

Ontologías en la Web Semántica

Adolfo Lozano Tello ¹

¹ *Área de Lenguajes y Sistemas Informáticos. Departamento de Informática. Escuela Politécnica. Avda de la Universidad s/n, 10071 Cáceres. Universidad de Extremadura. España. Tfno: +34 927257195. alozano@unex.es*

Resumen La enorme revolución que supuso el uso generalizado de Internet impulsó el intercambio de información personal, académica y comercial. La Web está a punto de sufrir un nuevo cambio: la información que aparece en Internet va a poder interpretarse por los ordenadores sin necesidad de intervención humana, es la denominada Web Semántica. Para que esto ocurra, es necesario que la información de las páginas web se codifique mediante ontologías. Las ontologías representarán el conocimiento de Internet, definiendo formalmente los conceptos de los diferentes dominios y sus relaciones, con capacidad para realizar deducciones con este conocimiento.

1. Introducción

Actualmente, la Web es un espacio preparado para el intercambio de información diseñado para el consumo humano. Las páginas web son creadas por personas para ser entendidas por personas. No existe un formato común para mostrar la información, por lo cual, los desarrolladores de páginas web crean sus páginas dependiendo de los potenciales usuarios que van a visitarlas.

Los actuales *browsers* de web realizan la búsqueda de información, con más o menos fortuna, mediante palabras clave que aparecerán en el código HTML de las páginas web dispersas en Internet. En los últimos años, algunas empresas están realizando anotaciones de datos introducidas dentro de este código HTML, siguiendo algún esquema de anotación común, normalmente basado en XML.

Otra carencia de la situación actual es que, con los estándares web del momento, no se puede diferenciar entre información personal, académica, comercial, etc. Es decir, cuando un buscador web realiza una consulta con algunas palabras clave, normalmente aparece información que no es útil

porque no corresponde a lo que estamos buscando. Además no todas las páginas proporcionan igual cantidad de información, debido precisamente a que no existe un formato o convenio que nos diga qué contenido debemos añadir a las páginas web.

Por otro lado, los agentes de búsqueda actuales no se diseñan para “comprender” la información que reside en la web, precisamente porque es prácticamente imposible conocer la representación de los datos ubicados en las diferentes páginas.

Si, por ejemplo, queremos buscar en la red a un dentista, que pertenezca a mi compañía de seguro médico “Segurín”, y que se encuentre en mi ciudad “Cáceres”, una forma podría ser indicarle al buscador tres palabras clave: *dentista, Segurín, Cáceres*. La información que puede encontrar el buscador puede estar relacionada con la prensa, demandas judiciales, información que reúna compañías de seguros, ... Normalmente gastamos mucho tiempo en seleccionar la información que nos puede ser útil, y navegando por las referencias URL hasta encontrar, con suerte, lo que estamos buscando. En el ejemplo, no es

seguro que los dentistas tengan páginas personales de su consulta, o que la compañía del seguro tenga un directorio de sus dentistas, y que la información de los dentistas esté relacionada con la ciudad que estoy buscando. Si por casualidad encontrásemos varias páginas personales de dentistas que tuvieran en su anotación HTML esas palabras clave, es bastante improbable que la información que nos proporcionen sea similar; algunos pueden tener datos informativos, otros publicaciones, especialidades, horarios, ...

Es indudable que las ventajas que ofrece Internet son enormes a la hora de buscar información, pero adolece de una manera de encontrar información de forma precisa y de poder realizar deducciones con la información existente. En la sección 2, mostraremos la visión de la web semántica para solucionar estos problemas indicados y veremos qué ventajas puede aportar a la situación actual. En la sección 3 se explicará el concepto de ontología, pieza fundamental para soportar la representación del conocimiento que necesita la web semántica. Y en la sección 4 se verá una perspectiva de los pasos necesarios que se deben abordar para hacer efectivo este adelanto en la comunicación del conocimiento.

2. El Intercambio de Conocimientos en la Web Semántica

En los últimos años, muchos investigadores están diseñando modelos para transformar la red desde un espacio de información a un espacio de conocimientos. Recientemente, Tim Berners-Lee, uno de los inventores de la Web, defiende el desarrollo de la Web con conocimientos [Ber01], y organizaciones como SemanticWeb¹ se encargan de estandarizar lenguajes y herramientas para hacer efectiva la web semántica. Pero, ¿qué se puede hacer en la web semántica?

