

## De la web actual a la web semántica

**Marlon Alberto Piñeres Melo Mg.\*, Ilma Azucena Bonilla BotiaMg(C)\*\***

*\*Ingeniero de Sistemas de la Universidad Simón Bolívar con diplomado en Ingeniería del Software.  
Magíster en Ingeniería de Sistemas y Computación de la Universidad del Norte. Actualmente profesor de la Universidad Autónoma del Caribe.  
Barranquilla, Colombia. Email: mpineres@uac.edu.co*

*\*\*Ingeniero de Sistemas Universidad del Norte, Especialista en Informática y Telemática, Magister© en Informática  
Educativa de la Universidad Metropolitana de Chile. Coordinador de opciones de grado de Ing. Sistemas y Profesor de  
Universidad Autónoma del Caribe, Barranquilla, Colombia. Email: ibonilla@uac.edu.co*

### RESUMEN

El propósito de este artículo es presentar a la comunidad académica e industrial, un breve estado del arte de la Web Semántica, en qué consiste y sus beneficios con respecto a la Web actual, mostrando su proyección hasta el año 2030. Además explicar conceptos básicos, Ontologías Web, utilidades de las Ontologías Web, lenguajes de las Ontologías Web y trabajos que se han realizado con la Web Semántica.

**Palabras clave:** Internet, Metadatos, Web Sintáctica, Ontologías Web, Web Semántica, OWL, RDF.

### ABSTRACT

The purpose of this paper is to present to the academic community and industry, a brief state of the art of Semantic Web and what their profits on the web today, showing their projection until the year 2030. Besides explaining the basics, Web Ontologies, Ontologies Utilities Web, Web Ontology Language and the work that has been done with the Semantic Web.

**Keywords:** Internet, Metadata, Syntactic Web, Web Ontologies, Semantic Web, OWL, RDF.

## I. INTRODUCCIÓN

Actualmente el contenido que se muestra en Internet es inmensurable y el uso de tecnologías como el HTML es totalmente indispensable para la visualización de los contenidos en un browser. Pero ante esto cabe anotar que la utilización de este lenguaje trae como desventaja una mala representación semántica de los datos. Esto conduce a que las páginas Web actuales proporcionen sintaxis en vez de ser añadidas la semántica. La Web actual es, un extraordinario medio económico para el acceso al conocimiento explícito, servicios, entretenimiento, comercio y negocios electrónicos, entre

otros. Para esto las tecnologías que la hacen posible han tenido una gran evolución para garantizar la facilidad de su uso, hasta el punto de vincular bases de datos con estas aplicaciones y la creación de algoritmos para la recuperación de contenidos. Sin embargo, la cantidad de información que se pueda encontrar en Internet se ha convertido en algo incalculable, por lo cual hace que los motores de búsquedas se vean obligados a cambiar la forma en la que se recupera dicha información. Es por esta razón que la w3c ha propuesto nuevas tecnologías que facilitan la incorporación de semántica a las páginas Web como el Resource Description Framework (RDF) y su extensión como lo es

Resource Description Framework Schema (RDF-S) y Ontology Web Language (OWL) permitiendo a través de esto modelar el conocimiento presente.

## 2. MARCO TEORICO

### 2.1 LA WEB SINTÁCTICA

En la actualidad se puede decir que los ordenadores son utilizados para visualizar los contenidos de las páginas escritas en HTML, el cual es una tarea frecuente y fácil. Por otro lado a las personas les corresponde interpretar el contenido que se está visualizando, lo cual tiene un mayor grado de dificultad en lo que se pretende encontrar a partir de la búsqueda. Entonces ¿Será posible que a través los ordenadores se puedan realizar estas actividades?

Este ejemplo permite ver el uso de la Web Sintáctica: cuando se escribe en Google una cadena de caracteres para realizar una búsqueda sobre algo en particular por ejemplo, libros de Marlon Piñeres, se visualizan una lista de links que no tienen nada que ver con lo que realmente se quiere buscar. Se visualizan links con combinaciones de las palabras de búsqueda, haciendo engorrosa la selección de lo que se pretende encontrar.

Para mejorar la búsqueda aparece una forma de extender este vocabulario que es a través del lenguaje de marcado extendido y con tecnologías para añadir significado a las páginas; Web según el consorcio de la World Wide Web.

