

APLICACIÓN DE LAS ONTOLOGÍAS PARA LA REPRESENTACIÓN DEL CONOCIMIENTO

Eva GARCÍA HERNÁNDEZ, evagahe@yahoo.es

Universidad Carlos III de Madrid (Getafe, España)

Resumen

La perspectiva de la construcción de ontologías, en un entorno tecnológico, permite abordar la problemática de la indización, recuperación y divulgación de la información depositada en internet, facilitando una gestión rápida, eficaz, pertinente y permanentemente actualizada de las necesidades a cubrir. La aplicación de ontologías terminológicas unifica la terminología de cada concepto y las relaciones entre ellos; y las ontologías de información unifican las estructuras de almacenamiento, de forma que pueden ser reutilizadas por varias aplicaciones informáticas con la misma fuente de información.

El conocimiento documental tradicionalmente ha sido organizado a través de la clasificación de Dewey o la clasificación decimal universal y los encabezamientos de materia. Actualmente, más orientado a la recuperación, se utilizan descriptores de un lenguaje controlado con el soporte de un tesoro. Dichos modos de almacenamiento y recuperación de información han jugado un papel fundamental en el uso generalizado y estandarizado de sistemas documentales, permitiendo a los documentalistas tener unos modelos comunes y aceptados en la descripción y almacenamiento de información. Los conceptos se describen explícitamente para entender su significado, mediante acuerdos ontológicos. De este modo, un usuario que desee reutilizar una ontología desarrollada por otros puede conseguir la información de todos los conceptos que soporta, su taxonomía y los axiomas.

Palabras clave

ontologías, herramientas para la recuperación de la información, metadatos, indización, web semántica

1 INTRODUCCIÓN

En los últimos años, se ha comenzado a experimentar el uso de técnicas de representación del conocimiento para extender la web actual con metadatos, en la que los ordenadores no sólo serán capaces de presentar toda la información contenida en ella, sino que además podrán entenderla y gestionarla de forma «inteligente». Esta web mejorada se conoce actualmente como web semántica.

Así, la web se convertirá en un espacio autonavegable y autocomprensible, siendo posible la relación entre los términos independientes de

una misma búsqueda, ofreciendo como resultado sólo aquellas páginas web que realmente se adaptan a los requisitos especificados. De este modo, la información se convierte actualmente en conocimiento codificando los datos con metadatos mediante las ontologías. Cada una de estas ontologías se crea atendiendo a los principios de reutilización y diseño de forma consensuada.

Los datos por si solos, sin tener una semántica asociada, no son de utilidad por resultar ambiguos. Por lo tanto, hay que introducir esa semántica añadiendo datos que describen a su vez los propios datos o metadatos. Sin embargo, para dar una descripción formal de contenidos, hace falta una herramienta más potente que los metadatos: las ontologías. Mientras los metadatos estructuran contenidos, las ontologías estructuran la semántica de un recurso.

La utilización de los metadatos es necesaria para poder definir en las ontologías qué partes pueden ser compartidas o reutilizadas. Sin embargo, las ontologías van más allá, pues son capaces de almacenar otro tipo de conocimiento, la semántica, a través de los conceptos y las relaciones entre ellos, permitiendo realizar una representación declarativa del conocimiento de un dominio, que puede ser comunicado entre personas y máquinas.

La información que aparece en internet va a poder interpretarse por los ordenadores, sin necesidad de intervención humana, al unificar los contenidos semánticos por medio de ontologías que formalicen el conocimiento de forma consensuada y reutilizable, para poder compartirlo. Es necesario un lenguaje común con suficiente capacidad expresiva y de razonamiento para representar la semántica de las ontologías.

Recientemente, con el desarrollo masivo de los recursos en las bases de conocimientos de la web, se habla de ontologías o taxonomías virtuales como modelo conceptual de organización y representación del conocimiento, capaces de servir de clasificaciones virtuales, de representar conocimientos y de razonar con ellos mediante reglas de inferencia.

La perspectiva de la construcción de ontologías del conocimiento, en un entorno tecnológico resuelve la problemática de la indización, recuperación y divulgación de la información depositada en internet, además, las tecnologías de la información pueden tener un importante impacto como factor de dispersión, e influir en su desarrollo.

El uso de ontologías para la recuperación de la información tiene ventajas sobre los métodos de acceso simples basados en palabras clave, al suministrar un vocabulario compartido común para expresar información sobre el contenido de los documentos. Además incluyen axiomas para especificar relaciones entre conceptos. Podemos utilizarlas para formular consultas complejas y recuperar exactamente la información que precisamos. Con los axiomas podemos derivar información intensional.

Los actuales buscadores, basados en indización humana en el caso de los índices de búsqueda, e indización automática en el caso de los motores de búsqueda, no ofrecen tasas de exhaustividad y precisión realmente aceptables en sus resultados. Este problema puede ser consecuencia de la falta de significado o semántica que para los ordenadores tienen los

documentos web, formateados mediante lenguaje HTML, que sólo expresa la forma de presentación de los contenidos.

