workshop Bridging the Gap between Semantic Web and Web 2.0, May 2007, pp. 57–70.

Bibliografía 35

[20] Weick, K.E., Sutcliffe, K.M., Obstfeld, D. (2005). Organizing and the Process of Sensemaking. Organization Science 16(4), 409–421.

- [21] Yeung, C.M.A., Gibbins, N., Shadbolt, N. (2009). Contextualising Tags in Collaborative Tagging Systems. In: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia.
- [22] ConExp: http://conexp.sourceforge.net/
- [23] Delicious: http://www.delicious.com/
- [24] Notación DOT: http://www.graphviz.org/Documentation.php
- [25] FIPA (Foundation for Intelligent Physical Agents): http://www.fipa.org/
- [26] Gephi: http://gephi.org/
- [27] JADE (Java Agent Development Framework): http://jade.tilab.com/