

LA WEB SEMÁNTICA

THE SEMANTIC WEB

María E. Chávez¹, Oscar Cárdenas¹, Oscar Benito¹

RESUMEN

Tenemos acceso a millones de recursos, independientemente de nuestra situación geográfica e idioma. Estos factores han propiciado el éxito de la web, pero también han originado problemas como: sobrecarga de información y heterogeneidad de fuentes de información. Si utilizamos un buscador, tan solo seremos capaces de recuperar conceptos aislados debido al algoritmo de emparejamiento que utilizan estos tipos de buscadores.

La Web Semántica propone facilitar la localización de los recursos a través de metadatos semánticos, que describirán así el contenido, el significado y la relación de los datos, permitiendo de esta manera la comunicación entre sistemas y programas, mejorando en la web la interoperatividad entre los sistemas informáticos y reduciendo la mediación de operadores humanos necesaria, delegando así las tareas en software. Gracias a la semántica en la web, el software es capaz de procesar su contenido, razonar con éste, combinarlo y realizar deducciones lógicas para resolver problemas cotidianos automáticamente. La Web Semántica va a suponer el futuro de la web, debido a estas ventajas que se distinguen claramente tras conocer su función.

En este trabajo se motivan y explican las propuestas de la Web Semántica, se describen las tecnologías más importantes desarrolladas para llevarlas a cabo, así como el punto en el que se encuentra este proyecto promovido por su propio creador.

Palabras clave: Metadatos semánticos, heterogeneidad, interoperatividad, Web Semántica.

ABSTRACT

We have access to million resources, independently of our geographic situation and language. These factors have caused the success of the Web, but also they have originated problems like: overload of information and heterogeneity of sources of intelligence. If we only used a finder we will be able to recover isolated concepts due to the emparejamiento algorithm that use these types of finders.

The Semantic Web it proposes to facilitate the location of the resources through semantic metadatos, that will describe therefore the content, the meaning and the relation of the data, and to allow to the communication between systems and programs, improving in the Web the interoperativity between the computer science systems and reducing the mediation of human operators necessary delegating therefore the tasks in software. Thanks to the semantics in the Web, software is able to process its content, to reason with this, to combine it and to make deductions logics to solve daily problems automatically. The semantic Web it is going to suppose the future of the Web, due to these advantages that are distinguished clearly after knowing their function.

In this work the proposals of the semantic Web are motivated and explained, describe the most important technologies developed to carry out them, as well as the point in which is east project promoted by its own creator.

Key words: Metadatos Semantic, heterogeneity, interoperatividad, Semantic Web.

¹ Universidad Nacional Mayor de San Marcos – UPG FISI.
E-mail: chavezbarces@yahoo.es, oeardenas@terra.com.pe, obenitop@gmail.com

1. INTRODUCCIÓN

La aparición de la WWW se puede situar en 1989, cuando Tim Berners-Lee presentó su proyecto de «*World Wide Web*» [Berners-Lee 1989] en el CERN (Suiza), con las características esenciales que perduran en nuestros días. Un año más tarde publicó el primer borrador de las especificaciones de HTML y HTTP; y en 1993 el lanzamiento de Mosaic, el primer navegador de dominio público, compatible con Unix, Windows, y Macintosh, por el *National Center for Supercomputing Applications* (NCSA), marca el momento en que la WWW se da a conocer al mundo, extendiéndose primero en universidades y laboratorios; y, en cuestión de meses, al público en general.

La *World Wide Web* se ha convertido en un instrumento de uso cotidiano en nuestra sociedad, es hoy un medio flexible y económico para la comunicación, el comercio y los negocios, ocio y entretenimiento, acceso a información y servicios, difusión de cultura, etc. Paralelamente al crecimiento espectacular de la web, las tecnologías que la hacen posible han experimentado una rápida evolución. La generación dinámica de páginas, el acoplamiento con bases de datos, la mayor interactividad con el usuario, la concepción de la web como plataforma universal para el despliegue de aplicaciones, la adaptación al usuario, son algunas de las tendencias evolutivas más marcadas de los últimos años.

En el año 2002, el mismo Tim Berners-Lee, inventor de la WWW, URIs, HTML y HTTP, lanza la Web Semántica como un complemento a su obra [Berners-Lee 2001]. Se trata de una corriente promovida por el propio creador y presidente del Consorcio W3C. Esta web estaría poblada por agentes o representantes de software capaces de navegar y realizar operaciones por nosotros para ahorrarnos trabajo y optimizar los resultados.

2. LA WEB SEMÁNTICA [1]

«*The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation*».

«La Web Semántica es una extensión de la web actual en la cual la información recibe un significado bien definido, permitiendo a los computadores y las personas trabajar en cooperación de mejor forma».

«*The Semantic Web is the representation of data on the World Wide Web. It is a collaborative effort*

led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming.» [2]

«Disponer de 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».

La Web Semántica es una web extendida, dotada de significado bien definido, que persigue ser más inteligente. Está basada fundamentalmente en el marcado semántico y descriptivo no sólo de los documentos, sino también de los datos. Funciona mediante la utilización de una infraestructura común denominados *metadatos semánticos* y se apoya en colecciones de enunciados llamadas *ontologías*, que definen las relaciones entre conceptos y especifican reglas lógicas para que las máquinas multipliquen su capacidad de procesar y comprender los datos. Nos encontramos frente a una potente herramienta facilitadora de la comunicación.

