

Gestión de la información a través de la Web Semántica: Iniciativas y dificultades

Marcano Aular, Yelitza Josefina*
Talavera Pereira, Rosalba**

Resumen

Actualmente los motores de búsqueda tradicionales devuelven listas de resultados que ofrecen escasa o nula información acerca de las relaciones semánticas existentes entre los documentos. Por su parte, los gestores del conocimiento gastan una cantidad considerable de tiempo revisando las fuentes e índices para descubrir el modo en que se relacionan entre sí los documentos. Sin embargo, el trabajo de los gestores comienza a ser productivo sólo cuando éstos comienzan a localizar las similitudes y las diferencias entre partes de información que les permiten construir relaciones para crear nuevo conocimiento. A partir de esta descripción se pretende en este artículo, destacar el apoyo que puede llegar a ofrecer la Web Semántica a la Gestión de la información, iniciativas en el área y algunas dificultades a las que se enfrenta la nueva Web. La metodología de investigación está dedicada al análisis del contenido como principal herramienta en donde se derivan algunas categorías teóricas referidas a: la Web Semántica, Internet y la Gestión de la información. El artículo concluye señalando que la información está generalmente dispersa a lo largo de la Web y/o las Intranets en forma de recursos, y a través de la Web Semántica se puede facilitar su procesamiento, para generar sitios Web que se adapten y configuren a los diferentes perfiles de usuario, preferencias y requerimientos.

Palabras clave: Web Semántica, Internet, Gestión de la información.

Recibido: 06-02-03. Aceptado: 06-06-06

* Ingeniero de Sistemas. Especialista en Costos. Maestrante del Postgrado Gerencia de Empresas de la Universidad del Zulia. Cursante del segundo año del Doctorado en Informática de la Universidad Politécnica de Madrid. Profesora Asociada de la Universidad del Zulia, Núcleo Punto Fijo. E-mail: ymarcanoa@hotmail.com.

** Ingeniero en Computación. Cursante del segundo año del Doctorado en Informática de la Universidad Politécnica de Madrid. Profesora Agregada de la Universidad del Zulia, Núcleo Punto Fijo. E-mail: talavera_p@hotmail.com.

Administration of the Information Through Web Semantics: Initiatives and Difficulties

Abstract

Nowadays, traditional search mechanisms produce lists of results that offer little or no information about existing semantic relationships among documents. On the other hand, the agents of knowledge spend a considerable quantity of time checking sources and indexes to discover the way in which the documents relate to one another. However, the work of the agents begins to be productive only when they begin to locate the similarities and the differences among these pieces of information that allow them to build relationships to create new knowledge. Starting from this description, the purpose of this article is to highlight the support that Web Semantics offers to the information management initiatives in this area and the difficulties that the new Web faces. The research methodology is devoted to the analysis of content as the main tool wherein certain theoretical categories are derived. These are referred to as: Web semantics, Internet and Administration of information. The article concludes pointing out that information is generally dispersed along the Web and/or the intranets in the form of resources, and through Web semantics its processing can be facilitated in order to generate Web sites that adapt and configure to the different user profiles, preferences and requirements.

Key words: Web semantics, internet, administration of the information.

1. Introducción

La Web Semántica es una extensión de la Web actual en la cual se dota a la información de un significado (explícito) bien definido (inteligible por los computadores), mejorando las posibilidades de que los computadores y las personas cooperen (Berners-Lee, 2001a). En sí, la idea es definir y enlazar la información disponible en la Web de modo que ésta pueda ser utilizada de manera más efectiva para el descubrimiento, automatización, integración y reutilización entre varias aplicaciones. Para ser más explícita la definición se destaca la conceptualización que ofrecen Berners-Lee y otros (Berners-Lee, et al, 2001) los cuales dan a entender que la Web Semántica es una extensión de la Web actual en la cual se dota a la información de significado bien

definido para que tanto personas como computadoras puedan trabajar cooperativamente (Berners-Lee, et al, 2001). Es decir, utilizando la propia estructura tecnológica de la Web se pretende dotar a los recursos que en ella residen de algún elemento que añada información sobre la propia información, de manera que esta primera facilite la comunicación "con sentido semántico" entre las máquinas que forman la red y, por lo tanto, redunde en que las personas que la utilizan puedan obtener lo que desean con mayor rapidez y fiabilidad. Así pues, el objetivo último de esta nueva Web es que se produzca un intercambio de información efectivo y eficiente (Miller, 2001).