La idea es que los datos puedan ser utilizados y “comprendidos” por los ordenadores sin necesidad de supervisión humana, de forma que los agentes web puedan ser diseñados para tratar la información situada en las páginas web de manera semiautomática. Se trata de convertir la información en conocimiento, referenciando datos dentro de las páginas web a metadatos con un esquema común consensado sobre algún dominio. Los metadatos no sólo especificarán el esquema de datos que debe aparecer en cada instancia, sino que además podrán tener información adicional de cómo hacer deducciones con ellos, es decir, axiomas que podrán aplicarse en los diferentes dominios que trate el conocimiento almacenado.

Con ello, se mejorará la búsqueda de información y se potenciará el desarrollo de aplicaciones de comercio electrónico, ya que las anotaciones de información seguirán un esquema común, y los buscadores web compartirán con las anotaciones web los mismos esquemas. Empresas que traten con clientes y proveedores, podrán intercambiar sus datos de productos siguiendo estos esquemas comunes consensuados.

Los agentes web no sólo encontrarán la información de forma precisa, si no que podrán realizar inferencias automáticamente buscando información relacionada con la que se encuentra situada en las páginas, y con los requerimientos de la consulta indicada por el usuario.

3. Las Ontologías como Soporte de la Web Semántica

Para que esto pueda llevarse a cabo, se necesita que el conocimiento de la web esté representado de forma que sea legible por los ordenadores, esté consensado, y sea reutilizable. Las ontologías proporcionan la vía para representar este conocimiento.

El término ontología proviene de la filosofía; pero en IA, tiene diferentes

¹ www.semanticweb.org

connotaciones. La definición declarativa más consolidada es la propuesta por Gruber [Gru93a] y extendida por Studer y colegas [Stu98] que la describe como “una especificación explícita y formal sobre una conceptualización compartida”. La interpretación de esta definición es que las ontologías definen conceptos y relaciones de algún dominio, de forma compartida y consensuada; y que esta conceptualización debe ser representada de una manera formal, legible y utilizable por los ordenadores.

Las ontologías tienen los siguientes componentes que servirán para representar el conocimiento de algún dominio [Gru93b]:

Conceptos: son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.

Relaciones: representan la interacción y enlace entre los conceptos del dominio. Suelen formar la taxonomía del dominio. Por ejemplo: *subclase-de*, *parte-de*, *parte-exhaustiva-de*, *conectado-a*, etc.

Funciones: son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como *categorizar-clase*, *asignar-fecha*, etc.

Instancias: se utilizan para representar objetos determinados de un concepto.

Axiomas: son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: “*Si A y B son de la clase C, entonces A no es subclase de B*”, “*Para todo A que cumpla la condición C1, A es B*”, etc.

Estos últimos componentes, los axiomas, permiten junto con la herencia de conceptos, inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos.

Por ejemplo, con el conocimiento anotado en las páginas web mediante ontologías, podremos utilizar un agente web al que podamos preguntar sobre los dentistas que se encuentren a una cierta distancia de mi casa. Una de las posibles respuestas que me podría ofrecer el agente sería:

* Dra. Macías – Dentista - ortodoncia

- Consulta Particular (85€)-
c/Rosa 4 - a 2 km
- Hospital Clínico “La Paz”-
c/Principal 17 - a 5 km.

Incluso si la página web de la Dra. Macías no tuviera especificada la dirección del hospital, el agente web de búsqueda podía utilizar un atributo de los dentistas (*Trabaja_en_Hospital*), y con el valor de este atributo encontrar el atributo *Dirección* del concepto *Hospital* y mostrar esta información del hospital concreto de esa instancia de *Dentista*.

Además de utilizar estas propiedades de relaciones y herencia de los conceptos especificados, mediante los axiomas se tendría mayor capacidad expresiva del dominio almacenado. Por ejemplo, si en este dominio tenemos declarado el axioma:

“*Si el médico trabaja en un hospital de mi seguro no tendré que pagar minuta.*”

El agente podría, utilizando el conocimiento representado en los conceptos, sus relaciones y utilizando el axioma, aconsejarnos sobre los dentistas que cumplieran este requisito.

