

Aproximación a la web semántica desde la perspectiva de la Documentación

Eduardo Peis

Enrique Herrera-Viedma

José M. Morales-del-Castillo *

Artículo recibido:
22 de noviembre de 2005.

Artículo aceptado:
23 de agosto de 2006.

RESUMEN

El proyecto *web semántica* se perfila cada vez más como una nueva alternativa que viene a resolver algunos de los problemas que presenta el actual modelo de Web. Este trabajo pretende, mediante el planteamiento de ejemplos sencillos y desde el punto de vista del “profesional de la información”, hacer una aproximación a las tecnologías básicas que permiten definir una infraestructura semántica sobre la cual desarrollar los sistemas de información del futuro. Además, se analiza el impacto que puede suponer la implantación de estas tecnologías en las metodologías y técnicas de trabajo de una disciplina como la documentación. Por último se discute brevemente la viabilidad de este nuevo modelo.

* Los tres autores pertenecen a la Universidad de Granada, España, (eduardo: epeis@ugr.es); (Enrique: viedma@decsai.ugr.es); (José: josemdc@ugr.es).

Palabras clave: Web semántica; Documentación; XML; RDF; OWL.

ABSTRACT

Approaching the Semantic Web from a Documentation perspective

Eduardo Peis, Enrique Herrera-Viedma and José M. Morales del Castillo

The Semantic Web is arising as a new alternative that could solve some of the problems of the actual Web model. By means of simple examples from an information professional perspective, this paper attempts to approach the basic technologies that will define the semantic infrastructure for developing the information systems of the future. Furthermore, the impact that the implementation of these technologies may have on work methodologies and techniques of a research area such as Information Science, is analyzed. Finally, the viability of this new model is briefly discussed.

Keywords: Semantic Web; Information Science; XML; RDF; OWL

I. INTRODUCCIÓN

Partamos de un hecho constatable: estamos inmersos en la sociedad de la información. Lo queramos o no, muchos de nuestros hábitos más cotidianos nos remiten irremisiblemente a la utilización de los nuevos sistemas de información. El acceso a Internet se ha convertido en algo habitual y casi rutinario para leer el periódico, reservar vuelos o habitaciones de hotel, presentar la declaración del impuesto sobre la renta, etcétera. A modo de eslogan publicitario podríamos decir que tenemos *'todo un mundo de información a un clic de distancia'*.

Sin embargo, a pesar de que la introducción de las nuevas tecnologías en nuestra vida ha sido muy rápida, gracias principalmente a su facilidad de uso y transparencia para el usuario (no podemos obviar lo sencillo que resulta para cualquier persona navegar por Internet o utilizar un buscador), comprobamos que no todo lo que se nos ofrece es tan maravilloso como cabría esperar. Valga como ejemplo una simple consulta en uno de los muchos buscadores que podemos encontrar en la red. Cuando introducimos una serie

de palabras, su motor de búsqueda se limita a aplicarles a éstas un algoritmo que mediante técnicas estadísticas nos devuelve un listado de documentos ordenados por relevancia. Suena realmente bien, pero la cruda realidad nos demuestra que no siempre encontramos lo que estamos buscando. Si por ejemplo se nos ocurre buscar algo tan ambiguo como una definición del término *documentación* en Google, una ecuación lógica de búsqueda podría ser introducir las palabras clave *definición* y *documentación*. Como resultado obtenemos la nada despreciable cantidad de 187.000 referencias, de las cuales no hay ninguna relevante, al menos entre las 30 primeras (sinceramente, no llegamos a pasar de la tercera página de resultados). Esto sucede porque el motor de búsqueda es incapaz de comprender qué le pretendemos decir con eso de que nos busque una definición de documentación. El sólo *ve* un par de cadenas de caracteres que tienen una cierta frecuencia de aparición entre los documentos que tiene indizados y se limita a devolvernos aquellos documentos en los que esa frecuencia es mayor. Realmente no *comprende* qué le estamos preguntando. Coloquialmente podríamos decir que para el motor de búsqueda es lo mismo un documento sobre los *bancos* de atunes del Atlántico, que una noticia sobre el *Banco* Mundial, o que el sitio web de una fábrica de *bancos* de madera. Por lo tanto no sería descabellado afirmar que el modelo actual se está colapsando y ya no es capaz de manejar de forma eficiente la avalancha de información que circula por la red (y que, por otro lado, no cesa de crecer exponencialmente).

¿Pero cómo solucionar este problema? En la actualidad la información está representada fundamentalmente en lenguaje natural, de forma que la podemos leer los humanos pero no puede ser directamente interpretada por los ordenadores. Para desarrollar nuevas herramientas y servicios capaces de resolver los problemas que nos encontramos hoy día es necesario que se produzca un salto cualitativo que permita una especie de ‘razonamiento automatizado’ por parte de las máquinas. ¿Sería posible representar la información de forma que los ordenadores fueran capaces de interpretarla y ayudarnos de una manera automática a realizar nuestras búsquedas de una manera más precisa? La respuesta a estas preguntas podría ser la tan traída y llevada *web semántica*.

En este trabajo pretendemos describir la estructura y las tecnologías básicas que componen el proyecto web semántica (proponiendo una serie de sencillos ejemplos aplicados), y analizar el impacto que su implantación puede tener no sólo en la propia Web actual, sino también en las metodologías y técnicas de trabajo de una disciplina como la Documentación. Además discutimos algunos de los factores que pueden afectar la viabilidad de la implantación del modelo.

El artículo se estructura de la siguiente manera: en la sección 2, ofrecemos una visión general de las diferentes capas que componen el proyecto web semántica, sus características y sus vocabularios básicos. En la sección 3, ponemos un ejemplo del funcionamiento conjunto de estas tecnologías. La sección 4 analiza el impacto que puede suponer para la documentación la implantación de estas tecnologías. En la sección 5 se discute brevemente la viabilidad del modelo, y por último en la sección 6 se presentan las conclusiones.

2. ¿QUÉ ES LA WEB SEMÁNTICA?

La web semántica supone una extensión de la web actual, donde la información está dotada de un significado bien definido que permite una mejor cooperación entre humanos y máquinas (Berners-Lee, Hendler, Lassila 2001). Esto se basa en dos ideas principales: el etiquetado semántico de recursos (lo que implica una separación formal entre el contenido y la estructura de los documentos), y la creación de aplicaciones de software “inteligentes” (también conocidas como *agentes*) capaces de procesar y operar con estos recursos a nivel semántico (Hendler 2001).

Los agentes son entidades software que realizan procesos de recolección, filtrado, procesamiento de información e inferencia de una forma semiautónoma, los cuales cumplen el papel de *infomediarios*, es decir, intermediarios entre las necesidades de los usuarios y las fuentes de información distribuidas disponibles en la red. De hecho, el verdadero potencial de la web semántica se desarrollará cuando la web esté *poblada* por un gran número de agentes que sean capaces de recoger información de diferentes fuentes distribuidas, procesarla e intercambiar resultados con otros agentes.