La Web Semántica mantiene los principios de descentralización, compartición, compatibilidad, la máxima facilidad de acceso y contribución al crecimiento. En este contexto, un problema clave es alcanzar un entendimiento entre las partes que han de intervenir en la construcción y explotación de la web: usuarios, desarrolladores y programas de muy diverso perfil.

La tecnología que se ha creado para hacer posible la web semántica incluye lenguajes para la representación de ontologías, parsers, lenguajes de consulta, entornos de desarrollo, módulos de gestión (almacenamiento, acceso, actualización) de ontologías, módulos de visualización, conversión de ontologías, y otras herramientas y librerías.

Hay un equipo de gente en el World Wide Web Consortium (W3C) dedicada a mejorar, extender y estandarizar el sistema y muchos lenguajes; publicaciones y herramientas han sido ya desarrolladas. A pesar de ello, las tecnologías de la Web Semántica aun están en un estado muy temprano, y aunque el futuro del proyecto en general parece ser brillante, parece haber poco consenso sobre la dirección que debe tomar.

Generalmente los datos ocultos en archivos HTML, suelen ser útiles en algunos contextos, pero no en

otros. El problema con la mayoría de los datos en la Web de esta forma es que es difícil de usar en una larga escala, porque no hay un sistema global para publicar datos de tal forma que pueda ser fácilmente procesada por cualquiera.

3. PUNTOS FUNDAMENTALES

La Web Semántica se basa en dos puntos fundamentales:

- a) **La descripción del significado.** Donde definimos los conceptos; y
- b) **La manipulación automática de estas descripciones.** Se efectúa mediante lógica y motores de inferencia.

En la descripción del significado se articula la **semántica**, los **metadatos** y las **ontologías**.

3.1. La semántica. Es el estudio del significado de los términos lingüísticos.

En nuestro contexto de la Web Semántica, se pretende dotar de significado interpretable por parte de las máquinas, como información adicional que pueda ser comprendida y procesada por una computadora.

3.2. Los Metadatos. Son datos que describen otros datos, en este contexto los datos que describen recursos de la web.

La distinción entre datos y metadatos es relativa, pues depende de la aplicación. Los metadatos de una aplicación pueden ser los datos que maneja otra aplicación.

3.3. Las Ontologías [3]

Una ontología es una jerarquía de conceptos con atributos y relaciones, que define una terminología consensuada para definir redes semánticas de unidades de información interrelacionadas y con ciertas reglas.

Una ontología se puede definir como una «*especificación explícita y formal de una conceptualización compartida*» según [Gruber, 1993] de esta manera:

- Es *explícita*, porque define los conceptos, propiedades, relaciones, funciones, axiomas y restricciones que la componen.
- Es *formal*, porque es legible e interpretable por las máquinas.
- Es una *conceptualización*, porque es un modelo abstracto y una vista simplificada de las entidades que representa.

- Finalmente, es *compartida* porque ha habido un consenso previo sobre la información que ha sido acordado por un grupo de expertos.

Las ontologías capturan conocimiento consensuado de un modo genérico y formal, de tal manera que pueda ser compartido y reutilizado por distintos grupos de personas y aplicaciones de software.

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

- **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 la clase C, entonces A no es subclase de B*», «*Para todo A que cumpla la condición C1, A es B*», etc.

Los axiomas permiten, junto con la herencia de conceptos, inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos, con estas propiedades de relaciones y herencia de los conceptos especificados se tendría mayor capacidad expresiva del dominio almacenado.

Una ontología proporciona un vocabulario de clases y relaciones para describir un dominio, cuyo objetivo es que sean comprensibles y compartidos para diferentes aplicaciones y permitan la interoperabilidad. Por ejemplo, una ontología sobre arte podría incluir clases como: *Pintor*, *Cuadro*, *Estilo* o *Museo*; y relaciones como: *autor* de un cuadro, pintores *pertenecientes* a un estilo artístico u obras *localizadas* en un museo.

Las ontologías se empiezan a utilizar a finales de los 80 en el campo de la inteligencia artificial, como medio para la compartición y reuso de conocimiento. En la segunda mitad de los 90 se empiezan a aplicar a la

web para la inclusión de descripciones semánticas explícitas de recursos, facilitando descripciones, búsqueda semántica y razonamiento.

Podemos representar con ontologías en la web: productos, catálogos, recursos humanos, material educativo, servicios, plataforma, dispositivos y perfil de usuarios.

Su áreas de aplicación se dan en la ingeniería del conocimiento, gestión del conocimiento, procesamiento del lenguaje natural, sistemas de información cooperativos, recuperación de información, comercio electrónico, etc.

En la Figura N.º 1 se muestra las ontologías en la Web Semántica para la descripción del significado.

Ontologías en la web semántica

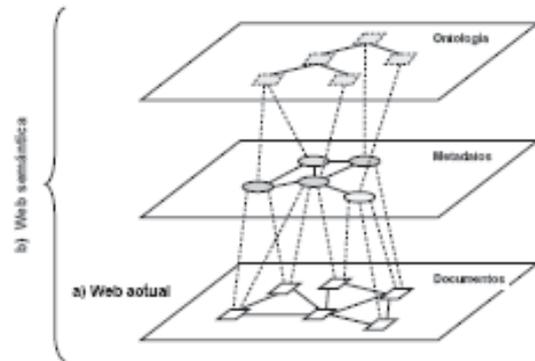


Figura N.º 1: Ontologías de la Web Semántica.

3.4. Editores de Metadatos

La creación de ontologías no es una tarea trivial, para facilitarla aparecen los editores de metadatos. Los más importantes son: Reggie, que es muy sencillo; OntoWeb, más completo; y Protégé.