El proyecto denominado web semántica se presenta como un área pujante en la confluencia de la Inteligencia Artificial y las tecnologías web, proponiendo nuevas técnicas y paradigmas para la representación del conocimiento que faciliten la localización e integración de recursos a través de la WWW (BERNERS-LEE 2001). La web semántica rescata la noción de ontología del campo de la Inteligencia Artificial como vehículo para cumplir este objetivo (GRUBER 1993).

Una ontología es una taxonomía de conceptos con atributos y relaciones que proporciona un vocabulario consensuado para definir redes semánticas de unidades de información interrelacionadas. Concretamente, estará formada por una taxonomía relacional de conceptos y por un conjunto de axiomas o reglas de inferencia mediante los cuales se podrá inferir nuevo conocimiento. En los últimos años se han desarrollado diversos lenguajes y estándares para su definición.

La aplicación de ontologías terminológicas unifica la terminología de cada concepto y las relaciones entre ellos; y las ontologías de información unifican las estructuras de almacenamiento de forma que pueden ser reutilizadas por varias aplicaciones informáticas que utilizan la misma fuente de información.

El conocimiento documental tradicionalmente ha sido organizado a través de la clasificación de Dewey o la clasificación decimal universal y los encabezamientos de materia. Actualmente, más orientado a la recuperación, se utilizan descriptores de un lenguaje controlado con el soporte de un tesoro. Estos modos de almacenamiento y recuperación de información han jugado un papel fundamental en el uso generalizado y estandarizado de sistemas documentales, permitiendo a los documentalistas tener unos modelos comunes y aceptados en la descripción y almacenamiento de información.

Las ontologías se sitúan dentro de los lenguajes controlados. Se deduce cierta analogía con otras estructuras conceptuales de organización de la información como tesauros o clasificaciones conceptuales, que igualmente establecen asociaciones entre sus conceptos. Pero existen diferencias fundamentales en cuanto a la representación del conocimiento, debido a que las ontologías permiten representar axiomas (conocimientos ciertos e inmutables) y razonar con ellos mediante las reglas de inferencia definidas en los agentes que filtran la información de las redes distribuidas de Internet.

Una ontología para una base de conocimientos debe abarcar todo tipo de documentos, las descripciones conceptuales, sus relaciones y los diferentes problemas científicos. Además debe incluir índices, descripciones bibliográficas, tesauros, códigos clasificatorios, formalizaciones de validez e información terminológica. Se la emplea para unificar vocabularios y definir estructuras comunes entre diferentes aplicaciones federadas que tengan como objetivo la representación del conocimiento. Sus relaciones son tanto estructuras lógicas como sociales, cuya estabilidad depende de la madurez del conocimiento abarcado.

La construcción de ontologías lleva implícito que cada término y cada relación entre términos se define formalmente. Los conceptos se describen explícitamente para entender su significado, mediante acuerdos ontológicos. Con ello un usuario que desee reutilizar una ontología desarrollada por otros, puede conseguir la información de todos los conceptos que soporta, su taxonomía y los axiomas. A diferencia de los tesauros, las ontologías poseen una mayor variedad de relaciones entre conceptos. Estas relaciones se crean a propósito del modelo conceptual existente en el dominio que se está formalizando. El objetivo último del desarrollo de ontologías es el de mejorar la representación de la información y los sistemas de recuperación de información. Se puede asegurar que la web proporciona un perfecto lugar de testado para aplicaciones conducidas mediante ontologías.

Este texto intenta reflejar el estado y evolución, presente y futuro, de esta línea de investigación -ontologías para la representación del conocimiento- bajo el marco organizacional del conocimiento, ofreciendo con ello una visión general sobre el tema.

2 MARCO DE IMPLANTACIÓN

La construcción de ontologías se ajusta perfectamente a las tareas tradicionales de la Biblioteconomía y la Documentación, debido a que las ontologías almacenan el conocimiento en un formato que puede ser utilizado por sistemas automáticos para realizar deducciones, y además, su propósito de reutilización exige que sus conceptos estén descritos explícitamente, de modo que también sean asimilables y comprensibles por el entendimiento humano.

Un análisis de un ejemplo que puede ser extrapolado al caso de reutilización de un sistema documental, es la iniciativa del KA (*Knowledge Acquisition*), que modeliza a la Comunidad de Adquisición de Conocimientos en una ontología llamada *Ontology*, accesible a toda la comunidad a través del *Ontology Server*.

En el contexto de la reutilización de la información, una ontología (GRUBER 1993a, 1993b) es una descripción de los conceptos (asimilable a la especificación formal de un programa informático) y de sus asociaciones, existentes en un agente (o *softbot*) o federación de agentes de software de la web. Lo que significa que una ontología proporciona una estructura y unos contenidos que son independientes del fin y del dominio de la aplicación en la que se reutilizarán sus definiciones.

Se estudia refundir las categorías de varios sistemas ontológicos de manera que se integren y se comuniquen entre sí. Para la representación de su conocimiento se perfila un consenso a favor de un modelo trilateral, con un esquema ontológico o estructura semántica central, un esquema de representación y un esquema de aplicación.