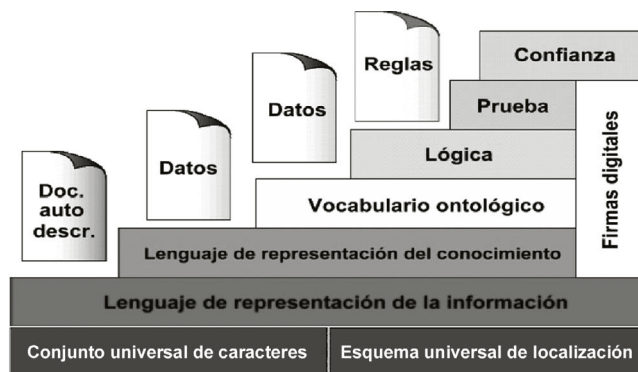


Fig. 1. Niveles de la web semántica

No obstante, para que los agentes puedan operar con la información es necesario que ésta sea interoperable; es decir, que debe estar representada (tanto a nivel sintáctico como semántico) de forma que pueda ser reutilizada tantas veces como se quiera. Basado en este principio, Berners-Lee (2000) desarrolló un modelo multicapa (ver Fig. 1), en el que se procura la máxima interoperabilidad de cada capa con la inmediatamente inferior y la inmediatamente superior (esto implica que los recursos definidos en una capa determinada puedan ser reutilizados íntegramente por las capas superiores, pero sólo parcialmente por las inferiores).

El modelo está formado por seis capas diferentes: en las tres primeras se establecen las bases para poder representar la información de una manera *semánticamente* accesible, mientras que en las tres capas superiores se definen los elementos que les permiten a los agentes software autentificar y comprobar la confiabilidad de los diferentes elementos del modelo (recursos, agentes, inferencias obtenidas, etcétera).

A continuación pasamos a describir cada una de estas capas y los elementos básicos que en ellas se definen.

2.1 La capa sintáctica

Esta capa supone el basamento del modelo de la web semántica y en ella se definen una serie de elementos que permiten el intercambio y reutilización de recursos de fuentes de información heterogéneas y distribuidas. Así, se establece el estándar *unicode*¹ como patrón de codificación de caracteres universal, y las URI (*uniform resource identifier*)² como esquema para identificar recursos de una forma unívoca. Esta identificación se puede realizar tanto a través de URL (*Universal Resource Locator*), que describen la localización física de un recurso determinado, como de URN (*Universal Resource Name*) que identifican el recurso unívocamente, independientemente de su ubicación física.

También es necesario definir un formato estándar válido para representar la información. Para ello recurrimos al lenguaje de marcado XML (*eXtensible Markup Language*) (Bray *et al.* 2004), que va a conformar la base sintáctica de todo el modelo. Este metalenguaje es una adaptación simplificada de SGML (*Standard Generalized Markup Language*) que permite de una forma simple y flexible la descripción normalizada de recursos mediante conjuntos de etiquetas, lo que lo convierte en un sistema adecuado para definir, validar y compartir recursos en la red.

1 <http://www.unicode.org/>. Consultado: 12-07-2006

2 <http://www.w3.org/Addressing/>. Consultado: 08-07-2006

Por ejemplo, si quisiéramos crear un catálogo de los libros que tenemos en casa podríamos crear un documento XML como el de la Fig. 2. En él se define el elemento *CatálogoLibros*, y dentro de éste otros sub-elementos denominados *libro* que vienen caracterizados por una serie de propiedades y atributos (título, autor, materia, resumen, etcétera), que a su vez toman un valor determinado (un literal).

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<CatalogoLibros xmlns="http://perso.wanadoo.es/kishinaru/Ejemplo">
  <libro isbn="84-473-0224-5">
    <título>Los dragones del Edén</título>
    <autores>
      <autor>Carl Sagan</autor>
      <traductor>Jaume Josa Llorca</traductor>
    </autores>
    <editorial>RBA</editorial>
    <materia>Divulgación</materia>
    <resumen>Relato de los hechos fundamentales en el desarrollo de la vida
      inteligente sobre la Tierra, en el que se analiza desde la cantidad
      de información contenida en un cromosoma, hasta posibles líneas de
      evolución del cerebro humano.
    </resumen>
  </libro>
</CatalogoLibros>
```

Fig. 2. Ejemplo de documento XML

Como se puede observar, hemos definido con total libertad el conjunto de etiquetas que nos ha parecido más adecuado para describir el contenido de los elementos de nuestro catálogo. Este conjunto de etiquetas necesita un nombre que las identifique unívocamente para distinguirlas de cualquier otro conjunto definido por otra persona. Para ello utilizamos los espacios de nombre (Bray, Hollander, Layman 1999), un sencillo método de identificación de conjuntos de etiquetas mediante URL. De esta forma es posible, por ejemplo, utilizar simultáneamente en un mismo documento varios vocabularios heterogéneos sin que exista colisión entre ellos, o darles a otras personas la opción de reutilizar un conjunto de etiquetas definido por nosotros mismos.

Una vez que tenemos descrita la información contenida en nuestros documentos XML, sería muy útil poder definir y validar tanto su contenido como su estructura. Para esta tarea podemos utilizar dos herramientas diferentes: las DTD (*document type definiton*) y el XML Schema.

Una DTD, tal y como su nombre lo indica, es la definición de un tipo de documento; es decir, la definición de una serie de requisitos para que nuestro documento XML sea considerado válido. Con una DTD podemos definir, por ejemplo, la validez de una etiqueta, su obligatoriedad, su ocurrencia e incluso el valor que

ésta puede tomar. Ejemplos de DTD serían HTML³, que no es más que un tipo de documento de SGML, o EAD (*Encoded Archival Description*)⁴ un vocabulario XML específico para descripción de archivos.

El diseño de una DTD es realmente sencillo, aunque su sintaxis difiere bastante de la utilizada en los documentos XML, como se puede ver en el ejemplo de la Fig. 3, donde se definen los elementos, atributos y valores permitidos en nuestro catálogo de libros.

```
<!ELEMENT CatalogoLibros (libro+)>
<!ELEMENT libro (titulo, autores, editorial, materia, resumen?)>
<!ATTLIST isbn CDATA #REQUIRED>
<!ELEMENT titulo {#PCDATA}>
<!ELEMENT autores (autor+, traductor)>
<!ELEMENT autor {#PCDATA}>
<!ELEMENT traductor {#PCDATA}>
<!ELEMENT editorial {#PCDATA}>
<!ELEMENT materia (Divulgación | Novela historica | Ciencia-Ficción)>
<!ELEMENT resumen {#PCDATA}>
```

Fig. 3. Ejemplo de DTD

Pero las DTD no son adecuadas para resolver problemas algo más complejos, como por ejemplo especificar el formato que debe tener un ISBN o una dirección de correo electrónico, ya que las DTD sólo soportan un número limitado de tipos de datos básicos. Por ello, para estructurar y validar documentos más complejos (o para hacerlo de una forma más precisa) será necesario recurrir a XML Schema (Thompson, *et al.* 2001), una herramienta más potente que las DTD, con capacidad para crear nuevos tipos de datos (extendiendo, restringiendo o reutilizando otros tipos de datos), y para controlar la recurrencia de los elementos y atributos de los documentos. Además cuenta con la ventaja añadida de utilizar sintaxis XML (lo que hace más fácil la interpretación de su código).

En el siguiente ejemplo vemos cómo se definiría el tipo de dato ISBN para los libros publicados en España (ejemplo extraído de (Costello, Sperberg 2002?)).

3 <http://www.w3.org/MarkUp/html-spec/html.dtd>. Consultado: 13-07-2006

4 <http://www.loc.gov/ead/>. Consultado: 10-07-2006