### 2.2 WEB SEMÁNTICA

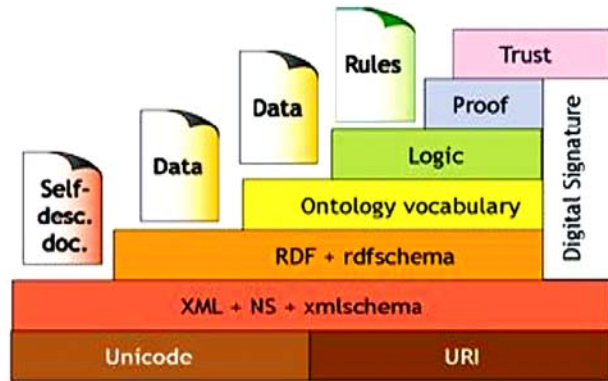
Teniendo en cuenta lo afirmado por la w3c la Web Semántica es:

- “Una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Al dotar a la Web de más significado y, por lo tanto, de más semántica, se pueden obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Esta Web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una Web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante” [1].
- “Disponer datos en la Web definidos y enlazados de forma que puedan ser utilizados por las máquinas no solamente para visualizarlos sino también para: automatizar tareas, integrar y

reutilizar datos entre aplicaciones (Tim Berners-Lee)” [2].

Con base en lo anterior se puede decir que ya existe una solución para agregar mayor significado a la Web. La figura siguiente es la arquitectura propuesta por la W3C para la Web Semántica [3]: ver figura 1.

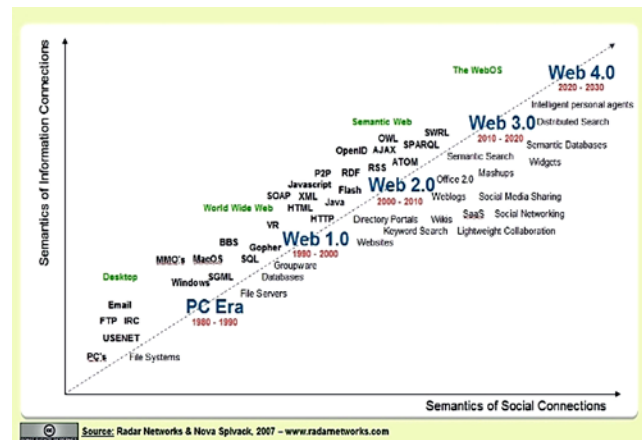
Figura 1. Arquitectura de la Web Semántica



### 2.3 ILUSTRACIÓN DEL ESTADO DEL ARTE DE LA WEB SEMÁNTICA

El siguiente gráfico muestre un breve estado del arte de la Web Semántica (ver figura 4).

Figura 2. La tendencia futura de la Web [16]



Con base a la figura anterior se puede decir que la Web Semántica es una realidad y para los próximos años tendrá una mayor utilización en el campo de los agentes personales inteligentes de búsquedas.

### 2.4 METADATOS

Para el desarrollo de este tema es menester precisar que los Metadatos son datos los que describen a otros datos para facilitar su catalogación. Los metadatos agregan semántica a los documentos lo cual facilitará a los motores de búsqueda la indexación de los datos y

lograr así, encontrar lo que estamos buscando. El Lenguaje de Marcado de Hipertexto, a pesar de que es el famosamente utilizado para la visualización de contenidos Web, carece de una semántica en el documento, es por eso que se hace necesario tener un lenguaje que facilite la descripción de la información de su contenido. Para esto se existen lenguajes estandarizados como XML que permite la descripción de contenidos a través de metadatos facilitando así la comprensión y el intercambio de información entre humanos y máquinas.

## 2.5 ONTOLOGÍAS WEB

Las ontologías provienen de la Inteligencia Artificial. Según Gruber [4] "una ontología es una especificación formal y explícita de una conceptualización compartida", esto se entiende mejor de la siguiente manera:

- **Formal:** se refiere al hecho de que la ontología debe ser legible por un ordenador, excluyendo el lenguaje natural.
- **Explícita:** significa que los conceptos que se utilizan y sus limitaciones se definen explícitamente.
- **Conceptualización:** se refiere a la identificación de los conceptos más relevantes de un fenómeno del mundo.
- **Compartida:** quiere decir que una ontología captura un consenso de conocimiento, es decir que el conocimiento no proviene de un solo individuo sino que es aceptada por un grupo.

Dicho de otro modo la información puede ser representada como una taxonomía de conceptos, relaciones y axiomas, permitiendo así describir y representar un área del conocimiento. Más específicamente, una ontología está formada por una taxonomía relacional de conceptos y por un conjunto de axiomas o reglas de inferencia mediante el cual se puede inferir conocimiento nuevo [5].