Protégé

Creado en la Universidad de Stanford, basado en un modelo de conocimiento compatible con OKBC (*Open Knowledge Base Connectivity*). Es mucho más desarrollado, muy completo, escrito en Java, y con código fuente disponible. Permite editar ontologías: texto ASCII, base de datos JDBC, esquema RDF, crear herramientas de adquisición de conocimiento mediante formularios relacionados con las ontologías descritas. Creación de base de datos de conocimiento mediante la entrada de instancias particulares de los datos de la ontología. Ejecución de aplicaciones que operen sobre la base de conocimiento.

La Figura N.º 2 muestra al editor de ontologías Protégé.

Su interface es bastante sencilla y comprensible. Se basa en clases y sus dependencias jerárquicas, slots, que es la definición de cada clase, formularios e instancias entrada de datos concretos en las fichas anteriores.

4. REPRESENTACIÓN DEL CONOCIMIENTO

Los metadatos y las ontologías forman parte del campo de la representación del conocimiento. Es un campo muy prometedor, pero que todavía no ha conseguido desplegarse ampliamente. Para representar el conocimiento contenido en las bases de datos se necesita: La definición de la semántica, metadatos y ontologías, al igual que un conjunto de reglas lógicas y motores de inferencia [4].

Los **motores de inferencia** concretizan el conocimiento abstracto que posee el sistema para obtener las conclusiones (base de conocimiento) y tomar las decisiones correspondientes.

Si el motor de inferencia no puede obtener una conclusión al interpretar las reglas, el sistema podrá solici-

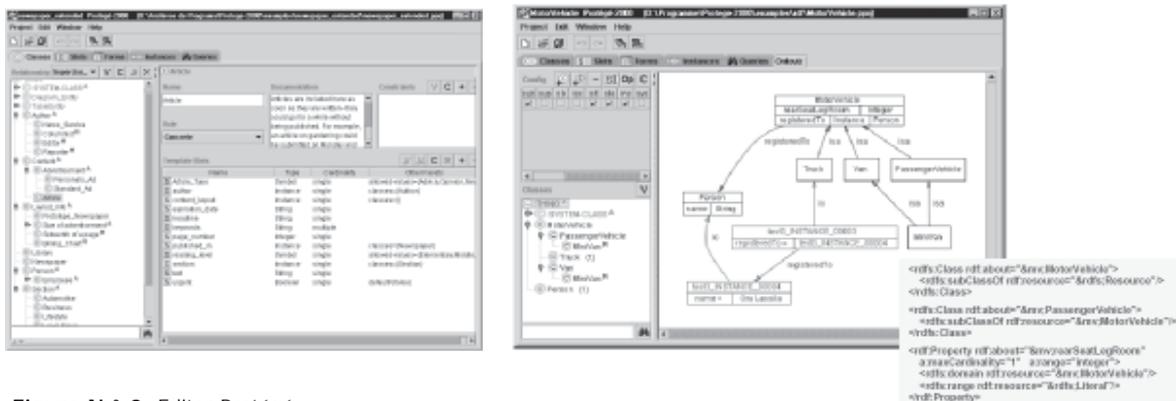


Figura N.º 2: Editor Protégé.

tar, a través de los módulos de adquisición, más conocimiento o más hechos, respectivamente, a fin de obtener alguna conclusión válida.

La base de conocimiento contiene el conocimiento general sobre el dominio en el que se trabaja. El método más común para representar el conocimiento es mediante las reglas de deducción.

4.1. Lenguajes de Representación

Para describir la semántica se requiere de un lenguaje apropiado llamado lenguaje de representación, estos lenguajes tienden a estar basados en XML.

El primer lenguaje para la construcción de la Web Semántica fue SHOE, creado por Jim Hendler en la Universidad de Maryland en 1997. Desde entonces, se han definido otros lenguajes y estándares con finalidad similar, como XML, RDF, DAML+OIL, y más recientemente OWL, por citar los más importantes.

XML (eXtensive Markup Language)

Estándar emergente para el intercambio de datos en la web.

XML no añade semántica al HTML (desde un punto de vista computacional), ello implica metadatos y ontologías.

XML permite estructurar datos y documentos en forma de árboles de etiquetas con atributos. Con XML Schema (XMLS) se pueden acordar de antemano las estructuras que se van a utilizar, así como manejar tipos de datos primitivos y derivados. Desde la aparición de XML en 1998, se han definido multitud de estándares para modelar información en dominios específicos como las finanzas, el periodismo, la enseñanza, o la medicina, entre otros muchos campos. XML es un primer paso en la dirección de avanzar hacia una representación explícita de los datos y la estructura de los contenidos de la web, separada de su presentación en HTML. XML proporciona una sintaxis para hacerlo posible, pero ofrece una capacidad limitada para expresar la semántica. El modelo de datos XML consiste en un árbol que no distingue entre objetos y relaciones, ni tiene noción de jerarquía de clases.

XML es un subconjunto de SGML (*Standard Generalized Markup Language*), y define un formato de texto diseñado para la transmisión de datos estructurados. Al ser un subconjunto de SGML mantiene sus características de validación, estructurado y especialmente facilita la extensibilidad, porque es un metalenguaje que permite describir lenguajes de

marcas, tanto la definición de etiquetas como la relación estructural que existen entre ellas.

SGML es una norma que pretende establecer una manera genérica de especificar, definir documentos, la cual permitiese a su vez usar formatos de mayor flexibilidad y portabilidad. Con lo cual reunía tres condiciones básicas:

1. *Formal*. Pues permite establecer la validez de los documentos.
2. *Estructurado*. Para que fuese capaz de manejar documentos complejos.
3. *Ampliable*. Para facilitar la gestión de grandes depósitos de información.