2.1 Metodologías

Desarrollar una ontología requiere definir las clases que forman un dominio, organizar las clases en una jerarquía taxonómica, definir las propiedades de cada clase e indicar las restricciones de sus valores y asignar valores a las propiedades para crear instancias. Actualmente, no existe un acuerdo entre la comunidad investigadora que permita la formulación de una metodología estándar para la construcción de ontologías, pero para poder desarrollar una metodología viable es necesario el conocimiento de las mismas.

Adaptando el estándar establecido por la IEEE 1074-1995 (Institute of Electrical and Electronics Engineers)¹ para el desarrollo de un proceso de construcción de software, el desarrollo de las ontologías se clasifica en tres categorías. Actividades ligadas al manejo del proyecto (planificación, seguimiento de la planificación y asegurar la calidad), actividades orientadas al desarrollo de la ontología (especificación, conceptualización, formalización, implementación y mantenimiento), y actividades integrales (adquisición de conocimiento, integración con otras ontologías, evaluación y documentación).

Estas metodologías proporcionan unas directrices que indican cómo realizar las actividades en el proceso de desarrollo, qué técnicas son las más apropiadas en cada actividad y qué produce cada una de ellas. Existen metodologías tales como IDEF5, la metodología de Unschold y King, la metodología de Gruninger y Fox, *Methontology, Chemicals* (GÓMEZ-PÉREZ 1999), *Environmental pollutants ontologies* (ROJAS 1998), *The Reference-Ontology* y la versión estructurada de la ontología (KA).

3 APLICACIONES DE LAS ONTOLOGÍAS

Las ontologías proporcionan una comprensión compartida y consensuada del conocimiento de un dominio, que puede ser comunicada entre personas y sistemas heterogéneos. Para poder reutilizar conocimientos de otros sistemas es necesario conocer y estar conforme con la terminología y su significado. Por ello, se llaman acuerdos ontológicos a los acuerdos terminológicos necesarios para reutilizar conocimiento. Se trata de convertir la información en conocimiento.

El amplio desarrollo de la investigación en el campo de las ontologías, permite un aumento notable de los usos de éstas, manifestándose los más frecuentes en el campo de la denominada web semántica, la ingeniería del conocimiento o los sistemas de información.

1. (IEEE1074, 1995) ANSI/IEEE Std.1074-1995. IEEE Standard for Developing Software Life Cycle Processes. The Institute of Electrical and Electronics Engineers, Inc. 6, 1995.

3.1 Aplicación de las ontologías en la Web Semántica

- *Indización de documentos*: La indexación de un sitio web con apoyo de una ontología terminológica comienza con la extracción de los términos más relevantes de cada página, y después de asociar a estos términos conceptos candidatos, se evalúa la capacidad de representación de la página de cada uno de estos conceptos, que determina su nivel de representatividad, y finalmente se construye el índice. De esta manera, las consultas se procesan a un nivel conceptual, lo que reportará un mayor grado de acierto.
- *Agrupamiento*: Las técnicas de *clustering* permiten el crecimiento de un sistema mediante la adición de procesadores o CPU (*central processing units*) a la unidad primitiva. Las ontologías aportan las herramientas para que los distintos equipos puedan entenderse entre sí y funcionar como si fuera uno sólo.
- *Servicios web*: Las ontologías representarán los datos en la red de tal forma que puedan ser utilizados y comprendidos por las máquinas sin necesidad de la intervención humana web semántica.
- *Comercio electrónico*: Hay ontologías orientadas a aplicaciones que facilitan el comercio electrónico, destacando el sistema MKBEEM (*Multilingual Knowledge Based European Electronic Marketplace*), que es un portal multilingüe que combina procesamiento basado en ontologías y procesamiento de lenguaje humano.

3.2 Aplicación de ontologías en ingeniería del conocimiento

- *Ingeniería del conocimiento*: Por un lado en el modelado conceptual se crea un glosario de la terminología del dominio de la aplicación (los conceptos), las relaciones entre dichos términos y las restricciones de uso. Este modelo conceptual explícito es la ontología. Por otro, la construcción de la base de conocimiento usa la ontología definida en la etapa anterior como un conjunto de esquemas o contenedores de conocimiento.
Las ontologías, junto con los métodos de solución de problemas (PSM), prometen la posibilidad de reutilizar componentes en ingeniería de conocimiento. Mientras las ontologías definen el conocimiento declarativo del dominio a un nivel genérico, los PSM especifican conocimiento de razonamiento sobre el dominio.
- *Procesamiento del lenguaje natural*: Una ontología puede mantener la definición de elementos gramaticales del lenguaje y sus relaciones, permitiendo, por ejemplo, el análisis sintáctico de un texto.

