

WEB SEMÁNTICA: Principios y Estándares

Samanta Patricia Cueva

Abril del 2008

Resumen

Actualmente en Internet se tiene acceso a demasiada información la cual no se encuentra clasificada correctamente lo que influye en los resultados de búsquedas que los usuarios requieren a través de los motores de búsquedas y que no siempre se ajustan a lo que se requiere, es por eso que desde 1999 se está hablado de la Web Semántica, la cual va a permitir proporcionar al usuario respuestas específicas que concuerden con los requerimientos de los mismos. Pero no solo los motores de búsqueda deben estar listos para enfrentar a esta nueva web, sino que también los portales y sitios web deben empezar a migrar de la web sintáctica a la web semántica, ya que esto influirá en el ranking y posicionamiento web que obtengan dichos sitios.

En este artículo se revisan conceptos, arquitectura, estándares que están inmersos en la web semántica y los pasos que se deberían seguir para migrar los sitios web a la web semántica.

Palabras Clave: Web Semántica, W3C, RDF, OWL, Accesibilidad

1. INTRODUCCIÓN

El concepto de web semántica fue creado por Tim Berners-Lee creador de la WWW, quien la define así:

“El primer paso es colocar los datos en la Web de un modo en que las máquinas puedan entenderlos naturalmente o convertirlos a esa forma. Esto crea lo que yo llamo una Web Semántica: una red de datos que pueden ser procesados directamente o indirectamente por máquinas.” [1]

“The Web of data with meaning in the sense that a computer program can learn enough about what the data means to process it.” [2]

“La Web Semántica es una extensión de la actual web en la cual la información se da mediante un significado bien definido, lo que facilita que los ordenadores y la gente trabajen en cooperación”. [6]

La Web Semántica aspira a mejorar sustancialmente la interacción entre los sistemas informáticos y los seres humanos, dotando a los primeros de una mayor inteligencia y autonomía, y a los segundos de un nuevo infouniverso conceptual ergonomizado intelectivamente que satisfaga la necesidad de convertir la información en conocimiento. [3]

En palabras de Berners-Lee, “no se está hablando de una inteligencia artificial mágica que permita a los ordenadores entender las palabras de los usuarios, sino únicamente de la habilidad de una máquina para resolver problemas bien definidos, a través de operaciones bien definidas que se llevarán a cabo sobre datos bien definidos”. (W3C, 1999).

En la actualidad la mayor parte de la información que se encuentra en la web, carece de semántica ya que están basados en HTML y por ende cuando se realizan búsquedas los resultados no son lo suficientemente concretos para ajustarse a las necesidades de los usuarios.

En conclusión la Web Semántica es añadir metadatos semánticos a los contenidos que existen en la web, estos metadatos deben ser entendibles tanto por los humanos como por las máquinas.

La web semántica permitirá el uso de agentes personales encargados de extraer información de múltiples fuentes, proporcionándole al usuario respuestas específicas que se ajusten a los requerimientos de los mismos.

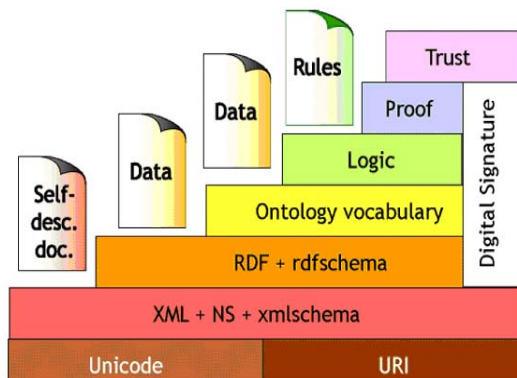


Figura 1: Arquitectura de Web Semántica presentada por Tim Berners-Lee

2. ARQUITECTURA DE LA WEB SEMÁNTICA

En el año 2000 Tim Berners definió la arquitectura de la web semántica en diferentes niveles, los cuales se encuentran representados en la Figura 1:

En el gráfico anterior, se pueden observar diferentes niveles, los que se proceden a explicar brevemente, basados en el artículo "The Semantic Web Revisited":[5]

Primer Nivel: En este nivel se realiza la identificación de recursos web, para lo cual se basa en:

1. **UNICODE:** Es un estándar que proporciona el medio por el cual un texto en cualquier forma e idioma pueda ser codificado para el uso informático.
2. **URI: (Uniform Resource Identifier):** Sirve para identificar un recurso web.²

Segundo Nivel: Es el nivel sintáctico, soluciona el problema de definición de distintos lenguajes de etiquetados para lo cual añade contenido semántico a las páginas.