## 2.6 UTILIDAD DE LAS ONTOLOGÍAS WEB

Según el documento de los casos de uso y requisitos de la w3c [6] se han clasificado seis áreas principales:

**Portales Web:** un portal Web es un sitio que proporciona el contenido de un tema común facilitando así la creación de comunidades, encontrar enlaces a otros sitios de interés común, recibir noticias, entre otros. Sin embargo, para que un portal Web tenga éxito debe permitir la buena localización de contenidos, y esto se puede lograr definiendo una ontología para describir el contenido. Un ejemplo de esto es OntoWeb que sirve a la comunidad académica y a la industria que esté interesada en una ontología de investigación.

**Colecciones de Multimedia:** aquí se pueden usar las ontologías para hacer anotaciones semánticas para co-

lecciones de imágenes, audio, video u otros objetos no textuales, con la finalidad de que las máquinas puedan extraer significado acerca de los elementos multimedia.

**Administración de sitios Web Corporativos:** las ontologías pueden servir para indexar los documentos. En este caso, trabajar con múltiples ontologías aumentaría la capacidad de descripción.

**Diseño de Documentación:** aquí las ontologías pueden ser utilizadas para construir un modelo de información que permite la exploración del espacio de información en términos de los artículos que están representadas, las asociaciones entre los elementos, las propiedades de los temas, y los enlaces a la documentación que describe y define.

**Agentes Inteligentes:** la web semántica puede facilitar a los agentes inteligentes la capacidad de comprender y tomar decisiones basadas en primordialmente en las preferencias o intereses de los usuarios.

**Servicios Web y computación ubicua:** proporcionar la capacidad de entender otros dispositivos, y la razón acerca de sus servicios / funcionalidad es necesaria en el uso de estas arquitecturas, por motivos a que se convierte en una tarea inmanejable. Para esto una ontología utiliza el idioma de los periféricos para describir las características de los dispositivos, los medios de acceso a los dispositivos de este tipo, la política establecida por el propietario para el uso de un dispositivo, y otras limitaciones técnicas y los requisitos que afectan a la incorporación de un dispositivo en una computación ubicua.

## 2.7 LENGUAJES DE LAS ONTOLOGÍAS WEB

Entre los lenguajes de las ontologías Web analizados en el marco de este trabajo de grado están:

**Simple HTML Ontology Extensions (SHOE):** consistió en el primer lenguaje de etiquetas para diseñar las ontologías Web de University of Maryland [7]. Se originó antes de que se pensara en la Web Semántica. Este proyecto a pesar de que permitía definir clases y reglas de inferencias (de forma limitada) fue abandonado por el desarrollo de OIL y DAM [8]. Ver: <http://www.cs.umd.edu/projects/plus/SHOE/>

**Ontology Inference Layer (OIL):** este proyecto fue en parte derivado de SHOE, utiliza sintaxis de XML y está basada en RDFS. El problema de este consistía en su falta de expresividad para la declaración de axiomas [8]. Ver: <http://oil.semanticweb.org/>

**Darpa Agent Markup Language (DAML):** desarrollado por Defense Advanced Research Projects Agency

(DARPA) para la web semántica. Ver: <http://www.daml.org/>

DAML-OIL: se origina por la unificación de Darpa Agent Markup y Ontology Inference Layer, basándose en los estándares de la w3c para extender la expresividad de RDFS. Hereda muchos elementos de OIL, teniendo en cuenta que se fortalece la lógica de descripción. Ver: <http://www.w3.org/TR/daml-oil-reference>

OWL Web Ontology Language (OWL): nace del lenguaje DAML-OIL. Básicamente se trata de una recomendación de la w3c para representar ontologías Web de forma explícita y facilitar agregar significado de conceptos en vocabularios y las relaciones existentes entre aquellos términos. Ver: <http://www.w3.org/TR/owl-features/>

RDF: el Resource Description Framework o Marco de Descripción de Recursos, consiste en un lenguaje que ha sido diseñado para apoyar a la Web Semántica, similar a la forma en que HTML apoyó a las páginas Web. Pero este a diferencia del HTML, permite describir recursos o metadatos para la Web definiendo estructuras comunes para el intercambio de datos XML. El primer trabajo en borrador de RDF fue construido en 1997 por Ora Lassila y Ralph Swick [9].

XML TOPIC MAPS: es un consorcio independiente [10] que aprovecha la familia de especificaciones de XML, que ofrece un modelo para la representación de la gramática y la estructura de los recursos de información utilizados para definir temas (llamados topics) y las relaciones entre los temas.