- *Distribución del conocimiento desde el punto de vista de la reutilización:* Uno de los fines principales de la utilización de una ontología en un sistema es poder reutilizar conocimiento para sistemas futuros. Se podrán integrar ontologías para la constitución de una nueva, más grande, que mejore la conceptualización que aportaban todas ellas por separado. Sistemas distintos podrán entender la información almacenada en una ontología sobre un dominio y trabajar con dicha información aunque no haya sido generada por y para ellos.
- *Distribución de conocimiento como un camino para resolver la integración de sistemas basados en conocimiento:* Integración inteligente de información. Como se ha señalado en puntos anteriores, la distribución de un conocimiento de forma estandarizada se traducirá en importantes mejoras en el desarrollo de agentes inteligentes que tendrán a su disposición un mayor número de bases de conocimiento disponibles.
- *Implementación de agentes software inteligentes:* Una posible meta de esta tecnología es poder disponer de un agente software para cada dominio o cada tarea que tenga que realizar un humano facilitándole la obtención de resultados.
- *Creación y construcción de cursos de manera colaborada destinado al desarrollo y la integración de sistemas de aprendizaje*

3.3 Aplicación de ontologías en sistemas de información

- *Interoperatividad entre sistemas heterogéneos:* Las ontologías se presentan como una solución para lograr una integración inteligente. Con una ontología terminológica, se pueden organizar los términos que son usados en interacciones entre sistemas heterogéneos, de manera que reconozca cuándo una aplicación está usando un término que es más general o más específico que otro que está en uso por otra aplicación.
- *Sistemas de información cooperativa:* El objetivo es que múltiples sistemas de información sean capaces de trabajar de forma cooperativa combinando sus datos y funcionalidades, con ayuda de las ontologías.
- *Medio de distribución de conocimiento dentro de aplicaciones de software y entre aplicaciones de software:* Mediante la comunicación entre aplicaciones sin intervención humana, utilizando estándares y protocolos para el entendimiento recíproco.
- *Recuperación de información, enfocado a mejorar la formulación de consultas:* Si se añade semántica a las consultas y no sólo se efectúan por palabras clave se proporciona una calidad superior en los resultados de una búsqueda. Las consultas serán tratadas desde un punto de vista

conceptual. De este modo, se reducirá el ruido y el silencio en los resultados de una búsqueda, lo que permitirá que no se omitan aquellos resultados, que aún siendo conceptualmente sinónimos al de la consulta, no se encuentran por ser distintos terminológicamente.

- *Normalización de sistemas documentales*: La documentación generada por los nuevos sistemas contará con características como la identificación del contenido documental (información) mediante el uso de códigos de identificación y descriptores, el etiquetado de la documentación legible en las fuentes y en los instrumentos de representación (catálogos, listados, bases de datos, etc.), el establecimiento de una estructura base que permita la relación de los términos empleados en la identificación documental (tesauros) y la creación de instrumentos auxiliares para la recuperación de la información como índices, tablas, etc.

4 LENGUAJES DE ESPECIFICACIÓN DE ONTOLOGÍAS

Las ontologías requieren de un lenguaje lógico formal para ser expresadas. En Inteligencia Artificial se han desarrollado algunos lenguajes para este fin, basados en la lógica de predicados de primer orden como *Cycl*, basados en macros, como *Ortolingua* y *Frame Logia* o en estrategias de razonamiento, como *Classic* y *Loom*.

En la elección de un lenguaje para la definición de una ontología debe considerarse que éste debe poseer una sintaxis bien definida para poder «leer» con facilidad la ontología definida, debe tener una semántica bien definida para comprender perfectamente el funcionamiento de la ontología y la suficiente expresividad para poder capturar varias ontologías, además de ser fácilmente mapeable desde/hacia otros lenguajes ontológicos y ser eficiente a la hora de realizar un razonamiento.

Por otra parte destacan dos lenguajes intermedios, KIF (*Knowledge Interchange Format*) y PIF (*Process Interchange Format*), que usan aplicaciones heterogéneas e independientes cuando necesitan intercambiar conocimiento y se han desarrollado con el propósito de resolver el problema de la heterogeneidad en los lenguajes de representación del conocimiento.

Entre las herramientas dedicadas a la implementación de ontologías destacan las de anotación, que permiten estructurar la información publicada mediante su clasificación sobre la base de conceptos semánticos, y añadir información adicional, estructurada o no, a los contenidos actuales de la web y a los de nueva creación. Podemos considerar como herramientas de anotación los editores de ontologías, los servidores de anotaciones, las herramientas de anotación externa y las herramientas del propietario.

Mientras que los editores de ontologías facilitan la definición, el desarrollo distribuido y la unión de ontologías, los servidores de anotaciones mantienen los repositorios de conocimiento generado a través de las anotaciones.

Las herramientas de anotación externa permiten asociar metainformación a páginas web que ya están presentes en Internet, haciendo posible la anotación de las páginas visitadas por los usuarios. Esa metainformación, que también puede estar basada en una ontología, se almacena en un repositorio externo, que suelen ser bases de datos RDF y se implementan en la actualidad sobre sistemas gestores de bases de datos relacionales (SGBDR). La metainformación relacionada con la página web queda almacenada en un repositorio común al que podrán acceder todos los usuarios de la comunidad.