1. **XML (Extensible Markup Language):** Permite estructurar datos y documentos en forma de árboles de etiquetas con atributos.³ Se lo ha diseñado para reemplazar el estándar HTML.

² <http://www.w3.org/Addressing/>
³ <http://www.w3.org/XML/>

2. **XMLS (XML Schema):** Se pueden fijar anticipadamente las estructuras que se van a utilizar, así como manejar tipos de datos primitivos y derivados.⁴

Tercer Nivel: En este nivel se realiza la descripción de Recursos utilizando básicamente RDF que es un estándar recomendado por la W3C para la definición y uso de descripciones de metadatos

1. **RDF (Resource Description Framework)**⁵: Es un lenguaje para la definición de ontologías y metadatos en la web, siendo el estándar más extendido en la comunidad de web semántica.
2. **RDF Schema:** Es una extensión semántica de RDF⁶

Cuarto Nivel: Es el nivel de las Ontologías, en la que se fundamenta principalmente la propuesta de Berner-Lee.

1. **Ontología:** Define los términos utilizados para describir y representar áreas de conocimiento, se utiliza en aplicaciones que necesitan compartir información de un determinado tema que se lo conoce como dominio.

Una ontología es una definición formal explícita, y estructurada en taxonomías, de un cuerpo de conocimiento. [4]

El proyecto DAML⁷ para el desarrollo de ontologías ofrece un directorio de las mismas que pueden ser utilizadas en definición de ontologías.

OWL Web Ontology Language es el lenguaje estándar de la Web semántica para expresar y codificar ontologías. De acuerdo con el W3C:

OWL está concebido para ser utilizado cuando la información contenida en los documentos necesita ser procesada por aplicaciones informáticas, en oposición a las situaciones donde el contenido solamente debe ser presentado a seres humanos. OWL puede ser utilizado para representar explícitamente el significado de términos en vocabularios y las relaciones [semánticas] entre esos términos. ⁸

⁴ <http://www.w3.org/XML/Schema>
⁵ <http://www.w3.org/RDF/>
⁶ <http://www.w3.org/TR/rdf-schema/>
⁷ <http://www.daml.org/ontologies/>
⁸ <http://www.w3.org/TR/2004/REC-owl-features-20040210/>

Quinto Nivel: Es el nivel de la *Lógica* en la cual se pretende dar flexibilidad a la arquitectura para realizar consultas e inferir conocimiento a partir de las ontologías.

Sexto Nivel: Es el nivel de prueba en el cual se considera que un ordenador alcanza la máxima fiabilidad en sus razonamientos cuando es capaz de realizar demostraciones o cuando es capaz de justificar el motivo por el cual tomó una decisión.

Séptimo Nivel: Es el nivel de confianza en el cual está permitido asignar grados de confianza y seguridad a los diferentes recursos web, a través de la utilización de firmas digitales y técnicas de autenticación de red. En estos niveles todavía no hay recomendaciones por parte del W3C, sin embargo existe un Grupo de Trabajo de Contexto Web Seguro que el 3 de Abril del 2008 publicó un borrador sobre Contexto Web Seguro; experiencia, indicadores y confianza⁹. En el cual se tratan recomendaciones sobre cómo los agentes de usuario web han de presentar información sobre su identidad y la forma de actuar ante errores en protocolos de seguridad, de tal forma que se minimice el número de decisiones de seguridad que deben tomar los usuarios.

Actualmente solo los tres primeros niveles disponen de desarrollo, los otros niveles tienen una gran cantidad de literatura técnica pero para efectos prácticos todavía está en sus inicios. Por lo tanto los pilares más importantes hasta el momento del proyecto de web semántica son XML, RDF y OWL.

3. LA WEB SEMÁNTICA EN LA ACTUALIDAD

La Web semántica aún no es una realidad, de acuerdo con las estimaciones del W3C, el despliegue total de la Web semántica puede prolongarse más allá del año 2010. Sin embargo, la Web semántica ya está entre nosotros de diversas formas:

- A través de una idea que ha sido capaz de agrupar a varias personas y empresas, las mismas que trabajan para hacer realidad la web semántica. Es así que empresas importantes del mundo como Microsoft, IBM, SUN, Adobe;

⁹ <http://www.w3.org/TR/2008/WD-wsc-xit-20080403/>

han apostado a realizar investigación en esta área.

- Por otro lado la creación de nuevos estándares que ya se están utilizando actualmente como el lenguaje XML y la creación de una nueva generación de navegadores, buscadores y editores de páginas web.