```

<xsd:pattern value="84-\d{0-9}|-){6}\d-[0-9x]">
  <xsd:annotation>
    <xsd:documentation>
      grupo/país ID = 84 (guión despues del segundo dígito)
      País = España. Dígito de control 0-9 o 'x'
    </xsd:documentation>
  </xsd:annotation>
</xsd:pattern>
<xsd:pattern value="84\s\d{0-9}|\s){6}\d\s[0-9x]">
  <xsd:annotation>
    <xsd:documentation>
      grupo/país ID = 84 (espacio después del segundo dígito)
      País = España Dígito de control 0-9 o 'x'
    </xsd:documentation>
  </xsd:annotation>
</xsd:pattern>

```

Fig. 4. Ejemplo de definición de un tipo de dato con XML Schema

2.2 La capa semántica

La infraestructura sintáctica está planteada, pero además de la flexibilidad, independencia, consistencia, capacidad descriptiva y discriminatoria que proporcionan las herramientas ya comentadas, necesitamos un medio para dotar de *semántica* a los recursos de la Web. Para definir un modelo semántico es necesario identificar los recursos de una forma unívoca, caracterizar los metadatos de los documentos y crear un modelo lógico de metadatos con el que se pueda operar, y a partir del cual se pueda inferir conocimiento. Con este fin, el W3C ha desarrollado las recomendaciones RDF (*resource description framework*) y RDFS (*RDF Schema*).

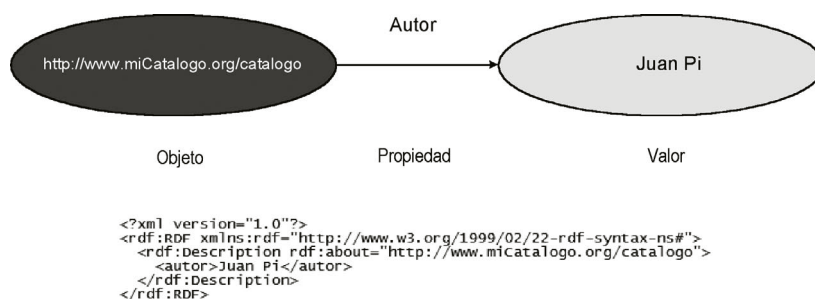


Fig. 5. Ejemplo de grafo RDF y su representación con sintaxis XML.

RDF (Beckett 2004) es un lenguaje que permite codificar, intercambiar y reutilizar metadatos estructurados. Se puede considerar como la piedra angular sobre la que se vertebra la estructura semántica de la Web. Este lenguaje le añade semántica a la información estructurándola en forma de tripletes objeto-propiedad-valor (donde el valor puede ser otro recurso o un literal)

que pueden ser representadas en forma de grafos (de hecho es habitual referirse a los documentos RDF como grafos (ver Fig. 5)).

La ventaja de utilizar RDF frente a lenguajes de etiquetado *no semánticos*, como XML, es que mientras éstos se limitan a definir un vocabulario que permite describir la información para que sea directamente interpretable por humanos, RDF da la vuelta de tuerca necesaria para que esa información sea también interpretable y procesada por los agentes software.

RDF puede ser expresado utilizando diferentes sintaxis, como Notation3 (Berners-Lee 1998) o XML, pero no cabe duda de que la serialización XML facilita la migración entre estos dos formatos, ya que sólo es necesario adaptar los documentos XML preexistentes en una nueva estructuración lógica. En la Fig. 6 vemos una posible forma (no la única) de expresar el documento XML que describe nuestro catálogo de libros en formato RDF:

```
<Catalogo rdf:ID="CatalogoLibros"
  xsi:schemaLocation="http://perso.wanadoo.es/kishimaru/Ejemplo_CataSchemaRDF.xsd">
  <registro>
    <Libro rdf:ID=" 84-473-0224-5">
      <titulo>Los dragones del Edén</titulo>
      <autores>
        <rdf:Description>
          <autor>Carl Sagan</autor>
          <traductor>Jaume Josa Llorca</traductor>
        </rdf:Description>
      </autores>
      <editorial>RBA</editorial>
      <materia>Divulgación</materia>
      <resumen>Relato de los hechos fundamentales en el desarrollo de la vida
        inteligente sobre la Tierra, en el que se analiza desde la
        cantidad de información contenida en un cromosoma, hasta posibles
        líneas de evolución del cerebro humano.
      </resumen>
    </Libro>
  </registro>
</Catalogo>
```

Fig. 6. Ejemplo de documento RDF

Como vemos, dentro del recurso *Catálogo* se anida el atributo *registro* cuyo valor es el recurso *Libro*. Éste a su vez contiene los atributos *título*, *resumen*, etcétera, cada uno con su valor correspondiente (ya sea un literal u otro recurso). Y seguiríamos anidando así recursos, atributos y valores sucesivamente según el grado de complejidad lógica de la información que estemos representando.

RDF Schema (Brickley, Guha 2002) es una extensión semántica de RDF, y define un lenguaje con el que es posible construir taxonomías de dominios particulares, definiendo clases y propiedades, relaciones básicas entre ellas (básicamente jerárquicas), y restricciones de rango y dominio para esas propiedades.

Sin embargo, y a pesar de lo que pudiera parecer, RDFS no es un lenguaje lo suficientemente potente como para poder definir ontologías (Ossenbruggen, Hardman, Rutledge 2002). Por lo tanto, RDFS se podría definir como un lenguaje de esquema *semántico* (para diferenciarlo de un lenguaje de esquema *sintáctico* como XML Schema) cuyo papel consiste en ofrecer información adicional sobre la interpretación que se debe hacer de los asertos definidos en los documentos RDF.

Veamos un ejemplo para explicar su función dentro del modelo. Supongamos un sitio web académico cuya línea de investigación se centra en estudiar los lenguajes de etiquetado orientados a la web. Una posible taxonomía de este dominio del conocimiento podría ser la que representamos en el árbol jerárquico de la Fig. 7:

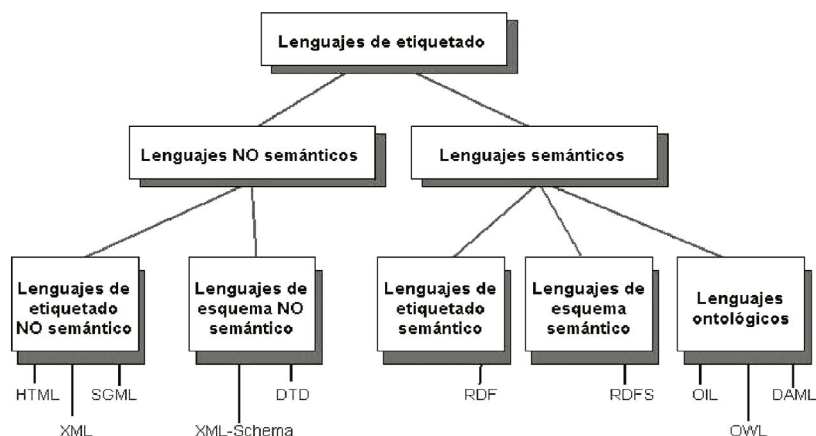


Fig. 7. Ejemplo de clasificación jerárquica

Cada rectángulo representa una clase, las líneas la relación de dependencia existente entre ellas y las palabras conectadas a los rectángulos representan algunas instancias (elementos o individuos) de cada una de estas clases. En la Fig. 8 presentamos un ejemplo de esquema RDF aplicado a esa taxonomía (por simplicidad, solo mostramos cómo expresar las entidades y dependencias jerárquicas que establecen entre si las clases de la rama derecha del árbol):