Las herramientas del propietario, partiendo de una ontología definida a priori, permiten incluir la metainformación en la misma página web y facilitan la estructuración de la información publicada en la web, ya sea en el ámbito de intranet, extranet o internet, haciendo posible la inclusión de metainformación en los documentos publicados. La información suele incluirse en alguno de los nuevos lenguajes de marcado promovidos por el W3C. Para la definición de ontologías se usa DAML (*DARPA Agent Mark-Up Language*), aunque actualmente este Consorcio está trabajando en la definición de un lenguaje estándar para definir ontologías, *Ontology Working Group* (OWL).

Entre las herramientas que podemos usar para mantener y construir ontologías sobresalen *Ontology Server*, *Ontolingua*, *WebOnto*, *ProtégéWin*, *OntoSaurus*, *ODE*, *KADS22*, *Apollo*, *LinkFactory*, *OILED*, *OntoEdit*, *OpenKnoME*, *Protégé-2000* y *SymOntoX*.

Para definir ontologías de manera estándar en diversos entornos se utilizan lenguajes de marcado como XML (*eXtensible Markup Language*) (BRAY 2000), XML Schema, RDF (*Resource Description Framework*) (LASILLA 1999), RDF Schema, DAML+OIL (*DARPA Agent Markup Language + Ontology Inference Layer*) (CONNOLY 2001), y más recientemente OWL (*Web Ontology Language*) (DEAN 2002), respaldados por el consorcio W3C. También se pueden confeccionar modelos de conocimiento del tipo entidad-relación o modelos de datos orientados a objetos, aunque ambas opciones presentan limitaciones en cuanto a la interpretación y representación. Se podría afirmar que es posible expresar una ontología en XML Schema, pero su definición debería ser realizada previamente en un lenguaje ontológico.

En la actualidad existe una serie de sistemas que utilizan ontologías para su funcionamiento. El sistema MESIA se presenta como un Modelo Computacional para Extracción Selectiva de Información de textos cortos, cuyo objetivo es la recuperación de documentos de la web de la Comunidad Autónoma de Madrid utilizando recursos lingüísticos para mejorar los resultados que ofrecen los buscadores tradicionales, y transforma una consulta en lenguaje natural en una consulta formal.

En el sistema MKBEEM (*Multilingual Knowledge Based European e-Market*), se utilizan las ontologías para la clasificación e indexación de catálogos, para facilitar los diálogos multilingües hombre-máquina entre el usuario y el agente software, para el filtrado de las consultas del usuario y para inferir información relevante que satisfaga las peticiones del usuario.

PEGASUS (*Presentation modeling Environment for Generis Adaptive hypermedia Support Systems*) es un sistema genérico de presentación para sistemas hipermedia de enseñanza adaptativa independiente de la representación del conocimiento del dominio y del mantenimiento del estado de la aplicación. La generalidad se consigue proporcionando un marco de aplicación para la definición de las ontologías que mejor se ajustan a un dominio o un autor concreto.

SCASEM (Sistema de catalogación semántica) es un proyecto del Grupo Planeta para adaptar al nuevo mercado la gestión de sus fondos, que constituyen el Banco de Contenidos Planeta (BCP), junto con la Base de Datos de Conocimiento (BDCon) y un conjunto de herramientas lingüísticas. Esta base de conocimiento contiene una ontología multidimensional que constituye la herramienta central del sistema de clasificación del BCP.

5 CONCLUSIONES

En la actualidad se aprecia una progresiva sustitución de los modelos convencionales del proceso de recuperación de la información por modelos cognitivos que aplican el modelo humano de memoria y aprendizaje a los agentes inteligentes de recuperación de la información, proponiendo las ontologías para estructurar las bases del conocimiento.

Los servicios de ontologías ofrecerán un marco contextualmente rico y moderno para elaborar modelos, prestar servicios y gestionar la terminología en cualquier materia. Cuando se integren con instrumentos de búsqueda basados en Internet, facilitarán mucho la recuperación de recursos, no sólo proporcionando acceso a los documentos específicos, sino también ofreciendo sugerencias relativas a otros recursos conexos potencialmente pertinentes para el tema de interés.

La iniciativa de la web semántica, en su estado actual, sólo ofrece descripciones para los recursos, pero nada sobre cómo presentar esos recursos a los usuarios de una forma conceptualmente clara. Por ello se avanza en la definición de una web conceptual que no sólo proporcione información semántica para la máquina, sino también información conceptual para el usuario. En este sentido no se puede olvidar que el objetivo fundamental de cualquier servicio debe ser la satisfacción de los usuarios.