Para que las máquinas puedan llevar a cabo esta función necesitan acceder a colecciones estructuradas de información y a formalismos actualmente ba-

sados en lógica matemática que les permitan tener un cierto grado de razonamiento automático; estas necesidades pueden cubrirse utilizando ontologías¹ para anotar los recursos Web.

La Web Semántica mantiene los principios que han hecho un éxito de la Web actual, como son los principios de descentralización, compatibilidad, o la apertura al crecimiento y uso no previstos de antemano. A su vez, rescata la noción de ontología del campo de la Inteligencia Artificial como vehículo para cumplir este objetivo (Gruber, 1993). Con base a lo expuesto se puede decir que 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. Durante los últimos años se han desarrollado diversos lenguajes y estándares para la definición de ontologías, entre ellos destacan el XML (Lenguaje eXtensible de Marcado), RDF (Resource Description Framework o Marco de Descripción de Recursos), DAML+OIL (Darpa Agent Markup Language + Ontology Interchange Language), y más recientemente OWL (sucesor de DAML+OIL), respaldados por el consorcio W3C, uno de los principales promotores de la Web Semántica. Algunos campos de aplicación de la Web Semántica son: el comercio electrónico, la bús-

queda de información eficiente en la Web, el procesamiento del lenguaje natural, el diseño instruccional, la gestión de librerías digitales, el turismo y la gestión del conocimiento corporativo, entre otros.

En muy pocos años se ha consolidado una comunidad internacional con un importante respaldo de los programas de financiamiento de agencias norteamericanas y europeas, para promover la investigación y desarrollo en el campo de la Web Semántica. Entre los grupos líderes de esta comunidad se pueden citar al consorcio W3C, el Knowledge Systems Laboratory de la Universidad de Stanford, la Universidad Libre de Amsterdam (Dieter Fensel), el Knowledge Management Research Group de la Universidad de Karlsruhe en Alemania, el Information Management Group de la Universidad de Manchester, el Knowledge Media Institute de la Open University en el Reino Unido, el Consejo Nacional de Investigación (Nicola Guarino) en Italia, y el Laboratorio de Inteligencia Artificial de la Universidad Politécnica de Madrid.

De igual manera, existe un gran interés desde el entorno corporativo, el sector público y el mundo académico por hacer de la Web Semántica una realidad, ya que se piensa que puede ser una pieza importante para el progreso de la sociedad de la información. Entre las principales líneas de trabajo que están siendo ob-

1 El término Ontología utilizado en filosofía para hablar acerca de una teoría sobre la existencia, ha sido adoptado por la comunidad de investigadores de inteligencia artificial para definir una categorización y las relaciones entre sus términos. Es en esta última concepción que se abordarán los planteamientos de este artículo. De esta manera las ontologías tratan de capturar conocimiento consensuado en un modo genérico y formal, de tal manera que puedan ser reutilizadas y compartidas por distintas aplicaciones de software y grupos de personas.

jeto de atención cabe citar: los lenguajes de definición de ontologías, las metodologías de desarrollo de ontologías, la integración de ontologías, el aprendizaje de ontologías, el desarrollo de vocabularios en dominios concretos, los agentes inteligentes y los servicios Web (Castells, 2001).

El presente artículo enfatiza a partir un enfoque descriptivo y exploratorio, la construcción de un sistema teórico en relación a las variables de investigación: Web Semántica, Internet y Gestión de la información. Para discernir sobre estos tópicos se señala en primer orden el estudio de las referencias teóricas existentes, continuando con una búsqueda de relaciones entre los diferentes componentes teóricos y operativos de cada una de las variables. En correspondencia con estos planteamientos se hace uso del análisis de contenido buscando disgregar en un contexto teórico el tratamiento que le dan a éstos, distintos autores.