```

<rdfs:Class name="Lenguajes_Etiquetado"/>
<rdfs:Class name="Lenguajes_Semanticos">
<rdfs:subclassOf rdf:resource="#Lenguajes_Etiquetado"/>
</rdfs:Class>
<rdfs:Class name="Etiquetado_Semantico">
<rdfs:subclassOf rdf:resource="#Lenguajes_Semanticos"/>
</rdfs:Class>
<rdfs:Class name="Lenguajes_Eschema_Semantico">
<rdfs:subclassOf rdf:resource="#Lenguajes_Semanticos"/>
</rdfs:Class>
<rdfs:Class name="Lenguajes_Ontologicos">
<rdfs:subclassOf rdf:resource="#Lenguajes_Semanticos"/>
</rdfs:Class>
...

<rdf:Property rdf:ID="Define_Ontologias_Modelado">
<rdfs:domain rdf:resource="#Lenguajes_Ontologicos"/>
<rdfs:range rdf:resource="#Etiquetado_Semantico"/>
</rdf:Property>
...
    
```

Fig. 8. Ejemplo de documento RDFS

Comprobamos cómo por un lado expresamos las relaciones jerárquicas de las diferentes clases mediante las etiquetas *Class* y *subclassOf*, y por otro definimos propiedades entre clases especificando su dominio y rango de aplicación.

Pero no sólo basta con definir una estructura para dotar la información de semántica, sino que es necesario poder realizar búsquedas y extraer información de los grafos RDF. Para ello se han desarrollado diferentes lenguajes de interrogación cuya principal diferencia con los lenguajes de búsqueda tradicionales radica en su capacidad para equiparar tanto literales como estructuras semánticas (Shah, *et al.* 2002) (Guha; McCool; Miller 2003). Sin embargo, no se puede hablar aún de un lenguaje de interrogación estándar ya que no existe definida una sintaxis y semántica común. La propuesta del W3C es SPARQL (*SPARQL Protocol and RDF Query Language*) un lenguaje de interrogación y protocolo de acceso de datos que tiene una estructura similar a SQL, y que dispone de herramientas capaces de manejar diferentes tipos de datos y de realizar búsquedas en múltiples fuentes distribuidas (Prud'hommeaux, Seaborne 2006). Actualmente ya existen diferentes librerías de programación y aplicaciones, como RAP⁵ o ARQ⁶, que pueden procesar consultas en SPARQL para realizar búsquedas en bases de datos RDF.

5 <http://www.wiwiiss.fu-berlin.de/suhl/bizer/rdfapi/>. Consultado: 06-07-2006

6 <http://jena.sourceforge.net/ARQ/>. Consultado: 07-07-2006

2.3 La capa ontológica

Una vez que hemos estructurado la información de forma que pueda ser procesada por máquinas es necesario contextualizarla dentro de un dominio concreto para poder inferir conocimiento a partir de ella. Es en la capa ontológica donde podemos definir las ontologías, entendidas como la suma de una serie de conceptos relevantes del conocimiento compartido por los miembros de un dominio concreto; las relaciones que establecen entre sí estos conceptos; y los axiomas definidos sobre estos conceptos y estas relaciones. De una manera más simple, podríamos definir *ontología* como la conceptualización de una parcela de realidad. Estas ontologías le dan sentido pleno a la información situándola en un contexto, permitiéndonos dar un salto cualitativo muy importante: pasaríamos de utilizar motores de búsqueda que trabajan con palabras clave, a utilizar agentes software inteligentes que trabajan con conceptos. Esto supondría, a su vez, pasar de la mera recuperación de información a la obtención de respuestas precisas a consultas concretas, es decir, a recuperar conocimiento.

Los lenguajes utilizados para el diseño de ontologías en la Web deben de reunir una serie de características (Fensel *et al.* 2001):

- 1 Deben ser intuitivos para el usuario humano.
- 2 Deben tener una semántica formal bien definida (las máquinas deben ser capaces de interpretarlos).
- 3 Deben estar bien conectados con lenguajes web ya existentes como XML o RDF para asegurar la interoperabilidad.

Existen diferentes lenguajes formales para la definición e instanciación de ontologías en la Web, como DAML (*DARPA Agent Markup Language*)⁷, OIL (*Ontology Inference Layer*) (Horrocks, *et al.* 2000), DAML+OIL (Connolly, *et al.* 2001), y OWL (*Web Ontology Language*) (McGuinness; Harmelen 2004), que es la recomendación del World Wide Web Consortium (W3C). OWL se define como una extensión semántica de RDFS, que permite definir clases y propiedades complejas para diseñar ontologías web y que mantiene la convención sintáctica de RDF. Existen tres versiones (o especies) de OWL (van Harmelen; McGuinness 2004) que se diferencian por su capacidad expresiva: OWL Lite es la versión más simple y en ella se definen una serie de elementos básicos para crear de una manera simple ontologías fáciles de procesar; OWL DL incluye todas las funcionalidades del lenguaje pero imponiendo una

serie de restricciones sobre las propiedades de RDF que pueden ser usadas; la versión más completa es OWL Full, que incluye todas las funcionalidades del lenguaje y admite propiedades típicas de RDF, como la reificación, que permite definir sentencias lógicas sobre sentencias lógicas. A la hora de desarrollar cualquier aplicación semántica, es importante tener en cuenta las características de cada versión para elegir aquella que mejor se ajusta a nuestras necesidades específicas, ya que cuanto mayor es la capacidad expresiva de una versión, mayor será también la dificultad que tendrán los agentes para operar con ellas de una forma eficiente. Dado el nivel de desarrollo actual del modelo, para garantizar la operatividad de las aplicaciones se recomienda utilizar OWL Lite u OWL DL.