Existen otras líneas de investigación en el campo de la Inteligencia Artificial, en la que los investigadores de esta innovadora ciencia han encontrado en las ontologías el modelo del conocimiento ideal para describir formalmente los recursos web y su vocabulario y, por tanto, hacer explícito el significado subyacente de los términos incluidos en las páginas web. La semántica ontológica (NIREMBURG 2001) es una teoría que estudia el significado del lenguaje humano o natural, así como una aproximación al procesamiento del lenguaje natural (PLN) que utiliza un modelo abstracto del mundo -la ontología- como recurso central para extraer y representar el significado de textos en lenguaje natural, al razonar con el conocimiento que se deriva a partir de estos textos. Asimismo, la

ontología es también el eje central a la hora de generar textos en lenguaje natural basados en las representaciones de su significado.

Con la semántica ontológica como punto de partida para la anotación de recursos web con información ontológica, se pretende conseguir el acceso inteligente a dichos recursos, facilitar la búsqueda y navegación en la Web y explorar nuevos enfoques de inferencia a partir de estos recursos. Se han desarrollado muchos sistemas y proyectos en este sentido: la iniciativa (KA) (BENJAMINS 1999); *PlanetOnto* (MOTTA 1999); *SHOE* (LUKE 2000) y el proyecto *Semantic Community Web Portals* (STAAB 2000). También han visto la luz hasta la fecha varias herramientas de anotación semántica: *COHSE*, *MnM* (VARGAS-VERA 2001), *OntoMat-Annotizer*, *SHOE Knowledge Annotator* y *AeroDAML*.

Las ontologías son por definición consensuadas y, por tanto, están más cerca de convertirse en estándar que muchos otros modelos de conocimiento.

Muchos de los grupos de investigación que desarrollan ontologías se caracterizan por tener un enfoque interdisciplinar que combina la Informática, la Lingüística y, a veces, la Filosofía. En este sentido se trabaja en el desarrollo de anotaciones lingüísticas en el nivel semántico que, según los expertos, son más ambiciosas y potencialmente más amplias que las basadas exclusivamente en ontologías y, por ello, más complejas y con un coste mayor. Parece una mejora prometedora establecer un enlace entre la anotación semántica y la del discurso, siguiendo el enfoque de la Teoría retórica de la estructura (MANN 1988), que ya se ha aplicado en la generación de textos.

El significado no está sólo en la semántica. La información contenida en una página web se presenta principalmente en sintagmas nominales o expresiones. Todos los niveles lingüísticos interactúan estrechamente para determinar el significado de una oración, una declaración o una expresión completa. Aunque los componentes básicos de una expresión sean los significados de las palabras, su significado estará caracterizado no sólo por esos significados, sino también por el orden de combinación que presenten. Dado que los modos de combinación están muy determinados por la estructura sintáctica de la lengua, necesitaremos reflejar el significado que aporta cada regla sintáctica subyacente en la expresión que se está analizando, es decir, la operación semántica que combina los significados. Por tanto, para determinar el significado de una expresión se necesita realizar su análisis sintáctico.

De ahí que autores como Dik (DIK 1989), Aguado y Pareja-Lora (AGUADO 2000) y Vargas-Vera (VARGAS-VERA 2001) consideren de gran importancia la interrelación entre sintaxis y semántica. Por otro lado, se observa la necesidad de una mayor integración entre la semántica oracional y las teorías de la estructura textual o discursiva, incluyendo aspectos como la composición y el escenario del texto o diálogo, o la intencionalidad de los hablantes. Así, para ayudar a determinar el significado de una expresión en un texto, conviene llevar a cabo un análisis pragmático, explícito o implícito. Por tanto, podemos llegar a la conclusión de que sería muy útil para la comunidad de la web semántica tener algún

modelo de anotación que permita no sólo anotar y hacer explícito el nivel semántico, sino también los otros niveles, con el fin de facilitar el tratamiento automático de las páginas web.

Existe una propuesta hecha por el World Wide Web Consortium (W3C) *SKOS-Core*,² cuyo objetivo fundamental es proporcionar un modelo para la migración de sistemas de organización de conocimiento al entorno de la web semántica. Además sirve para construir esquemas de conceptos simples para su utilización en la web. *SKOS-Core* está pensado como un complemento a OWL, ya que proporciona un marco básico para la construcción de esquemas de conceptos pero sin la definición semántica tan estricta que exige la utilización de OWL. Supone en cierta medida una simplificación mayor de la que encontramos ya de por sí en OWL-Lite, lo cual permite acceder a un mayor número de personas a este tipo de tecnologías para la representación del conocimiento.

Finalmente, cabe resaltar que en la actualidad se aprecia un gran movimiento en los grupos de trabajo de World Wide Web Consortium (W3C) y de Semantic Web, especialmente a nivel académico y de estandarización. Hay una considerable proliferación de proyectos para la definición de editores de ontologías, y de motores de inferencia. Hay incluso navegadores «inteligentes» de pago como Amblit.³