En la Figura N.º 3 se muestra un ejemplo de cómo funciona XML.

```
<agenda>
  <persona>
    <nombre>Kike</nombre>
    <telefono>638002993</telefono>
    <comentario>Es un bombon</comentario>
  </persona>
  <persona>
    <nombre>Maria</nombre>
    <telefono>956-78.90.12</telefono>
    <telefono>652135792</telefono>
  </persona>
</agenda>
```

Figura 3. Lenguaje XML.

XML Schema es un lenguaje para definir la estructura de los documentos XML.

RDF

Las ontologías son colecciones de enunciados redactados en un lenguaje, como el RDF, que define las relaciones entre conceptos y especifica reglas lógicas para razonar con ellos. Los ordenadores «comprenderán» el significado de los datos semánticos de una página siguiendo vínculos con ontologías especificadas. XML no es suficiente, sólo modela documentos, y el mundo real no es un documento, sino una red de relaciones.

Los RDF (*Resource Description Framework – Mecanismo para Describir Recursos*) permiten la descripción y el procesamiento de los metadatos, tiene la capacidad de describir metadatos de cualquier dominio. La información es un grafo dirigido etiquetado que modela las relaciones entre objetos.

Su sintaxis y estructura es similar al de los lenguajes orientados a objetos: clases y subclases, éstas se disponen en una jerarquía, y las subclases pueden heredar propiedades de las clases; es posible la herencia múltiple, ésta permite la mezcla de diferentes esquemas semánticos.

RDF proporciona también reglas para facilitar técnicamente la manera de explicar conceptos de modo que los ordenadores puedan procesarlo rápidamente y proporciona un medio que posibilita la edición de vocabularios con propiedades definidas para la descripción de los recursos de una comunidad.

En el **modelo** se describe parejas de nombre y valores de una propiedad determinada, con tres tipos de objetos:

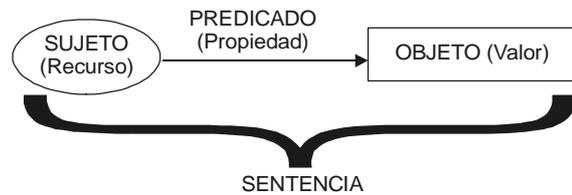
Recurso. Es cualquier objeto: una página web, bases de datos, dirección de la empresa, etc. Un recurso tiene un único identificador. Éstas, sean de la naturaleza que sean, se describen con un URI (*Universal Resource Identifier*) abstracto.

- *URL Localizador*, una dirección concreta.
- *URN Nombre*, permite cambiar la dirección y aún así encontrarla.

Propiedad. Es la característica, atributo, relación o aspecto que describe un recurso. Cada propiedad tiene: significado, define sus valores posibles, define los tipos de recursos a los que es aplicable y también define la relación con otras propiedades.

Sentencia. Agrupa un recurso determinado y una propiedad con un nombre y valor asociado al recurso. La sentencia tiene los siguientes elementos:

- Sujeto: recurso
- Predicado: nombre de la propiedad
- Objeto: el valor de la propiedad



RDF Sintaxis básica

El modelo define el marco abstracto y se requiere una sintaxis concreta. RDF usa XML como lenguaje base, también XML *namespace facility* para asociar cada propiedad con el esquema que la define. La estructura de clases aparece en el esquema.

Ejemplo: Dürsteler es el creador de la web:

<http://www.infovis.net>

Sujeto: <http://www.infovis.net>

Predicado: creado (propiedad, tiene un creador)

Objeto: Dürsteler

```

<Painter about=»vangogh» name=»Vincent van Gogh»
    birth=»1853" death=»1890" nationality=»Dutch»>
    ...
</Painter>

<Painting about=»starrynight» ...>
<author resource=»vangogh»>
    ...
</Painting>                                Instancias

<Class about=»Painter»>                    Clases
<subclassOf resource=»Artist»/>
</class>

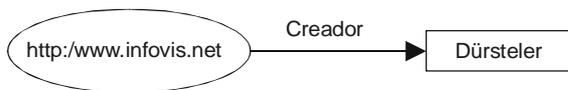
<Class about=»Painting»>
<subclassOf resource=»Artwork»/>
</Class>

<Class about=»Artist»/>
<Class about=»Artwork»/>

<Property about=»author»>
    <domain resource=»Artwork»/>
    <range resource=»Artist»/>
</Property>
    
```

Figura N.º 4: Sintaxis de RDF(s).

Visto en forma de grafo:



RDF Schema

Un esquema es un conjunto de clases que definen un dominio o aplicación.

El esquema se encarga de proporcionar el mecanismo para declarar las propiedades y definir las relaciones entre propiedades y recursos, jerarquía de clases.

RDF Schema es un vocabulario para describir las propiedades y las clases de los recursos RDF, con una semántica para establecer jerarquías de generalización entre dichas propiedades y clases.

OWL (Web Ontology Language). Lenguaje de Ontologías Web

OWL añade más vocabulario para describir propiedades y clases, tales como: **relaciones entre clases** (por ejemplo, disyunción), **cardinalidad** (por ejemplo, «únicamente uno»), **igualdad**, **tipologías de propiedades** más complejas, **caracterización de propiedades** (por ejemplo, simetría) o clases enumeradas. OWL es un mecanismo para desarrollar temas o vocabularios específicos en los que asociar estos recursos. Proporciona un lenguaje para definir ontologías estructuradas que pueden ser usadas a través de diferentes sistemas. La ontología se encarga de definir los términos utilizados para describir y representar un área de conocimiento. Las ontologías son utilizadas por usuarios, bases de datos y aplicaciones que necesitan compartir información específica, de un determinado campo; por ejemplo, de finanzas, medicina, deporte, etc. Las ontologías incluyen definiciones de conceptos básicos en un campo determinado y la relación entre ellos.