```
<owl:Class rdf:ID="Sintaxis_NO_XML"/>
<owl:Class rdf:ID="Sintaxis_XML">
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdf:about="Lenguajes_Etiquetado"/>
    <owl:Class>
      <owl:complementOf rdf:resource="#Sintaxis_NO_XML"/>
    </owl:Class>
  </owl:intersectionOf>
</owl:Class>
```

Fig. 9. Ejemplo de clase descrita en OWL

Para ejemplificar la capacidad de OWL para componer clases complejas, en la Fig. 9 definimos la clase *Sintaxis XML* como la intersección de la clase *Lenguajes de etiquetado* y el complemento lógico de la clase *Sintaxis NO XML*. O dicho de otra forma, estamos considerando dentro de la clase *Sintaxis XML* a todos aquellos lenguajes de etiquetado que no pertenecen a la clase *Sintaxis NO XML*.

2.4 La capa lógica

Sobre la capa ontológica se apoyan las tres últimas capas del modelo, en las que se tratan básicamente cuestiones relacionadas con la seguridad, veracidad y confianza de la información que intercambian entre sí los agentes software. Dado que el modelo de web semántica está a penas en su fase inicial, el nivel de desarrollo de estas tres capas es escaso aún, pero no obstante vamos a repasar los elementos fundamentales que se definen en cada una de ellas.

La primera de estas capas es la capa lógica, y en ella es en la que se establecen las diferentes reglas de producción con las que hay que describir el conjunto de deducciones que se pueden hacer a partir de un conjunto determinado de

datos. En otras palabras, en esta capa se establecen los pasos que un agente debe seguir para llegar a una conclusión con las inferencias obtenidas en un proceso de búsqueda de información. Estas reglas adoptan la forma de implicaciones entre un antecedente y un consecuente, y se pueden definir usando diversos lenguajes como RuleML (*Rule Markup Language*)⁸ o SWRL (*Semantic Web Rule Language*) (Horrocks 2003), (que permite integrar las reglas dentro del mismo código de las ontologías definidas en OWL Lite o DL).

Veamos un ejemplo. Supongamos que queremos definir una regla en un sistema de información, que defina el procedimiento que debe seguir el agente del sistema cuando un usuario solicite un libro cuya materia sea *Divulgación*. Cuando se dé esta situación, el agente del sistema deberá recomendar un libro (distinto del que ha escogido el usuario) que pertenezca a la categoría *Novedades de divulgación*. De una forma informal esta regla se podría expresar de la siguiente manera:

$$[\text{Solicita}(\text{usuario}, \text{libro1})] \wedge [\text{Materia}(\text{libro1}, \text{Divulgación})] \wedge [\text{libro2} \in \text{Novedades_Divulgación}] \wedge [\text{libro1} \neq \text{libro2}] \Rightarrow \text{Recomendar}(\text{usuario1}, \text{libro2})$$

Como vemos, hay un antecedente donde se definen una serie de requisitos que es necesario que sean ciertos (es decir, que se satisfagan) para que el consecuente también sea cierto y por lo tanto la regla se pueda aplicar. Si la traducimos a SWRL, la regla tendría esta forma:

```

<swrl:Variable rdf:ID="usuario"/>
<swrl:Variable rdf:ID="libro1"/>
<swrl:Variable rdf:ID="libro2"/>
<ruleml:imp>
  <ruleml:body rdf:parseType="Collection">
    <swrl:individualPropertyAtom>
      <swrl:propertyPredicate rdf:resource="solicita"/>
      <swrl:argument1 rdf:resource="#usuario"/>
      <swrl:argument2 rdf:resource="#libro1"/>
    </swrl:individualPropertyAtom>
    <swrl:DatavaluedPropertyAtom>
      <swrl:propertyPredicate rdf:resource="materia"/>
      <swrl:argument1 rdf:resource="#libro1"/>
      <swrl:argument2 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        Divulgación
      </swrl:argument2>
    </swrl:DatavaluedPropertyAtom>
    <swrl:ClassAtom>
      <swrl:classPredicate rdf:resource="Novedades_Divulgacion"/>
      <swrl:argument1 rdf:resource="#libro2"/>
    </swrl:ClassAtom>
    <swrlb:notequal>
      <swrl:argument1 rdf:resource="#libro1"/>
      <swrl:argument2 rdf:resource="#libro2"/>
    </swrlb:notequal>
  </ruleml:body>
  <ruleml:head rdf:parseType="Collection">
    <swrl:individualPropertyAtom>
      <swrl:propertyPredicate rdf:resource="recomendar"/>
      <swrl:argument1 rdf:resource="#usuario"/>
      <swrl:argument2 rdf:resource="#libro2"/>
    </swrl:individualPropertyAtom>
  </ruleml:head>
</ruleml:imp>
    