De igual manera, se conceptualiza a la Web Semántica como un instrumento que permita utilizar de manera más efectiva la información disponible en la red, para el descubrimiento, automatización, integración y reutilización de ésta entre varias aplicaciones, de manera colaborativa. Además se presenta una revisión sobre los aportes en Web Semántica derivados de las investigaciones de autores reconocidos en el área de la Tecnología de la Información y las Comunicaciones, elementos que permiten al investigador reflexionar sobre las iniciativas y dificultades para el tratamiento de la información a través de la Web Semántica.

2. Gestión de la información en la Web

Desde sus inicios el hombre ha buscado mecanismos de cómo representar y tratar la información que genera su entorno y por supuesto él mismo. Hoy día a, través de la gestión de la información, se han desarrollado herramientas, técnicas y procesos que permiten gestionar de la forma más efectiva y eficiente posible los activos intelectuales de una organización apoyados por la Web, haciendo posible que el hombre y las organizaciones tengan a su disposición herramientas para mejorar y hacer más eficiente el procesamiento de la información proveniente de múltiples fuentes.

A continuación se presentan algunos beneficios que acarrea el contar en la Web, con millones de páginas en: Intranets corporativas, documentos en la World Wide Web, millones de usuarios de todas las edades; como punto de apoyo para el auge de la Web Semántica:

- Posibilidad de desarrollo de una visión holística de la información en la organización, donde la tecnología juegue un papel importante en la cultura organizativa.
- Proliferación de Servicios Web: publicación, localización, composición y orquestación de servicios basados en la Web Semántica.
- Posibilidad de contar con un conocimiento distribuido, dinámico y ubicuo, haciendo uso de servidores de conocimiento, motores semánticos y sistemas de gestión de ontologías. Cabe destacar con respecto a este aspecto que la www es el instrumento de despliegue ideal para este conocimiento

ubicuo y sus tecnologías asociadas (Hayes-Rot, 2002). Varios proyectos realizados en la segunda mitad de los 90, financiados por el US DARPA y el programa IST de la UE (e.g. On-To-Knowledge, IST-1999-10132) demostraron la utilidad de ligar ontologías machine-readable a recursos en la Web. Se prevé que la Web Semántica supondrá una revolución tan importante como la Web original en sí (Hayes-Rot, 2002).

Diversas opiniones se han producido en torno al tema de Internet y la Web Semántica y cómo podría mejorar la gestión de la información y a futuro del conocimiento, una de ellas tuvo lugar en la II Jornada de en.red.ando, realizada en los Estados Unidos en el año 2001, en ella se reunieron investigadores de primera línea procedentes de EEUU y Europa, que están trabajando para conseguir que Internet se convierta en una red inteligente por dos vías diferentes; una vía apuesta por un sistema que permita que la búsqueda y recuperación de información sea más intuitiva y precisa a partir de la aplicación de la semántica, de ontologías y de agentes inteligentes (Berners-Lee, et al, 2001), la otra vía de investigación persigue que los servidores sean capaces de aprender el recorrido virtual que hacen los internautas en su navegación por la red para que ésta acabe funcionando como un cerebro global.

Desde la perspectiva presentada se puede decir aun cuando existen varias iniciativas para la mejorar la gestión de la información y a futuro del conocimiento, actualmente el impacto potencial de toda esta tecnología no ha modificado sustancialmente la vida en la Web. Los estándares avanzan lentamente, y el esfuerzo de diseñar ontologías interoperables y, sobre todo, de introducir la enorme cantidad de metadatos para indexar apropiadamente los datos existentes, hacen que su avance sea lento.