4.2. Definir un Conjunto de Reglas Lógicas y los Motores de Inferencia

Es la forma de extraer nuevos conocimientos a partir de los existentes. El motor de inferencia extrae conclusiones a partir de un conjunto de reglas o un conjunto de premisas o axiomas.

Aproximaciones:

- Motores de inferencia basados en diferentes órdenes de lógica.
- Métodos de resolución de problemas.

En la Figura N.º 5 se muestra la estructura en capas de la Web Semántica.

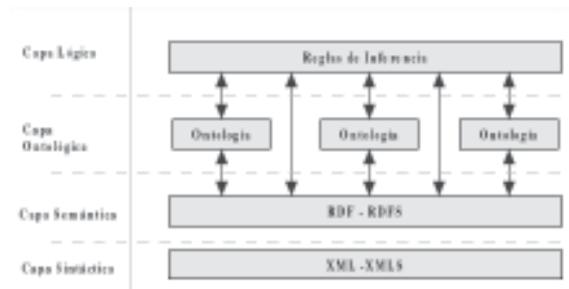


Figura N.º 5: Capas para la representación del conocimiento.

Semántica de contenidos: niveles de representación

Reglas y lógica	<i>Razonadores</i>
Objetos interrelacionados	<i>Ontologías</i>
Metadatos estructurados	<i>Clases de metadatos</i>
Metadatos (pares atributo/valor)	<i>Data Elements</i>
Palabras clave	<i>Thesaurors, diccionarios</i>

Web del Conocimiento

Cuyo sistema de razonamiento es el conjunto de axiomas + las reglas de inferencias.

Ejemplo:

Razonamiento:	Juan Pérez es pediatra, por lo tanto atiende a niños.
Prueba:	Credenciales de Juan Pérez + certificado de actualización + definición de pediatra
Validación:	Determinar si Juan Pérez es pediatra, dadas sus credenciales del Colegio Médico + certificado + etc.

Propiedades

- Expresividad: Capacidad de representar un problema.
- Consistencia: Todo lo que se deduce es correcto.
- Completabilidad: Todo lo que es correcto puede deducirse.
- Decidibilidad: Existe un algoritmo para decidir si se cumple una conclusión.
- Tratabilidad: El algoritmo de decisión tiene una complejidad.

5. ARQUITECTURA DE LA WEB SEMÁNTICA [5]

En la Figura N.º 6 se muestra una posible arquitectura de la Web Semántica, constituida generalmente por:

Unicode. Es una codificación de textos que permite utilizar los símbolos de diferentes idiomas sin observar caracteres extraños. Esto permite expresar información en la Web Semántica en cualquier idioma.

URI. Es el acrónimo de *Uniform Resource Identifier* (Identificador Uniforme de Recursos). Es un poco confusa la diferencia entre URI y URL, por lo que vamos a decir que un URL es un recurso que puede ser accedido vía Internet.

XML + NS + XML Schema. Tal vez la capa más técnica de la Web Semántica. En esta capa se agrupan las diferentes tecnologías que hacen posible que los agentes puedan entenderse entre ellos.

RDF + RDF Schema. Directamente basada y apoyada en la capa anterior; esta capa define el lenguaje universal con el cual podemos expresar diferentes ideas en la Web Semántica.

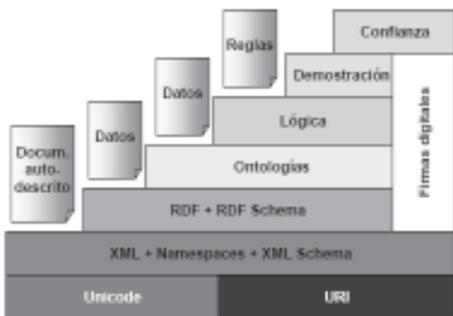
LENGUAJES DE ONTOLOGÍAS. Los lenguajes de ontologías nos permiten extender la funcionalidad de la Web Semántica, agregando nuevas clases y propiedades para describir los recursos. [6]

Lógica
Prueba
Confianza
Firma Digital

Las firmas digitales son simplemente pequeños bits de código que uno puede usar para verificar, sin lugar a dudas, que alguien escribió cierto documento. Muchas personas están familiarizadas con la tecnología: Es el mismo sistema basado en llaves PGP, los cuales usan mensajes encriptados. Usamos simplemente estas tecnologías a RDF.

PICS

Los PICS (*Platform for the Internet Content Selection*) nos indican lo adecuado o conveniente de determi-



Fuente: Tim Berners-Lee. Semantic Web-XML 2000. Arquitectura.

Figura N.º 6: Componentes de la Arquitectura de la Web Semántica.

nados ficheros de datos, según la comunidad en la que se encuentre el usuario. Es una infraestructura para asociar las etiquetas con los contenidos de Internet. Aunque en un principio estaba destinado al control del acceso de los niños a Internet, su uso se puede extender a otras etiquetas que incluyan privacidad, licencias, etc. PICS es una plataforma sobre la cual se han construido otros servicios de clasificación que no sólo define una manera de construir etiquetas, sino que es un mecanismo para realizar las valoraciones. Este mecanismo está formado, al menos, por lo siguiente:

- Las *etiquetas*, que son los metadatos que indican la valoración de un documento.
- *Los servicios de valoración*, es decir, las organizaciones, grupos o personas que realizan una valoración.
- Los *perfiles*, que son las reglas que da el usuario para definir el filtro para evitar recibir documentos no deseados.