Friend Of A Friend (FOAF): es un acrónimo de "amigo de un amigo" [11], este proyecto consiste en la especificación de un vocabulario que aporta a la Web Semántica, descripciones en las relaciones soportadas en XML, RDF, OWL, facilitando la extensibilidad de RDF. Si las personas publicaran información en el formato FOAF, las máquinas podrán tener acceso a esta información. Por ejemplo: si existe una imagen en la Web con una descripción de FOAF, el ordenador puede saber toda la información contenida en sus metadatos.

DOAP (Description Of A Project): se trata de la especificación de un vocabulario para describir proyectos de Software Libre en RDF desarrollado por Edd Dumbill [12].

DCMI (Dublin Core Metadata Initiative): es una organización encargada del desarrollo de estándares para metadatos interoperables, que facilitan la descripción de recursos. Actualmente este vocabulario tiene quince definiciones (título, claves, descripción, autor, fecha, entre otros.) que pretende proporcionar signi-

ficado semántico a cualquier recurso sin importar su formato [13].

SIOC (Semantically-Interlinked Online Communities o comunidades online semánticamente interconectadas): consiste en una organización, cuya finalidad es permitir la integración de comunidades en línea, ofreciendo una ontología Web para la representación de datos en redes sociales soportadas en RDF y OWL-DL, reutilizando el vocabulario de las ontologías como FOAF y Dublin Core [15].

### 3. RESULTADOS

En el siguiente ejemplo [17] se ilustran los resultados obtenidos con un buscador normal y un buscador semántico:

Figura 2. Resultados obtenidos con un buscador actual



Figura 3. Resultados utilizando la Web Semántica



#### 3.1 TRABAJOS REALIZADOS CON LA WEB SEMÁNTICA

Actualmente existen muchísimos trabajos relacionados con esta temática (Bosch, 2005), a continuación se mencionan unos ejemplos:

SWoogole2006: disponible en <http://swoogle.umbc.edu/> consiste en un proyecto para recuperación de documentos de la Web Semántica, lo que se denomina en SWoogole (SWDs, Semantic Web Documents) para archivos escritos en RDF y OWL. Su interfaz es parecida a Google.

Spock: se trata del primer motor de búsqueda para encontrar información de las personas en la Web como los sitios web, fotos, videos y blogs. Ver: <http://www.spock.com/>.

ZOMM: Motor de búsqueda semántica para la indexación y búsquedas en lenguaje natural, organizando los documentos en base a las informaciones que lo estructuran. Ver: <http://www.conocimiento-semantico.com/zoom.htm>.

AKT (Advanced Knowledge Technologies): El de consorcio Tecnologías Avanzada de conocimiento, consiste en la generación de tecnologías para ayudar a las organizaciones a crear, gestionar y extraer valor de los activos de conocimiento. URL: <http://www.aktors.org/akt/>.

DBin: es una aplicación de la Web Semántica que permite a usuarios crear grupos de discusión para expresar y recuperar conocimiento compartiendo archivos e información estructurada semánticamente usando RDF. URL: <http://www.dbin.org/>.

DIP (Data, Information, and Process Integration with Semantic Web Services): Su finalidad es desarrollar y extender la Web Semántica junto con las tecnologías de Web Services para que diferentes Servicios Web puedan cooperar entre sí automáticamente. URL: <http://dip.semanticweb.org/>.

KW (Knowledge Web): busca fortalecer los proveedores de servicios y la industria Europea, a través de las tecnologías de la Web Semántica, facilitando el e-work y e-commerce. URL: <http://knowledgeweb.semanticweb.org/>.

NeOn. Consiste en un proyecto que tiene por finalidad, manejar múltiples ontologías en red, que existen en un objeto particular. Está disponible en la URL: <http://www.neon-project.org/web-content/>.

Powerset: consiste en una tecnología que permite que los ordenadores puedan entender nuestro idioma, siendo la primera aplicación de procesamiento del lenguaje natural de búsqueda. Ver: <http://www.powerset.com/>.

#### 4. CONCLUSIONES

Los Lenguajes de Ontologías Web, nos permiten definir un vocabulario legible por las computadoras y ser utilizado para representar un área del conocimiento. La Web Semántica apoya la gestión del conocimiento; debido a que facilita manejo del mismo de tácito en explícito, hacer tareas que automatizan procesos, interoperabilidad entre ordenadores y sobre todo el compartir el conocimiento.