3. Contribución de la Web Semántica al tratamiento eficaz de la información

La Web actual está basada en el lenguaje HTML, el cual especifica cómo dar formato a una página para que ésta pueda ser leída por humanos. Este lenguaje no puede explotarse mediante técnicas de recuperación de información para mejorar el resultado, el cual tiene que basarse en las palabras que forman el contenido de la página, de ahí que esté restringido básicamente al uso de palabras clave. En cambio en la Web Semántica, las páginas almacenan el contenido como un conjunto de palabras relacionadas en un documento, incluyen su significado y su estructura. En este mismo orden de ideas, cabe destacar que todos los lenguajes de la Web Semántica² están basados en XML, lenguajes mucho más ricos y potentes que el HTML³, los

2 Algunos de estos lenguajes son: el RDF, RDFS, OIL, DAML+OIL y OWL, entre otros.

3 Lenguaje que sólo permite dar formato a una página para que sea intellegible para la lectura de un ser humano.

cuales permiten, en mayor o menor medida, representar el significado y la estructura del contenido, es decir las interrelaciones entre conceptos.

Para obtener esa adecuada definición de los datos, según expresa el Consorcio de la World Wide Web (W3C, 2005), la Web Semántica utiliza dos estándares que ayudan a convertir la Web en una infraestructura global en la que es posible compartir, y reutilizar datos y documentos entre diferentes tipos de usuarios, estos son el RDF y el OWL. El RDF proporciona información descriptiva simple sobre los recursos que se encuentran en la Web y que se utiliza, por ejemplo, en catálogos de libros, directorios, colecciones personales de música, fotos, eventos, entre otros.

Por su parte, el OWL es un mecanismo para desarrollar temas o vocabularios específicos en los que asociar esos recursos. Es decir, proporciona un lenguaje para definir ontologías estructuradas que puedan ser utilizadas a través de diferentes sistemas. Las ontologías, que se encargan de definir los términos utilizados para describir y representar un área de conocimiento, son utilizadas por los usuarios, las bases de datos y las aplicaciones que necesitan compartir información específica, es decir, en un campo determinado como puede ser el de las finanzas, medicina, deporte o educación, de esta manera las ontologías incluyen definiciones de conceptos básicos en un campo determinado y la relación entre ellos.

Así pues, la reutilización de contenido entre sistemas distintos requiere que los objetos de aprendizaje estén estandarizados, por lo que muchas organizacio-

nes dedican sus esfuerzos al desarrollo de estándares, especificaciones y modelos de referencia que faciliten la interoperabilidad y la reutilización de objetos de aprendizaje (Santacruz et al, S/F). Dentro de estas organizaciones sobresale el IEEE que cuenta con el estándar de meta-datos LOM, que es el primer esquema de meta-datos acreditado para tecnología de aprendizaje. El propósito de este estándar es simplificar las operaciones de búsqueda, gestión e intercambio de objetos de aprendizaje dentro de la Web. Otra de las organizaciones sobresalientes es el IMS, constituido por diferentes grupos de trabajo dedicados al desarrollo de especificaciones relacionadas, entre otros temas, con el diseño de contenido reutilizable para sistemas de gestión de contenido de aprendizaje.

Algunos de los proyectos más destacados sobre la Web Semántica, son el RSS y el FOFA. El RSS es un vocabulario RDF basado en XML que permite la catalogación de información (noticias y eventos) de tal manera que sea posible encontrar información precisa, adaptada a las preferencias de los usuarios. Los archivos RSS contienen metadatos sobre fuentes de información especificadas por los usuarios cuya función principal es avisar a los usuarios de que los recursos que ellos han seleccionado para formar parte de esa RSS han cambiado sin necesidad de comprobar directamente la página, es decir, notifican de forma automática cualquier cambio que se realice en esos recursos de interés seleccionados. Otro proyecto que permite crear páginas Web, para describir personas, vínculos entre ellos, y cosas que hacen y crean, es el FOFA. Este proyecto se trata de un voca-

bulario RDF, que permite tener disponible información personal de forma sencilla y simplificada para que pueda ser procesada, compartida y reutilizada.