BIBLIOGRAFÍA CITADA

- (AGUADO 2000) AGUADO, G.; PAREJA-LORA, A. «A competition model for the generation of complementation patterns in machine translation». *International Journal of Translation*. Vol. 12, n. 1-2 (Jan-Dec 2000).
- (BENJAMINS 1999) BENJAMINS, V.R.; FENSEL, D.; DECKER, S.; GÓMEZ-PÉREZ, A. «Building Ontologies for the Internet: a Mid Term Report». *IJHCS: International Journal of Human Computer Studies*. n. 51 (1999), p. 687-712.
- (BERNERS-LEE 2001) BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. «The Semantic Web». *Scientific American*. Vol. 284, n. 5. (May, 2001).
- (BRAY 2000) BRAY, T.; PAOLI, J.; SPERBERG-MCQUEEN, C. M.; MALER, E. «Extensible Markup Language (XML) 1.0» (Second Edition) [recurso electrónico]. *W3C Recommendation 6 October 2000*. <<http://www.w3.org/TR/REC-xml>> [Consulta: 10 ene. 2005].
- (CONNOLLY 2001) CONNOLLY, D.; VAN HARMELEN, F.; HORROCKS, I.; MCGUINNESS, D. L.; PATEL-SCHNEIDER, P. F.; STEIN, L.A. «DAML+OIL Reference Description» [recurso electrónico]. *W3C Note*. (18 December 2001). <<http://www.w3.org/TR/daml+oil-reference>> [Consulta: 10 ene. 2005].
- (DEAN 2002) DEAN, M.; CONNOLLY, D.; VAN HARMELEN, F.; HENDLER, J.; HORROCKS, I.; MCGUINNESS, D. L.; PATEL-SCHNEIDER, P. F.; STEIN, L. A. «OWL Web Ontology Language 1.0 Reference». *W3C Working Draft 29 July 2002* [recurso electrónico]. <<http://www.w3.org/TR/owl-ref>> [Consulta: 10 ene. 2005].
- (DIK 1989) DIK, S.C. *The Theory of Functional Grammar. Part I: The Structure of the Clause*. Dordrecht: Foris Publications (Functional Grammar Series, 9).

2. <<http://www.w3.org/2001/sw/Europe/events/200406-esp/trabajo-final-extratesauros/node6.html>>

3. <<http://www.amblit.com/>>

- (GÓMEZ-PÉREZ 1999) GÓMEZ-PÉREZ, A.; BENJAMINS, R. «Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods». En: *Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods* (KRR5). (Morgankaufmann, 1999).
- (GRUBER 1993a) GRUBER, T. R. «A Translation Approach to Portable Ontologies». *Knowledge Acquisition*. Vol. 5 (1993), n. 2, p. 199-220.
- (GRUBER 1993b) GRUBER, T. R. «Toward Principles for the Design of Ontologies Used for Knowledge Sharing». En: GUARINO, N.; Poli, R. (eds.). *Formal Ontology in Conceptual Analysis and Knowledge Representation* (Boston: Kluwer Academic Press, 1993).
- (LASILLA 1999) LASSILA, O.; SWICK, R. «Resource Description Framework (RDF): Model and Syntax Specification» [recurso electrónico]. *W3C Recommendation 22 February 1999*. <<http://www.w3.org/TR/REC-rdf-syntax>> [Consulta: 10 ene. 2005].
- (LUKE, 2000) LUKE, S.; HEFLIN, J. «SHOE 1.01: Proposed Specification» [recurso electrónico]. En: *SHOE Project* (February 2003). <<http://www.cs.umd.edu/projects/plus/SHOE/spec1.01.htm>> [Consulta: 10 ene. 2005].
- (MANN 1988) MANN, W.; THOMSON, S. «Rhetorical Structure Theory: Toward a functional theory of text organization». *Text*. Vol.18 (1988), n. 3, p. 243-281.
- (MOTTA 1999) MOTTA, E.; BUCKINGHAM SHUM, S.; DOMINGUE, J. «Case Studies in Ontology-Driven Document Enrichment». En: *Proceedings of the 12th Banff Knowledge Acquisition Workshop* (Banff, Canada: the Workshop, 1999).
- (NIREMBURG 2001) NIRENBURG, S.; RASKIN, V. «Ontological Semantics» [recurso electrónico]. <<http://crl.nmsu.edu/Staff/pages/Technical/sergei/book/index-book.html>> [Consulta: 10 ene. 2005].
- (ROJAS 1998) ROJAS, M. D. «Ontologías de iones monoatómicos en variables físicos del medio ambiente». 1998. Final-Year Project- Facultad de Informática de la Universidad Politécnica de Madrid.
- (STAAB 2000) STAAB, S.; ANGELE, J.; DECKER, S.; ERDMANN, M.; HOTH, A.; MÄDCH, A.; SCHNURR, H. P.; STUDER, R. «Semantic Community Web Portals». En: *INTERNATIONAL WORLD WIDE CONFERENCE (9ª: 2000). WWW9: - Proceedings of the 9th International World Wide Web Conference*. (Amsterdam, Holanda: Elsevier, 2000), p. 473-491.
- (VARGAS-VERA 2001) VARGAS-VERA, M.; MOTTA, E.; DOMINGUE, J.; SHUM, S. B.; LANZONI, M. «Knowledge Extraction by Using an Ontology-based Annotation Tool». En: *Proceedings of the K-CAP'01 Workshop on Knowledge Markup and Semantic Annotation, Victoria B.C., Canada* (Victoria: K-CAP'01,2001).