Gracias a la Web Semántica, se pueden encontrar respuestas a las preguntas realizadas en un browser de una manera más exacta, sencilla y rápida. Para esto es necesario analizar e integrar cada una de las capas de la llamada Web 3.0, para lograr que se puedan realizar búsquedas por significado, permitiendo así que la información pueda ser comprensible por computadoras y humanos.

La Web Semántica facilitará a las organizaciones tener la información disponible, ordenada, relacionada, indexada por medio de la utilización de metadatos, teniendo como fundamento una arquitectura basada en el lenguaje XML (Extensible Markup Language) y XML Shema como la base sintáctica de la futura Web, RDF (Resource Description Framework), RDF Schema para dotar de semántica a la Web y OWL (Ontology Web Language) para representar cualquier área del conocimiento.

#### 5. BIBLIOGRAFÍA

[1] Guía Breve de Web Semántica. (s.f.). Recuperado el 23 de Diciembre de 2008, de <http://www.w3c.es/Divulgacion/Guiasbreves/WebSemantica>

[2] Oviedo, D. d.-U. (s.f.). SlideShare de Tecnologías de XML y Web Semántica. Recuperado el 1 de Noviembre de 2008, de <http://www.slideshare.net/dragonx/tecnologias-xml-y-web-semantica>

[3] Berners-Lee, T. (s.f.). Architecture - Semantic Web - XML2000. Recuperado el 1 de Noviembre de 2008, de <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>

[4] RUDI STUDER V., R. B. (s.f.). Knowledge Engineering: Principles and Methods. Recuperado el 1 de Noviembre de 2008, de <http://www.das.ufsc.br/~gb/pg-ia/KnowledgeEngineering-PrinciplesAndMethods.pdf>

[5] Juan Federico Sequeda Sanclemente, D. P. (s.f.). Propuesta del Uso de Ontologías para la Búsqueda Semántica en Laboratorios de Investigación y Desarrollo: OLID. Recuperado el 1 de Noviembre de 2008, de <http://www.cs.utexas.edu/~jsequeda/pub/olid.pdf>

[6] OWL Web Ontology Language Use Cases and Requirements. (s.f.). Recuperado el 1 de Noviembre de 2008, de W3C Recommendation 10 February 2004: <http://www.w3.org/TR/2004/REC-webont-req-20040210/>

[7] Jeff Heflin, D. J. (s.f.). SHOE. Recuperado el 30 de Diciembre de 2008, de <http://www.cs.umd.edu/projects/plus/SHOE/>

- [8] La Web Semántica. (s.f.). Recuperado el 4 de Noviembre de 2008, de <http://www.lawebsemantica.com/contents/webSemantica/ontologias4.html>
- [9] O'Reilly. (2003). Practical RDF. O'Reilly & Associates.
- [10] XML Topic Maps (XTM) 1.0. (2001). Recuperado el Enero de 2009, de <http://www.topicmaps.org/xtm/index.html>
- [11] Dan Brickley, L. M. (2 de Noviembre de 2007). FOAF Vocabulary Specification 0.91. Obtenido de <http://xmlns.com/foaf/spec/>
- [12] Dumbill, E. (s.f.). DOAP - Description of a Project. Recuperado el Enero de 2009, de <http://usefulinc.com/doap/>
- [13] Dublin Core Metadata Initiative. (1995 - 2009). Recuperado el Enero de 2009, de <http://www.dublincore.org/>
- [14] John G. Breslin, U. B. (2004). Recuperado el Enero de 2009, de <http://sioc-project.org/>  
La Web Semántica. (s.f.). Recuperado el 4 de Noviembre de 2008, de <http://www.lawebsemantica.com/contents/webSemantica/ontologias4.html>
- [15] Bosch, M. (2005). La Web Semántica: Estado del arte. Recuperado el 3 de Octubre de 2008, de <http://www.webalice.it/melabosch/Contenidos/DocsPeriodio/BoschWebSemanticaEstadoActual.pdf>
- [16] How the WebOS Evolves. (s.f.). Recuperado el 9 de Enero de 2009, de [http://novaspivack.typepad.com/nova\\_spivacks\\_weblog/2007/02/steps\\_towards\\_a.html](http://novaspivack.typepad.com/nova_spivacks_weblog/2007/02/steps_towards_a.html)
- [17] Guía Breve de Web Semántica (2005). Recuperado el 19 de Mayo de 2009, de <http://www.w3c.es/Divulgacion/Guiasbreves/WebSemantica>