Aun cuando tal como se ha señalado, existen varias iniciativas para simplificar las operaciones de búsqueda, gestión e intercambio de objetos de aprendizaje dentro de la Web, la catalogación de la misma, entre otras; actualmente la Web Semántica presenta algunos inconvenientes que es necesario solventar, los cuales se describen a continuación.

4. Dificultades que enfrenta la Web Semántica

Entre las dificultades a las cuales se enfrenta la Web Semántica están:

- La disponibilidad de contenido. Actualmente, hay poco contenido relevante sobre la Web Semántica. El contenido de las Webs existentes debe ser actualizado a contenido para la Web Semántica, incluyendo páginas de HTML estático, contenido existente XML, y contenido dinámico, multimedia y servicios Web.
- Disponibilidad, desarrollo y evolución de ontologías. Actualmente se está haciendo énfasis en el desarrollo de ontologías que traten de capturar el conocimiento consensado en un modo genérico y formal de tal manera que pueda ser reutilizados y compartidos por distintas aplicaciones de software y grupos de personas, pero aun no se ha llegado a definir esto con precisión. De esta manera las ontologías se convertirán en una pieza clave, ya que permiten explícitamente la semántica del contenido de la Web Semántica. Por esta razón se debe

realizar un gran esfuerzo en la creación y evolución de ontologías comúnmente usadas para la Web Semántica, en proporcionar infraestructuras adecuadas para el desarrollo, gestión y mapeo de ontologías y en el entorno distribuido de la Web.

- Escalabilidad. Otra dificultad está en los procedimientos para organizar el contenido de la Web Semántica, almacenarlo y proporcionar los mecanismos para encontrarlo. Todas esas tareas deben ser realizadas y coordinadas de forma escalable, ya que las soluciones deben estar preparadas para el crecimiento de la Web Semántica.
- Soporte multilinguaje. Este problema ya existe en la Web actual, y también debe ser tenido en cuenta en la Web Semántica. Cualquier aproximación a la Web Semántica debe proporcionar mecanismos para acceder a la información en diferentes lenguajes, permitiendo la creación y el acceso al contenido independientemente del lenguaje original de los proveedores de contenido o los usuarios.
- Visualización. Este inconveniente se refiere a que los usuarios requerirán de procedimientos sencillos para el reconocimiento de la información relevante de la Web Semántica, para sus propósitos particulares. Por tanto deben explorarse nuevas técnicas que difieren de la visualización usual de la estructura mediante hipertexto de la Web actual.
- Estabilidad de los lenguajes de la Web Semántica. De igual manera representa un inconveniente la falta de estándares en este campo emergente, aun cuando se están realizando esfuerzos para ello.

- Sobrecarga de información. Debido al exceso de la información presente en Internet se hace cada vez más difícil para los usuarios encontrar información pertinente que aporte valor agregado a las búsquedas. Para resolver esto será necesario usar Agentes de software autónomos, que podrán mostrar varios niveles de comportamiento inteligente de estos agentes, que irán desde simplemente reactivo hasta comportamiento adaptativo y con aprendizaje, donde éstos aprendan qué les gusta o les disgusta a los usuarios, lo cual permita aislar a los usuarios de información irrelevante, y sin valor agregado. Estos agentes deberían de "aprender" a funcionar en "entornos sociales" donde será necesario colaborar, competir o negociar con otros agentes.

Por último y no menos importante, cabe resaltar las observaciones que hacen Santacruz y otros (Santacruz et al, S/F) los cuales expresan que la iniciativa de la Web Semántica en su estado actual no proporciona tal semántica, pues sólo ofrece descripciones para los recursos y no dice nada sobre cómo presentar los recursos a los usuarios en una forma conceptualmente clara. Es por ello, que según los autores 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.

5. Conclusiones

Una de las actividades más recientes en los desarrollos orientados a la Web, es la Web Semántica, cuya finali-

dad es dotar de significado a todas las clases de información sobre la Web. Un subconjunto importante de esa información lo representan los objetos de aprendizaje, que son recursos digitales que se pueden reutilizar en diferentes contextos para lograr un objetivo de aprendizaje particular.