BIBLIOGRAFÍA COMPLEMENTARIA

- AGUDO DE CEA, G.; ÁLVAREZ DE MON Y REGO, I.; PAREJA LORA, A. «Primeras aproximaciones lingüístico- ontológica de documentos de la Web Semántica: OntoTag». *Revista Iberoamericana de Inteligencia Artificial*. Vol. 1, n. 17 (otoño 2002), p. 37-49.
- ARANO, S. «La ontología: una zona de interacción entre la Lingüística y la Documentación» [recurso electrónico]. *Hipertext.net*, n. 2 (2004). <<http://www.hipertext.net>> [Consulta: 10 ene. 2005].
- BOSCH, M. «Documentos y lenguajes de marcado: conceptos, problemas y tendencias». *El profesional de la información*. Vol. 10, n. 11 (noviembre 2001), p. 4-9.
- CASTELLS, P.; MACÍAS, J. A. «Un sistema de presentación dinámica hipermedia para representaciones personalizadas del conocimiento». *Revista Iberoamericana de Inteligencia Artificial*. Vol. 4, n. 16 (verano 2002), p. 25-34.

- DÜRSTELER, Juan Carlos. «La telaraña o Web semántica se presenta como la nueva revolución de Internet: la promesa es convertir información en conocimiento» [recurso electrónico]. *La revista digital de InfoVis.net*. <<http://www.infovis.net/printMag.php?num=26&lang=1>> [Consulta: 12 nov. 2004].
- DÜRSTELER, Juan Carlos. «La Web Semántica hoy» [recurso electrónico]. *La revista digital de InfoVis.net*. <<http://www.infovis.net/printFicha.php?rec=revista&num=131&lang=1>> [Consulta: 12 nov. 2004].
- FONSECA, Frederico; EGENHOFER, Max; BORGES, Karla A. V. «Ontologias e Interoperabilidade Semântica entre SIGs». [recurso electrónico]. En: WORKSHOP BRASILEIRO DE GEOINFORMÁTICA (2º: Brasil: 2000). *GeoInfo*. <<http://www.spatial.maine.edu/~max/RJ40.html>> [Consulta: 12 nov. 2004].
- GARCÍA JIMÉNEZ, A. «Instrumentos de representación del conocimiento: tesauros versus ontologías». *Anales de documentación*. N. 7 (2004), p. 79-95.
- GÓMEZ-PÉREZ, A.; FERNÁNDEZ LÓPEZ, M.; CORCHO GARCÍA, O. *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. London: Springer, 2004.
- HENDLER, James; BERNERS-LEE, TIM; MILLER, Eric. «Integrating Applications on the Semantic Web». *Journal of the Institute of Electrical Engineers of Japan*. Vol. 122, n. 10 (october 2002), p. 676-680. <<http://www.w3.org/2002/07/swint>> [Consulta: 12 nov. 2004].
- LÓPEZ ALONSO, Miguel Ángel. «El modelo conceptual en los sistemas de procesamiento de la información». *Scire*, vol. 4 (1998), n. 1, p. 29-43.
- LÓPEZ ALONSO, Miguel Ángel. «Las estructuras conceptuales de representación del conocimiento en Internet». *Scire*. vol. 6 (2000), n. 1, p. 107-123.
- LOZANO TELLO, Adolfo. «Ontologías en la Web Semántica» [recurso electrónico]. En: *I Jornadas de Ingeniería Web' 01*. <<http://www.informandote.com/jornadasIngWEB/articulos/jiw02.pdf>>. [Consulta: 12 nov. 2004].
- NILSSON, M.; PÁLMER, M.; NAEVE, A. «Semantic Web Metadata for e-Learning-Some Architectural Guidelines» [recurso electrónico]. <<http://www2002.org/CDROM/alternate/744/>> [Consulta: 12 nov. 2004].
- PEIS, E.; HERRERA-VIEDMA, E.; HASSAN, Y.; HERRERA, J.C., «Análisis de la web semántica: estado actual y requisitos futuros. *El profesional de la información*. Vol. 12, n. 5, (septiembre-octubre 2003), p. 368-376.
- ROLDÁN GARCÍA, M.M.; MORENO VERGARA, N.; ALDANA MONTES, J.F. «Esquemas para representar el conocimiento basado en ontologías». <http://kybele.escet.urjc.es/RedBD/Articulos/Art_03.pdf> [Consulta: 12 nov. 2004].
- SANTACRUZ-VALENCIA, L. P.; AEDO, I.; DELGADO KLOOS C. «Learning Objects: Trends into Semantic Web». *Boletín de la Red Iris*, N. 66-67 (2003-2004).
- SILVA MUÑOZ, Lydia. «Estudo de ontologías para representação de conteúdos de ensino baseado na www» [recurso electrónico]. 2002. <http://www.inf.ufrgs.br/adapt/publicacoes/TI_Lydia.pdf>. [Consulta: 12 nov. 2004].