Para que el filtrado de documentos no deseados se lleve a cabo; también es necesario un software cliente y otro servidor que tengan implementados el sistema de valoración. Estas funciones se pueden realizar por separado, lo cual permite que por un lado los desarrolladores de software puedan realizar una aplicación informática sin suministrar un sistema de valoración; mientras que, por otro, una organización puede crear sistemas de valoración sin tener que desarrollar el software.

6. EL FUTURO DE LA WEB [7]

Con bastante certeza el futuro de la Web es la WEB SEMÁNTICA; sin embargo, algunos autores van más allá y hacen predicciones; según esto, todo va hacia una red global del conocimiento. Francis Heylighen, Cliff Joslyn y Johan Bollen hablan de la futura web



Fuente: La Web inteligente. II Jornada de en red ando. Barcelona, 26-10-2001.

Figura N.º 7.

como algo que piensa y aprende, como un superorganismo social que tiene un cerebro global.

A pesar de este tecno-optimismo, también existen numerosos críticos que aducen que este tipo de planteamientos son no sólo descabellados; sino, incluso, temibles. La realidad es que todavía estamos muy lejos de conseguir, incluso, algunos de los planteamientos menos complejos de la Web Semántica.

Pero la realidad es que, actualmente, la construcción y recuperación de estructuras semánticas no se puede llevar a cabo de forma automática. En los albores de la web nadie pensó en etiquetar categorías y añadir metadatos, pues la mayor parte del contenido de la web estaba diseñada para leer, no para que fuera manipulada por ordenadores, robots y agentes. Pronto se vio que, ante las colosales dimensiones de la web y la heterogénea información que contenía, era necesario no sólo organizar la información, sino también clasificarla y categorizarla con el fin de poder realizar búsquedas y recuperar la información de forma automática. Y esto sólo es posible mediante una indización y clasificación previas, esto es, con la generación de documentos previamente estructurados formal y semánticamente, y utilizando un lenguaje que sea independiente de la plataforma o aplicación empleadas.

La Web Semántica cuenta ya con una nutrida infraestructura de lenguajes y tecnologías para hacerla posible. La sintaxis se basa en el lenguaje XML y derivados, y la semántica en los lenguajes RDF (S) y OWL, y también están presentes otras muchas aplicaciones y tecnologías ya desarrolladas como URIs, Topic Maps, XFML, firmas digitales, etc. y numerosas empresas y centros de investigación están trabajando en ella. Sin embargo, para que la antigua tarea humana e intelectual de la identificación, indización y clasificación de documentos pueda hacerse mediante ordenadores sólo podrá llevarse a la práctica si colaboran de forma interdisciplinaria documentalistas, lingüistas, informáticos, arquitectos de la información, ingenieros y expertos en inteligencia artificial, etc.

7. APLICACIONES Y PROYECTOS [8]

Un gran número de aplicaciones de la Web Semántica pueden ser usadas para varias tareas, incrementando la modularidad de las aplicaciones en la web. A pesar de ello, las tecnologías de la Web Semántica aun están en estado muy temprano, y aunque el futuro del proyecto en general parece ser brillante, parece haber poco consenso sobre la dirección que debe tomar y las características de los co-

mienzos de la Web Semántica. Hay un equipo de gente en el World Wide Web Consortium dedicada a mejorar, extender y estandarizar el sistema, y han sido ya desarrollados muchos lenguajes, publicaciones y herramientas.

La Web Semántica es el futuro de la web, dentro de pocos años su uso será tan común como lo es ahora la web. Para lograr esto muchas instituciones tienen proyectos en marcha, algunos de ellos ya están en funcionamiento. Aquí una lista de ellos:

- **Haystack.** Personaliza la navegación según los intereses y gustos del usuario. Entre las funciones de este navegador podemos destacar las siguientes: importa archivos RDF/XML de la web o archivos del sistema, navega por las páginas web y los recursos semánticos, crea colecciones de recursos de la Web Semántica y permite navegar por ellos, lee weblogs basados en RSS, permite el etiquetado de cualquier objeto, maneja colecciones de fotos digitales y organiza archivos de música, monta la información en mini-portales, desarrolla ontologías a medida, etc.
- **SWoogle.** (Semantic Web Search) Fue desarrollado por la Universidad de Maryland (Baltimore, USA). Se trata de un buscador que busca ontologías, documentos y términos escritos, tanto en RDF como OWL, esto es, busca documentos de la Web Semántica o *Semantic Web Documents* (SWDs). Curiosamente, tanto su propio nombre como su interfaz de presentación son muy similares a los de Google.
- **CORESE.** Un proyecto del INRIA (Institut National de Recherche en Informatique et en Automatique) que se enfoca a visualizar la Web Semántica, puesto que se basa en gráficos conceptuales.
- **LOOM.** Es un lenguaje para construir aplicaciones inteligentes y que se basa en un sistema de representación que se utiliza para ofrecer un soporte deductivo y que permite hacer definiciones y establecer reglas, clasificar y consultar conocimiento, o CLIPS –*C Language Integrated Production System*– que ofrece una herramienta para gestionar una extensa variedad de conocimiento con el soporte de tres paradigmas de programación diferentes: basado en reglas, orientado a objetos y procesal.
- **WSML (Web Service Modeling Language)** Lenguaje de Modelado de Servicios Web basado en *WSMO (Web Service Modelling Ontology)*. WSML consta de una variedad de lenguajes llamados: *WSML-Core*, *WSML-DL*, *WSML-Flight*.