Así pues, los antecedentes y premisas esbozadas a lo largo de este artículo, permiten inferir que la Web Semántica, busca modificar la forma en que se presenta la información en la Web de un modo que facilite el procesamiento de la misma por parte de los computadores, para que se puedan generar automáticamente documentos y sitios Web adaptativos, a los cuales se les permita reconfigurarse dinámicamente de acuerdo a perfiles de usuario, preferencias, y requerimientos particulares de cada usuario. De igual manera, analizando la bibliografía consultada, se puede concluir que la Web Semántica tiene como visión la búsqueda de una Web más inteligente en la que se pueda conseguir una comunicación efectiva entre computadores y centra sus esfuerzos principalmente en la búsqueda de descripciones enriquecidas semánticamente para los datos en la Web.

Para terminar, cabe destacar que aunque la Web Semántica no es todavía una realidad absoluta, sí se prevén muchas ventajas, ya que las búsquedas al ser más precisas arrojarán mejores resultados cuando la web semántica sea una realidad; ya que será posible combinar informaciones que residen en diferentes páginas web que ahora no tienen ninguna conexión entre ellas. Y al dotar a la Web de más significado, es decir de más semántica, se podrían obtener soluciones a pro-

blemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, sea posible compartir, procesar y transferir información de forma sencilla, mejorando así la gestión de la información. Cabe destacar además que esta Web extendida y basada en el significado, se podría apoyar en lenguajes universales que resuelvan los problemas ocasionados por una Web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea ardua y frustrante.

Referencias Bibliográficas

- Berners-Lee, Tim, James Hendler, Ora Lassila (2001). **The Semantic Web**. Scientific American.
- Berners-Lee, Tim, Hendler, James (2001a). "Scientific publishing on the semantic web". **Nature**. April.
- Castañeda, Javier (2001). **La Web inteligente**. BAQUIA Knowledge Center.
- Catells, Pablo. Aplicación de técnicas de la Web Semántica. Universidad Autónoma de Madrid (Página Web en línea). Disponible: <http://giig.ugr.es/~mgea/coline02/Articulos/pcastells.pdf>. Consulta: 18-10-05.
- Consortio de la World Wide Web (2005). Disponible: <http://www.w3.org>. W3C. World WideWebConsortium. Oficina Española. Consulta: 11-10-05.
- García Barriocanal, Elena. Sicilia Urbán, Miguel Ángel (S/F). **Una propuesta para la búsqueda semántica de recursos web de nuevas tecnologías aplicadas a la educación**. Universidad de Alcalá de Henares/Universidad de Carlos III. (Página web en línea). Disponible: <http://tecnologiae-du.us.es/ticsxxi/comunic/egb-masu.htm>. Consulta: 21-10-05.
- Gruber, Tom (1993). **A Translation Approach to Portable Ontology Specifications**. Knowledge Acquisition. Vol. 5: 199-220.
- Hayes-Roth, Richard (2002). CTO for Software, Hewlett-Packard (Business Week).
- IMS Global Learning Consortium Inc. Overview of Specifications. Disponible: <http://www.imsglobal.org/overview.cfm>. Consulta: 15-09-05.
- LOM Standard. Draft Standard for Learning Object Metadata, IEEE P1484.12/D4.0. Disponible: http://ltsc.ieee.org/doc/wg12/LOM_Wd4.doc. Consulta: 12-09-05.
- Miller, Eric (2001). **The W3C semantic web activity**. Proc. International Semantic Web Workshop. Julio.
- Santacruz-Valencia, Liliana Patricia, Aedo, Ignacio., Delgado Kloos, Carlos (S/F). **Objetos de aprendizaje: Tendencias dentro de la Web Semántica**. Universidad Carlos III Madrid. (Página web en línea). Disponible: <http://www.rediris.es/rediris/boletin/66-67/ponencia18.pdf>. Consulta: 14-10-05.