```

Fig. 10. Ejemplo de regla descrita en SWRL

Como resultado el agente personal del usuario recibiría la referencia del libro que estaba buscando y además otra referencia de un libro recomendado por el agente del sistema.

2.5 Las capas de prueba y de confianza

Uno de los elementos clave de la web semántica es el establecimiento de mecanismos de seguridad adecuados que permitan autenticar y comprobar la confiabilidad de los diferentes elementos del modelo (desde los recursos, hasta los agentes, las entidades emisoras o los propios individuos) para de esta forma garantizar la fiabilidad de los resultados obtenidos en un proceso de búsqueda de información. Para desarrollar estos mecanismos se definen dos nuevas capas en el modelo: la capa de prueba y la capa de confianza.

La capa de prueba es la inmediatamente superior a la capa lógica y permite definir una infraestructura adecuada para que los agentes puedan establecer

relaciones lógicas complejas con otros agentes, e intercambiar pruebas sobre las inferencias obtenidas en una búsqueda (Finin, Joshi 2002). Esta infraestructura se apoya básicamente en tres elementos: i) las reglas de inferencia definidas en la capa lógica donde se especifican los requerimientos de seguridad; ii) la capacidad de los agentes de rastrear y probar el origen de una secuencia lógica gracias a la estructura en que está representada la información (Antinou, van Harmelen 2004); y iii) las firmas digitales.

Las firmas digitales son bloques de datos cifrados que permiten a los agentes verificar que determinada información proviene de una fuente fiable (Berners-Lee, Hendler, Lassila 2001). A todo documento que va a ser firmado digitalmente se le aplica un algoritmo de “hashing” que devuelve un bloque de datos que representa a dicho documento. Este bloque es posteriormente encriptado mediante un sistema de Clave Pública (Mohapatra 2000) que permite identificar al autor/emisor del documento y comprobar la integridad del contenido (es decir, que el documento no ha sido alterado fraudulentamente). Sin embargo, el uso de firmas digitales presenta dos inconvenientes: por un lado es necesario validar la encriptación de Clave Pública a través de un certificado emitido por una autoridad de certificación (centralizada), lo cual choca directamente con la naturaleza descentralizada de la web semántica. Por otro lado, las firmas digitales no permiten establecer la confianza que a un individuo le merece determinada fuente o agente (y consecuentemente el contenido de los recursos e inferencias que se pueden obtener de ellos).

La solución a estas deficiencias pasa por la definición de una última capa, la capa de confianza, en la que se establecen extensas redes sociales de confianza (Richardson, Agrawal, Domingos 2003) que les permitirán a los agentes determinar la confiabilidad de una determinada fuente o recurso. En estas redes todo usuario de la Web está identificado por su propia URI y expresa el grado de confianza o desconfianza (Guha *et al.* 2004) que le merecen aquellos individuos que conoce (o aquellos con los que ha tenido algún tipo de contacto). Estos grados de confianza podrán definirse, por ejemplo, utilizando una extensión de FOAF (*Friend Of A Friend*) (Brickley, Millar 2005), que es un vocabulario RDF específico para describir relaciones sociales en la Web (Hendler, Goldbeck, Parsia 2002).

De esta manera, aplicando conjuntamente las reglas de inferencia definidas en la capa lógica, las firmas digitales y las redes de confianza, sería posible construir la denominada *Web of Trust* (web de confianza) en la que se dispondría de mecanismos de autenticación y validación precisos. De esta forma, ya no sería necesario recurrir a la encriptación de datos y, consecuentemente, se podría prescindir de las autoridades de certificación, con lo que se resolvería el problema de la centralización (Reagle 2002).

3. TODA LA MAQUINARIA EN FUNCIONAMIENTO

Situémonos en el más optimista de los escenarios, un escenario en el que en Internet todos los recursos están descritos en formato RDF y todos los dominios de conocimiento disponen de su correspondiente ontología web.

Supongamos que somos los gestores de una web académica especializada en lenguajes de etiquetado orientados a la Web. Supongamos que nuestros recursos están accesibles en la red codificados en forma de documentos RDF y de acuerdo con la ontología que han desarrollado reputados organismos que son una autoridad en el tema. Como sólo nos interesan los artículos científicos y las contribuciones al Congreso sobre el tema, controlamos la estructura de nuestros recursos utilizando para ello varios XML Schemas que validan y definen ambos tipos de documentos.

En otro lugar del mundo, un estudiante de Documentación llamado Marco necesita información sobre las distintas especies de OWL, para realizar un trabajo de clase. Accede al interfaz de búsqueda de su agente software personal y le lanza la consulta, estableciendo unos criterios mínimos de fiabilidad y relevancia de los documentos que va a recuperar. El agente dispone de una ontología que controla el perfil de usuario de Marco y comprueba que debe realizar la búsqueda dentro del dominio de los lenguajes de etiquetado, y descartar otros dominios. De esta forma se asegura de no recuperar, por ejemplo, recursos sobre ornitología (recordemos que, en inglés, *owl* significa búho).

El agente de Marco se encarga de interactuar con otros agentes de confianza (entre ellos el nuestro) a los que traslada su consulta. El agente de nuestra web evalúa los criterios de búsqueda y comprueba que existen una serie de documentos o partes de documentos que son relevantes para la consulta. Esto se lo hace saber al agente de Marco, el cual le pide a nuestro agente pruebas del razonamiento seguido para comprobar que le estamos ofreciendo una información pertinente.

Una vez hecha esta comprobación, con todos y cada uno de los agentes consultados, el agente de Marco recopila e integra la información obtenida, presentando en pantalla el resultado de la búsqueda de acuerdo con los criterios establecidos en la ontología de visualización del interfaz del agente. Marco obtiene una respuesta precisa a su consulta: un enlace con un artículo en el que se repasan las características, ventajas e inconvenientes de cada una de las especies de OWL; otro que le remite a un apartado concreto de una contribución a un congreso de *Documentalistas en Internet* que trata específicamente sobre las diferentes especies de OWL; y un último enlace, para ampliar información, le remite a la especificación de OWL publicada por el W3C.

Esta situación es hoy por hoy ciencia ficción, pero ya se están colocando los cimientos para que en un plazo no demasiado largo sea una realidad.

4. DOCUMENTACIÓN Y WEB SEMÁNTICA

Como vemos, la adopción de la web semántica como nuevo paradigma tecnológico no solo implicará una profunda transformación en la Web que conocemos hoy día, sino que en disciplinas como la Documentación, cuyas materias primas de trabajo son la Información y el Conocimiento, su implantación podría suponer una revolución equiparable a la ocurrida al sustituir los catálogos manuales por los automatizados. Las diferentes tecnologías que subyacen a este nuevo modelo de Web le ofrecerán al profesional de la información la posibilidad de disponer de un arsenal de herramientas flexibles y potentes que, de forma análoga a los lenguajes documentales, le permitirán ejercer el control necesario sobre la información para procurar su descripción, acceso y recuperación eficaz, pero a niveles impensables hasta ahora.

En este nuevo escenario los documentalistas tendrán que desempeñar un papel fundamental en el diseño conceptual de ontologías dependientes de dominio, que servirán para desarrollar y controlar diferentes servicios y procesos dentro de los sistemas de información. Multitud de estos procesos, que se suelen realizar a diario de forma manual o semiautomática, se verán transformados de una forma radical. Y no sólo nos referimos a la recuperación de información (tal y como la hemos visto hasta ahora). Sin ir más lejos, en el proceso de catalogación de recursos se tenderá a utilizar lenguajes para el modelado de datos, como RDF, que sirvan como plataforma de descripción. Gracias al uso de editores “semánticos” la gran mayoría de los recursos estarán ya “catalogados” en su origen, antes de entrar al sistema de información. Progresivamente se impondrá el uso de perfiles de aplicación (espacios de nombre estándar que podremos combinar como deseemos de acuerdo con las necesidades de nuestro sistema), al tiempo que otros formatos de descripción como MARC (*Machine-Readable Cataloguing*) se verán avocados a desaparecer y a ser sustituidos por otros estándares de descripción de recursos mediante metadatos, como los estándares DCMI (*Dublin Core Metadata Initiative*)⁹ o PRISM (*Publishing Requirements for Industry Standard Metadata*).¹⁰