4. Cómo Alcanzar la Web Semántica

Para poder explotar la web semántica, se necesitan lenguajes de marcado apropiados que representen el conocimiento de las ontologías. Actualmente, mediante anotaciones

RDF-RDF Schema se pueden representar algunas facetas sobre conceptos de un dominio y permite, mediante relaciones taxonómicas, crear una jerarquía de conceptos. Pero se necesitan lenguajes de marcado (basados en RDF) con mayor expresividad y capacidad de razonamiento para representar los conocimientos que contienen las ontologías [Cor00]. De esta forma, existen ya disponibles herramientas como Protégé², OntoEdit³, o WebOnto⁴ para realizar anotaciones en páginas web con lenguajes de marcado propios. El lenguaje con gran capacidad expresiva que está emergiendo como un estándar para realizar anotaciones de ontologías en web es DAML⁵, aunque en este momento no tiene sus formatos totalmente definidos.

Por otro lado, se necesitan agentes y aplicaciones web que exploten este conocimiento anotado en las páginas web. Estos agentes de conocimientos web serán capaces de interpretar los esquemas ontológicos y axiomas de diferentes dominios, mantendrán la consistencia de las instancias que se inserten en las páginas web siguiendo los esquemas ontológicos definidos, realizarán una búsqueda con inferencias utilizando los axiomas situados en los esquemas, y podrán realizar ligaduras de los árboles taxonómicos de varias ontologías.

Para potenciar el uso de ontologías en la web, se necesitan aplicaciones específicas de búsqueda de ontologías, como (Onto)²Agent⁶ [Arp00], que indiquen a los usuarios las ontologías existentes y sus características para poder utilizarlas en su sistema.

5. Conclusiones

La Web Semántica proporcionará un salto cualitativo sobre el potencial de la Web. Las principales ventajas de esta nueva revolución en Internet serán el desarrollo de aplicaciones con esquemas de datos comunes, fomento de las transacciones entre empresas por comercio electrónico y búsqueda de información con inferencias.

Para poder lograr estos objetivos se necesita unificar los contenidos semánticos por medio de ontologías que formalicen este conocimiento de forma consensuada y reutilizable.

Se necesita un lenguaje común basado en web, con suficiente capacidad expresiva y de razonamiento para representar la semántica de las ontologías; este hecho parece que lo veremos en muy poco tiempo. Los futuros agentes se diseñarán para explotar el conocimiento de la web.

Referencias

- [Arp00] Arpirez J., Gómez-Pérez A., Lozano Tello A. and Pinto S. "Reference Ontology and (Onto)²Agent: The Ontology Yellow Pages", Knowledge and Information Systems, An International Journal, Springer-Verlag, 2 (2000) 4, 387-412. Mar. 2000.
- [Ber01] Berners-Lee T., Hendler J. and Lassila O, "The Semantic Web", Scientific American, Volume 284, Number 5 (May, 2001), pages 34-43.
- [Cor00] Corcho O., Gómez-Pérez A. "A Roadmap to Ontology Specification Languages". EKAW 2000 pp. 80-96
- [Gru93a] Gruber T., "Toward Principles for the Design of Ontologies Used for Knowledge Sharing" Technical Report KSL-93-04, Knowledge Systems Laboratory, Stanford University, CA, 1993.
- [Gru93b] Gruber T., "A Translation Approach to Portable Ontology Specifications", Knowledge Acquisition, 5 (2), 199-220, 1993b.
- [Stu98] Studer S, Benjamins R., and Fensel D., "Knowledge Engineering: Principles and Methods", Data and Knowledge Engineering, 25, 161-197, 1998.

² <http://protege.semanticweb.org>

³ <http://ontoserver.aifb.uni-karlsruhe.de/ontoedit/>

⁴ <http://kmi.open.ac.uk/projects/webonto/>

⁵ <http://www.daml.org>

⁶ <http://delicias.dia.fi.upm.es/OntoAgent>