Los proyectos más importantes en el campo de la Web Semántica son los siguientes:

- **AKT.** (*Advanced Knowledge Technologies*) El consorcio AKT agrupa 5 universidades del Reino Unido y fue fundado por el Engineering and Physical Sciences Research Council (EPSRC). Su fin es ayudar a desarrollar la próxima generación de tecnologías del conocimiento para dar soporte a la gestión del conocimiento de las organizaciones. AKT pretende desarrollar y extender métodos integrados y servicios para capturar, modelar, publicar, reutilizar, compartir y gestionar el conocimiento. Para ello se tienen en cuenta los recientes desarrollos en inteligencia artificial, psicología, lingüística, multimedia y tecnologías de Internet.
- **ASG.** (*Adaptive Services Grid*) Es un proyecto integrado dentro del 6º Programa Marco de la Comisión Europea. El proyecto que comenzó en septiembre de 2004 y dura 2 años, agrupa 22 participantes de 7 países. El objetivo es desarrollar un prototipo de plataforma abierta para la innovación, creación, composición y lanzamiento de servicios. ASG cuenta con las principales organizaciones de producción científica y tecnológica que hacen uso del conocimiento y con las instituciones europeas líderes en la investigación y desarrollo del software, las telecomunicaciones y la industria telemática.
- **DBin.** Es una nueva clase de aplicación web: un P2P Web Semántico con una filosofía para «compartir archivos» y/o un Cliente de Grupos de noticias semántico. Similar a un cliente para compartir archivos, DBin conecta directamente a otros peers. *Instead of files, sin embargo, éste bajará «información relevante»* sobre los temas que se especifiquen. De forma más técnica, permite compartir y recibir «información estructurada semánticamente» usando RDF y otros componentes de la Web Semántica.
- **DIP.** (*Data, Information, and Process Integration with Semantic Web Services*) El objetivo de DIP es desarrollar y extender la Web Semántica y las tecnologías de los Servicios Web para producir una nueva infraestructura tecnológica para los Servicios de la Web Semántica.
- **ELeGI.** (*The European Learning Grid Infrastructure*) Una red semántica para el aprendizaje humano para la puesta en marcha de escenarios futuros de aprendizaje basado en la ubicuidad y la colaboración, y centrados en la experiencia y el aprendizaje contextualizado a través del diseño, implementación y validación del aprendizaje en red.
- **Esperanto Project.** Es un proyecto que fue desarrollado entre el 2002 y el 2005 dentro del 5º Programa Marco de la Comisión Europea, y cuyo objetivo era hacer de puente entre la Web actual y la Web Semántica. La descripción del proyecto y los resultados se pueden encontrar en:
- **ESWS.** (*European Semantic Web Symposium*) El primero de estos simposios se celebró en Creta en el año 2004. Estos encuentros tienen como fin mostrar las tecnologías, desarrollos y aplicaciones de la Web Semántica a nivel europeo y mundial.
- **KW.** (*Knowledge Web*) Es una Red de Excelencia FP6 que ayuda a dar soporte de transición a las tecnologías de ontologías desde el sector académico a la industria. El consorcio actual está integrado por 18 participantes que incluyen líderes en Web Semántica, multimedia, tecnologías del lenguaje humano, agentes, etc.
- **IMS Global.** se trata de un consorcio en el que participan más de 50 organizaciones y empresas, que tiene como objetivo el aprendizaje global a través de la web. En este marco, se trabaja con esquemas XML y documentación estructurada en donde RDF juega un papel fundamental; por ejemplo, en la definición de vocabularios y taxonomías.
- **NeOn.** Proyecto fundado por la Comisión Europea dentro del 6º Programa Marco. Coordinado por la Open University. En el proyecto también participan instituciones líderes de Europa en el área del modelado del conocimiento y las ontologías. El fin de NeOn es crear el primer servicio orientado de infraestructura abierta y metodología asociada para soportar el desarrollo de aplicaciones de la Web Semántica, con el objetivo de extender el estado de la cuestión a soluciones viables económicamente. Estas aplicaciones se realizarán sobre una red de ontologías contextualizadas, mostradas localmente que no necesariamente tengan una consistencia global.
- **OpenKnowledge.** Tiene por objetivo crear una nueva forma de abrir y compartir arquitecturas de conocimiento de forma coordinada poniendo el foco de atención en la semántica para la interacción y uso en las comunidades web. «Open» en *OpenKnowledge* se emplea en 2 sentidos: Como un sistema abierto, en el cual cualquiera puede utilizarlo en cualquier momento; y en el sentido de abierto, para ser ensamblado, logrado por medio de la participación con un coste individual bajo.
- **ESSI.** (*European Semantic System Initiative*) Es la suma de dos proyectos SDK Project Cluster y

ASG: El SDK (SEKT, DIP, *Knowledge Web*) Cluster se ha unido estratégicamente con ASG (*Adaptive Services Grid*). Este nuevo ESSI Cluster combina los Servicios de la Web Semántica y las soluciones basados en sistemas semánticamente potentes con arquitecturas orientadas a servicios semánticos.