La automatización de catálogos ya produjo en su momento que los sistemas de clasificación clásicos como la nada intuitiva CDU o los encabezamientos de materia quedaran totalmente obsoletos en ese nuevo contexto

9 <http://www.dublincore.org/>. Consultado: 07-07-2006

10 <http://www.prismstandard.org/>. Consultado: 06-07-2006

de trabajo. De forma similar, la indización en un entorno “semántico” se verá modificada y pasará de ser un proceso intelectual a ser uno asistido por aplicaciones capaces de extraer términos representativos de cada recurso y hacerlos corresponder con los términos admitidos en el tesoro del sistema, que estará definido con SKOS Core (*Simple Knowledge Organisation System*) (Miles, Brickley 2005): una aplicación RDF específica para facilitar la migración, enriquecimiento e intercambio de tesauros en la Web.

Las tecnologías de web semántica abren también una nueva vía para encontrar solución a problemas endémicos en la generación de resúmenes automáticos, ya que en una Web donde los documentos estén descritos semánticamente se podrán desarrollar agentes capaces de distinguir en diferentes tipos de recursos aquellas partes o áreas susceptibles de ser resumidas, y además asistir en el análisis morfológico, sintáctico, semántico y pragmático del contenido, recurriendo para ello a las ontologías adecuadas.

Los nuevos sistemas de información aprovecharán su infraestructura semántica para aplicar técnicas de minería web semántica y buscar modelos o patrones en el contenido y la estructura intrínseca de sus recursos, en las relaciones que establecen con otros recursos y en la forma en que son utilizados por los usuarios. Esta información será, a su vez, de gran utilidad para el modelado de perfiles de usuario, donde aspectos mudables como las preferencias, costumbres de navegación, gustos, intereses, permisos de acceso a los recursos, etcétera, de un individuo concreto, podrán ser expresados mediante metadatos y actualizados dinámicamente para reflejar fielmente sus variaciones en el tiempo. De esta forma se podrán desarrollar servicios web, y sistemas de filtrado y recomendación más precisos (Herrera-Viedma, Peis, Morales-del-Castillo 2006).

Las capacidades que ofrecen estas tecnologías para compartir y reutilizar recursos permitirán a los sistemas de bibliotecas definir, por ejemplo, sus políticas de préstamo o incluso las de expurgo, al especificar diferentes conjuntos de reglas que permitan controlar estos procesos a través de una red de agentes y ontologías compartidas. Más aún, las bibliotecas digitales podrían adaptar algunos de los servicios que se ofrecen en las bibliotecas convencionales a este nuevo medio, y aprovechar algunas de estas tecnologías. Por ejemplo, se podrían crear servicios de difusión selectiva de información (DSI), definiendo por un lado perfiles de usuario enriquecidos en formato RDF y por otro una especie de *boletines de novedades* en los que se recogieran recursos novedosos adquiridos por la biblioteca, o aquellos que por algún otro motivo pudieran ser de interés para los usuarios registrados de la biblioteca. Estos boletines estarían definidos en forma de canales RSS 1.0 (Bege-Dov *et al.* 2001), vocabulario RDF que sirve para publicar y gestionar listados de hiperenlaces de una manera sencilla y flexible.

De nuevo puede parecer que estemos planteando meras hipótesis, pero no cabe duda de que si estas potentes tecnologías representan el futuro de los sistemas de información (Peis *et al.* 2003), las Ciencias de la Documentación (y sus profesionales) no pueden obviar esta circunstancia y deben apostar por adaptarse a esta nueva Web en la que, por sus propias características, el papel de los profesionales de la información puede ser más relevante que nunca.

5. LA VIABILIDAD DEL PROYECTO

A pesar de que el modelo teórico de la web semántica está aceptado hace tiempo por una gran parte de la comunidad científica, no han cesado de surgir voces críticas que, esgrimiendo diferentes argumentos, ponen en entredicho su viabilidad.

Algunos autores se aproximan a la web semántica como aplicación de la inteligencia artificial (IA) al ámbito web, y advierten sobre las inconsistencias semánticas que se pueden derivar de la definición de algunos elementos del modelo (Patel-Schneider; Fensel 2002), o sobre la dificultad de extraer semántica (más allá de la explícita) de los documentos expresados en lenguaje natural (Sparck-Jones 2004). Sin embargo, y a pesar de estas deficiencias (que no tienen por qué ser insalvables) es importante no perder la perspectiva de lo que un proyecto de esta envergadura puede llegar a ofrecer. Tomemos como ejemplo el caso de la propia inteligencia artificial, rama de las Ciencias de la Computación cuyo principal (y ambicioso) objetivo es conseguir diseñar entidades artificiales que exhiban algún tipo de comportamiento inteligente, aunque por el momento aún no se haya conseguido una computadora que sea capaz de emular fielmente el razonamiento humano. No obstante, diferentes áreas de la IA sí tienen una aplicación real y sirven para resolver eficazmente determinados problemas. De igual manera, la web semántica pretende aportar su propia visión del mundo y proponer algunas soluciones.

Aun así, hay que reconocer que existen escollos que ralentizan su desarrollo. Sin embargo, uno de los mayores escollos a salvar radica en la necesidad de que los recursos estén marcados utilizando RDF, y que además sean instancias de una determinada ontología aquellas donde se recoja la semántica de los elementos que los componen. Esta tarea puede ser tremendamente ardua y costosa si no se dispone de mecanismos que asistan al autor de contenidos a generar este etiquetado de una manera similar a como hoy día se crean los documentos HTML. Esta circunstancia hace que muchos de ellos se muestren reticentes a realizar este esfuerzo extra, más aún si lo tienen que realizar de forma altruista. Por esta razón gran parte de los esfuerzos de investigación se

están dirigiendo hacia el desarrollo de mecanismos y aplicaciones que permitan el etiquetado de recursos (Soo *et al* 2003) y la generación de ontologías (Tho, Cao 2006) de forma automática.

No obstante, no todo son dificultades ya que los primeros pasos tangibles para realizar la transición del actual modelo al de web semántica se están dando ya desde los movimientos de software libre que promueve la Web 2.0. Este nuevo concepto de Web (del que aún no se tiene una definición formal consensuada) se caracteriza principalmente por adoptar un enfoque orientado hacia el usuario, la interacción y el desarrollo de redes sociales, donde los sitios web actúan más como puntos de encuentro entre usuarios que como webs tradicionales (independientemente del uso de una tecnología en concreto)¹¹. Algunos ejemplos de aplicaciones Web 2.0 son Flickr¹², del.icio.us¹³ y CiteUlike¹⁴, donde los usuarios proponen sus propias categorías para describir fotos, listados de hiperenlaces favoritos o referencias bibliográficas, respectivamente, y las comparten de forma altruista con el resto de usuarios para facilitarles el acceso a esos recursos. Esta filosofía supone de facto la aplicación práctica, aunque a pequeña escala, de conceptos, principios y tecnologías que encajan dentro del proyecto web semántica (como los microformatos (Kare, Tantek 2006), o los vocabularios como FOAF (*friend of a friend*) y RSS (*RDF Site Summary*). Otras iniciativas son más ambiciosas y van más allá del simple marcado de recursos puesto que proponen generar ontologías simples a partir de las semánticas emergentes que se pueden extraer del conjunto de etiquetas usado por los usuarios de una determinada comunidad (Gruber 2005) (Mika 2005).