Sin embargo los avances de la web semántica son demasiado lentos debido a que involucra una combinación de técnicas de inteligencia artificial que simulen la capacidad humana de realizar tareas extremadamente complejas para un ordenador. Las primeras aproximaciones de estas combinaciones de técnicas son *Mechanical Turk de Amazon*¹⁰ y *Google Image Labeler*¹¹.

Por otro lado Yahoo ha lanzado *Yahoo! Search Open Platform*¹² la cual está soportando estándares RDF y microformatos; Google ha sacado una nueva API que hace uso de estándares XFN y FOAF con lo cual Google va a afrontar las búsquedas sociales y se llama *Social Graph API*.¹³

Como ejemplos de Web Semántica también se pueden citar a:

- **RSS:** Es un formato sencillo de datos que se utiliza para distribuir contenidos a suscriptores de un sitio web; se conoce como redifusión web o sindicación web.
- **FOAF (Friend of A Friend)**¹⁴: Es un proyecto dentro de la Web semántica para describir relaciones mediante RDF que puedan ser procesadas fácilmente por máquinas.
- **DOAP**¹⁵: Es un proyecto para crear vocabulario de XML/RDF para describir proyectos de software.
- **SIMILE**¹⁶: Desarrolla herramientas de fuente abierta que permiten que usuarios puedan tener acceso, manejar, visualizar y reutilizar activos digitales.

¹⁰ <http://www.mturk.com/mturk/welcome>

¹¹ <http://images.google.com/imagelabeler/>

¹² <http://www.ysearchblog.com/archives/000527.html>

¹³ <http://code.google.com/apis/socialgraph/>

¹⁴ <http://www.foaf-project.org/>

¹⁵ <http://trac.usefulinc.com/doap>

¹⁶ <http://simile.mit.edu/>

En los centros de investigación la web semántica se ha convertido en una de las áreas de investigación prioritaria entre los grupos de investigación más destacados se puede mencionar el Instituto Tecnológico de Massachusetts, la Universidad de Stanford, La Universidad de Maryland, la Universidad de Innsbruck, la Universidad de Karlsruhe, la Universidad de Manchester, en el Reino Unido la Open University, la Universidad Autónoma de Madrid, Universidad Politécnica de Madrid¹⁷ en donde se están llevando proyectos en esta área. Los resultados de estas investigaciones se refleja en un gran congreso internacional anual llamado el Internacional Semantic Web Conference¹⁸, revistas como el Journal of Web Semantics¹⁹, The Semantic Web de Electronic Transactions on Artificial Intelligence (ETAI)²⁰ y sobre todo hay que destacar el importante trabajo que está realizando la W3C²¹ en el liderazgo de grupos de trabajo de estandarización de lenguajes y tecnologías para la web semántica. En cuanto a estándares XML es el elemento de la web semántica al cual se le está dando mayor uso debido a que es un meta lenguaje de marcado que permite codificar páginas de un modo directo con etiquetas ad-hoc y por otro lado se pueden definir esquemas y tipos de documentos que pueden ser validados de forma automática por programas.

4. PASOS PARA MIGRACIÓN A WEB SEMÁNTICA

Según Lluís Codina y Cristòfol Rovira para realizar migración de sitios web sintácticos a sitios web semánticos se debería tener:

1. Calidad del código fuente: Es decir el código fuente de las páginas web debe ser consistente y sin errores.

2. Uso de metadatos: Las páginas web deben estar descritas mediante metadatos en suficiente cantidad y calidad.

3. Uso de RDF: La codificación de los metadatos se la puede hacer mediante RDF.

Referencias

- [1] Berners-Lee. Weaving the web.
- [2] Berners-Lee. Weaving the web: Glossary.
- [3] Rovira Cristòfol Codina Lluís. La web semántica. Technical report, E-prints in Library and Information Science.
- [4] T. R. Gruber. A translation approach to portable ontology specifications. Technical report, 1993.
- [5] Nigel Shadbolt, Tim Berners-Lee, and Wendy Hall. The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101, 2006.
- [6] James Hendler Tim Berners-Lee and Ora Lasila. The semantic web. Technical report, Scientific American, Mayo 2001.

¹⁷ <http://www.redwebsemantica.es/SemWeb/sewView/frames.jsp>

¹⁸ <http://iswc.semanticweb.org/>

¹⁹ http://www.elsevier.com/wps/find/journaldescription.cws_home/671322/description#description

²⁰ <http://www.etaj.org/seweb/>

²¹ <http://www.w3.org/2001/sw/>