- **SEKT.** Project (*Semantically -Enabled Knowledge Technologies*) Co-fundado por el 6º Programa Marco de la UE para llevar a cabo la Sociedad del Conocimiento. La visión de SEKT es desarrollar y explotar las tecnologías del conocimiento bajo unos nuevos parámetros de gestión donde se rompan las fronteras entre gestión de documentos, gestión del contenido y gestión del conocimiento, y donde la gestión del conocimiento se convierta en una actividad cotidiana sin gran esfuerzo.

Otros:

- Directorios y Catálogos Web
 - Open Directory Project (www.dmoz.org), catálogo de editores distribuidos.
 - Openguides.org, guía de editores distribuidos.
 - Dublín Core (dublincore.org)
 - Vocabulario liviano para representar metadatos sobre recursos: <http://purl.org/dc/elements/1.1/>
 - Propuesto desde la comunidad de bibliotecología.
 - Title, Contributor, Author/Creador, Publisher, Subject, Description, Date, Resource, Format, Resource Identifier, Language, Source, Coverage, Rights.
- a) Redes Sociales
- ¿Cómo encontrar:
 - personas con intereses similares a los de uno?
 - personas relacionadas a alguien?
 - autoridades en un tópico?
 - etc.
 - Redes sociales:
 - Información personal en la Web semántica.
 - Enlaces a información personal de personas conocidas.
- b) Búsqueda Semántica [9]
- Pediatras que vivan a menos de 10 cuadras de mi casa.
 - Hoteles 3 estrellas en Viña del Mar que estén cerca del casino.
 - Carreras del área biología en Universidad de Cauca.

- Nombre y dirección postal de profesores de escuelas primarias de Bogotá.

c) Consultas Avanzadas

- ¿Cuál es la relación entre dos personas?
 - Asociaciones semánticas.
- ¿Es esta persona «autoridad» en física teórica?
 - Ranking de objetos: PageRank, HITS.
- ¿Es válido documento el que me envió esta persona?
 - Inferencia: RDFS, OWL, XML Schema.

d) Asociaciones Semánticas

«Relaciones complejas que capturan conectividad y similitud de entidades en una base de conocimiento»

8. CONCLUSIONES

1. 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. Para que esto ocurra, es necesario que la información de las páginas web se codifiquen mediante ontologías.

2. En los últimos años, muchos investigadores están diseñando modelos para transformar la red desde un espacio de información a un espacio de conocimiento, referenciando datos dentro de las páginas web a metadatos con un esquema común consensuado sobre algún dominio.

3. Los metadatos no sólo especificarán el esquema de datos, sino que además podrán tener deducciones con ellos, es decir, axiomas que podrán aplicarse en los diferentes dominios que trate el conocimiento almacenado.

4. Los agentes web podrán realizar inferencias automáticamente buscando información relacionada con la que se encuentra situada en las páginas y con los requerimientos de la consulta.

5. Organizaciones como W3C se encarga de estandarizar lenguajes y herramientas para hacer efectiva la web semántica.

6. La Web Semántica es aún una visión que ha comenzado a caminar pero aún tiene mucho camino por delante.

7. La Web Semántica es un requerimiento evidente; todos los cibernautas en algún momento han pensado que «si tuviéramos respuestas más con-

- cretas, en vez de una gran lista de páginas ahorraríamos tiempo». El requerimiento de una Web «inteligente» se hizo evidente incluso antes del lanzamiento del WWW, ya que Tim Berners-Lee pensó en desarrollar un software que permita respuestas más concretas a los usuarios.
8. El futuro de la Web es la Web Semántica, esto permitirá a todos los cibernautas tener respuestas más rápidas e «inteligentes». Este cambio visto de manera global permitirá ahorrar millones de horas hombre, que se gastan en buscar información en la Web.
 9. Actualmente ya se están liberando algunos productos relativos a la Web Semántica y existen muchas instituciones y empresas que han proyectado el desarrollo, tanto en buscadores o el perfeccionamiento de las estructuras ontológicas que permitirán hacer viable la globalización de la Web Semántica.
 10. El futuro de la Web Semántica está marcado por el usuario que normalmente al ver satisfecho sus requerimientos informáticos queda maravillado, pero en poco tiempo se acostumbra a este nuevo estado de cosas, y trata de subir un escalón más. En este caso el usuario de la Web Semántica sentirá que lo «entienden» y que logra una comunicación más fluida, una vez que se acostumbra es probable que pida una «red del conocimiento global», de modo que todas sus consultas sean satisfactorias.

11. Las ontologías favorecen la comunicación entre personas, organizaciones y aplicaciones porque proporcionan una comprensión común de un dominio, de modo que se eliminan confusiones conceptuales y terminologías.

9. REFERENCIAS BIBLIOGRÁFICAS

- [1] The Semantic Web, Scientific American, May 2001, Tim Berners-Lee, James Hendler and Ora Lassila
- [2] Listas de artículos en la W3C <http://www.w3.org/2001/sw/>
- [3] Ontologías. <http://www.ontology.org> (2001)
- [4] John Davies, Dieter Fensel, Frank van Harmelen: Towards the Semantic Web: Ontology-Driven Knowledge Management, 2002.
- [5] Pablo Castells (2003) *La web semántica*. Universidad Autónoma de Madrid
- [6] Directorio de Ontologías. <http://www.schemaweb.info> (2003)
- [7] Andy Carvin: Tim Berners-Lee: Weaving a Semantic Web. Digital divide network artículos, 2005
- [8] Carolina García Cataño y David Arroyo Menéndez (2002) Biblioteca Digital y Web Semántica
- [9] Aaron Swartz interview January 23 2004 http://iron.wootest.net/aaron_swartz.php