Otra posible vía de evolución de la web semántica podría consistir en no circunscribirla dentro de la Web actual, sino desarrollarla como una red paralela en la que la prioridad sea ofrecer información muy específica y especializada principalmente a una comunidad científica e investigadora. No es necesario rasgarse las vestiduras ante la idea de dos Web paralelas pues de hecho ya existe el proyecto de diferentes instituciones gubernamentales y académicas para desarrollar Internet2¹⁵. La idea no sería crear una Internet de primera clase y otra de segunda, sino ofrecer diferentes servicios, cada uno desde la plataforma adecuada. ¿Acaso acudimos a una biblioteca pública cuando estamos buscando un libro de ecuaciones diferenciales?

En definitiva, sea cual sea la dirección en que evolucione finalmente la web semántica, queda claro que debido a la envergadura del proyecto, el desarrollo

11 http://es.wikipedia.org/wiki/Web_2.0. Consultado: 10-07-2006

12 <http://www.flickr.com/>. Consultado: 12-07-2006

13 <http://del.icio.us/>. Consultado: 12-07-2006

14 <http://www.citeulike.org/>. Consultado: 12-07-2006

15 <http://www.internet2.edu/>. Consultado: 10-07-2006

completo del modelo no se llevará a cabo a corto plazo (aunque ya se están dando los primeros pasos). No obstante, y a pesar de los grandes retos que hay que afrontar, confiamos plenamente en que el esfuerzo cooperativo de los numerosos colectivos y sectores implicados para hacerla viable dará sus frutos en un futuro no muy lejano.

6. CONCLUSIONES

En este trabajo hemos visto cómo el modelo actual de Web ya no es capaz de manejar el gran volumen de información que circula por la red, por lo que se hace imprescindible la aplicación de tecnologías que permitan representar, procesar y gestionar la información de una forma más precisa y eficaz.

El proyecto web semántica representa este nuevo paradigma tecnológico que supondrá una revolución no solo en la propia Web, entendida como sistema de información, sino que implicará necesariamente una profunda transformación en las metodologías y técnicas de trabajo tradicionales de disciplinas que, como la documentación, trabajan con la información y el conocimiento.

A pesar de lo ambicioso del proyecto y de las dificultades que hoy por hoy existen para su implantación a gran escala, desde diferentes sectores (instituciones científicas, empresas privadas y movimientos de software libre) ya se están dando los primeros pasos para sentar las bases de lo que será la Web del futuro.

BIBLIOGRAFÍA

- Antinou, G.; van Harmelen, F. (2004). "A Semantic Web Primer". MIT Press
- Beckett, D. (ed.) (2004). "RDF/XML Syntax Specification (Revised)". Consultado: 13-07-2006.
<http://www.w3.org/TR/rdf-syntax-grammar/>
- Beged-Dov, G. *et al.* (2001). "RDF Site Summary (RSS) 1.0". Consultado: 09-07-2006.
<http://web.resource.org/rss/1.0/spec>
- Berners-Lee, T. (1998). "Notation 3: An RDF language for the Semantic Web". Consultado: 17-07-2006.
<http://www.w3.org/DesignIssues/Notation3.html>
- Berners-Lee, T. (2000). "Semantic Web - XML2000". Consultado: 13-07-2006.
<http://www.w3.org/2000/Talks/1206-xml2k-tbl/>

- Berners-Lee, T.; Hendler, J.; Lassila, O. (2001) "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". Scientific American, May 2001.
- Bray, T.; *et al.* (eds.) (2004). "Extensible Markup Language (XML) 1.0 (Third Edition) W3C Recommendation 04 February 2004". Consultado: 19-07-2006.
<http://www.w3.org/TR/REC-xml/>
- Bray, T.; Hollander, D.; Layman, A. (eds.) (1999). "Namespaces in XML". Consultado: 12-07-2006.
<http://www.w3.org/TR/REC-xml-names>
- Brickley, D.; Guha, R.V. (eds.) (2002). "RDF Vocabulary Description Language 1.0: RDF Schema". Consultado: 15-07-2006.
<http://www.w3.org/TR/rdf-schema>
- Brickley, D.; Miller, L. (2005). "FOAF vocabulary specification". Consultado: 20-07-2006.
<http://www.xmlns.com/foaf/0.1/>
- Connolly, D., *et al.* (eds.) (2001). "DAML+OIL Reference Description. W3C Note. 18 December 2001". Consultado: 01-07-2006.
<http://www.w3.org/TR/daml+oil-reference>
- Costello, R., Sperberg, R. (2002?). "XML Schema simpleType Definition of an ISBN". Consultado: 20-07-2006.
<http://www.xfront.com/isbn.xsd>
- Fensel, D. *et al.* (2001). "OIL: An ontology infrastructure for the Semantic Web". IEEE Intelligent Systems, marzo/abril 2001, pp. 38-45.
- Finin, T., Joshi, A. (2002). "Agents, trust, and information access on the semantic web". ACM SIGMOD Record, vol. 31 (4), pp. 30-35
- Gruber, T. (2005). "Ontology of Folksonomy: A Mash-up of Apples and Oranges". Consultado: 20-07-2006.
<http://tomgruber.org/writing/ontology-of-folksonomy.htm>
- Guha, R., *et al* (2004). "Propagation of trust and distrust". Proceedings of the 13th international World Wide Web Conference, pp. 403-412
- Guha, R.; McCool, R.; Miller, E. (2003). "Semantic search". Proceedings of the 12th International World Wide Web Conference. pp. 700-709.
- Hendler, J. (2001). "Agents and the Semantic Web". IEEE Intelligent Systems, marzo/abril 2001, pp. 30-37.
- Hendler, J.; Golbeck, J.; Parsia, B. (2002). "Trust networks on the Semantic Web". Consultado: 05-07-2006.
<http://www.mindswap.org/papers/CIA03.pdf>
- Herrera-Viedma, E.; Peis, E.; Morales-del-Castillo, J.M. (2006). "A Fuzzy Linguistic Multi-agent Model Based on Semantic Web Technologies and User Profiles". Studies in Fuzziness and Soft Computing, vol. 197, Jan 2006, pp. 105-120

- Horrocks, I., *et al.* (2003). "SWRL: A Semantic Web Rule Language Combining OWL and RuleML". Consultado: 17-07-2006.
<http://www.daml.org/2003/11/swrl/>
- Horrocks, I.; *et al.* (eds.) (2000). "The Ontology Inference Layer OIL". Consultado: 12-07-2006.
<http://www.ontoknowledge.org/oil/TR/oil.long.html>
- Khare, R., Tantek, C. (2006). "Microformats: a pragmatic path to the semantic web". Proceedings of the 15th international conference on World Wide Web, pp. 865-866
- McGuinness, D. L.; van Harmelen, F. (2004). "OWL Web Ontology Language Overview". Consultado: 15-07-2006.
<http://www.w3.org/TR/2004/REC-owl-features-20040210/>
- Mika, P. (2005). "Ontologies are us: a unified model of social networks and semantics". Consultado: 20-07-2006.
<http://www.cs.vu.nl/~pmika/research/papers/ISWC-folksonomy.pdf>
- Miles, A.; Brickley, D. (2005). "SKOS Core Guide". Consultado: 30-06-2007.
<http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>
- Mohapatra, P. K. (2000). "Public key cryptography". Crossroads, vol. 7 (1), pp. 14-22
- Ossenbruggen, J.; Hardman, L.; Rutledge, L. (2002). "Hypermedia and the Semantic Web: A Research Agenda". Journal of Digital Information, vol. 3(1).
- Patel-Schneider, P. F.; Fensel, D. (2002). "Layering the Semantic Web: problems and directions". Proceedings of the ISWC 2002, pp. 16-29
- Peis, E.; Herrera-Viedma, E.; Hassan, Y.; Herrera, J. C. (2003). "Análisis de la web semántica: estado actual y requisitos futuros". El Profesional de la Información, vol. 12(5), 368-376.
- Prud'hommeaux, E., Seaborne, A. (eds.) (2006). "SPARQL Query Language for RDF". Consultado: 20-07-2006.
<http://www.w3.org/TR/rdf-sparql-query/>
- Reagle, J. M. (2002). "Key Free Trust in the Semantic Web: Finding Bacon's Key". Consultado: 17-07-2006.
<http://www.w3.org/2002/03/key-free-trust.html>
- Richardson, M.; Agrawal, R.; Domingos, P. (2003). "Trust Management for the Semantic Web". En: Proceedings of the 2nd International Semantic Web Conference 2003.
- Shah, U., *et al.* (2002). "Information Retrieval on the Semantic Web". Proceedings of the 10th International Conference on Information and Knowledge Management 2002.
- Soo, V. W., *et al.* (2003). "Automatic metadata creation: Automated semantic annotation and retrieval based on sharable ontology and case-based learning techniques". Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries, pp. 61-72

- Sparck-Jones, K. (2004). "What's new about the Semantic Web? Some questions", ACM SIGIR Forum, vol 38 (2), pp.18-23.
- Tho, Q. T., Cao, T. H. (2006). "Automatic Fuzzy Ontology Generation for Semantic Web". IEEE Transactions on Knowledge and Data Engineering, vol. 18 (6), pp. 1041-4347
- Thompson, H., *et al.* (eds.) (2001). "XML Schema Part 1: Structures. W3C Recommendation, 2 May 2001". Consultado: 12-07-2006. <http://www.w3.org/TR/xmlschema-0/>
- van Harmelen, F.; McGuinness, D. L. (eds.) (2004). "OWL Web Ontology Language Overview". Consultado: 15-07-2006. <http://www.w3.org/TR/owl-features/>

