

# Análisis de la web semántica: estado actual y requisitos futuros

Por E. Peis; E. Herrera-Viedma; Y. Hassan y J. C. Herrera



Eduardo Peis, doctor en documentación, es profesor de la Universidad de Granada desde 1991. Especializado en la investigación sobre metalenguajes normalizados como sgml/xml y su aplicación a los servicios de información. La confluencia entre la docencia y dicha línea de investigación ha producido numerosos trabajos (publicados en las principales revistas internacionales del área), alguno de los cuales se centra en el papel de los modelos metadatos en el diseño

de los modernos sistemas de gestión de archivos electrónicos. La representación de la información en la web, con xml/xsl, y la descripción de recursos, con rdf, completan su actividad investigadora, ahora especialmente dirigida hacia la futura web semántica.

**Resumen:** El gigantesco y creciente número de recursos web evidencia que las técnicas léxico-estadísticas, por sí solas, no pueden solucionar la problemática de la recuperación de la información. El proyecto de web semántica intenta paliar estos problemas (y mucho más) mediante la "comprensión" del contenido de los recursos por los ordenadores. Este trabajo analiza el estado de las "capas" propuestas por **Berners-Lee** como necesarias para el desarrollo de la web semántica: la capa de identificación uniforme, la capa de representación, la capa lógica y la capa de seguridad.

**Palabras clave:** Web semántica, Uri, Xml, Rdf, Ontologías, Taxonomías, Agentes inteligentes, Firma digital, Redes de confianza.

## Title: Semantic web analysis: current status and future requirements

**Abstract:** The vast and growing number of web resources show that lexical-statistical aids by themselves cannot resolve the problem of information retrieval. The "semantic web" project tries to solve these problems (and to do much more) by having the computers comprehend the content of the resources. This paper examines the current status of the required 'la-



Yusef Hassan Montero, es licenciado en documentación y estudiante de doctorado en la Facultad de Biblioteconomía y Documentación de la Universidad de Granada. Con larga experiencia en desarrollo web, compagina su labor como becario de apoyo técnico a la docencia con su función de arquitecto de información y administrador del magazín electrónico sobre diseño e investigación web NoSoloUsabilidad.com. Sus líneas de investigación se centran en visualización de

información, usabilidad de interfaces web y representación de conocimiento en la web semántica.

Enrique Herrera Viedma, doctor en informática por la Universidad de Granada desde 1996, actualmente es profesor titular del Dpto. de Ciencias de la Computación e I. A. y vicedecano de ordenación académica en la Facultad de Biblioteconomía y Documentación de la Univ. de Granada. Entre sus principales temas de investigación están el modelado de información lingüística mediante técnicas difusas, los sistemas de toma de decisiones con múltiples criterios, los

sistemas de recuperación de información así como la representación y evaluación de información en la web usando metalenguajes como xml. Sobre estos temas ha publicado en revistas internacionales, congresos internacionales y nacionales, editado 5 números especiales en revistas internacionales, organizando diferentes sesiones especiales en congresos internacionales. Actualmente imparte la asignatura de sistemas expertos y recuperación de información en la licenciatura de documentación de la Univ. de Granada.



Juan Carlos Herrera Torres, es licenciado en biblioteconomía y documentación por la Universidad de Granada. Actualmente es becario de investigación con cargo a un proyecto de digitalización del fondo antiguo de la Biblioteca Universitaria del Hospital Real de la citada universidad, donde está cursando el doctorado en documentación e información científica. Sus líneas de investigación versan sobre los metalenguaje normalizados (como xml), su aplicación a los servicios de información, la representación de la información en la web y la web semántica.



yers' put forward by **Berners-Lee** for the development of the Semantic web: namely, the uniform identification layer, the representation layer, the logical layer, and the security layer.

**Keywords:** Semantic web, Uri, Xml, Rdf, Ontologies, Taxonomies, Intelligent agents, Digital signatures, Trusted networks.

**Peis, E.; Herrera-Viedma, E.; Hassan, Y.; Herrera, J. C.** "Análisis de la web semántica: estado actual y requisitos futuros". En: *El profesional de la información*, 2003, septiembre-octubre, v. 12, n. 5, pp. 368-376.

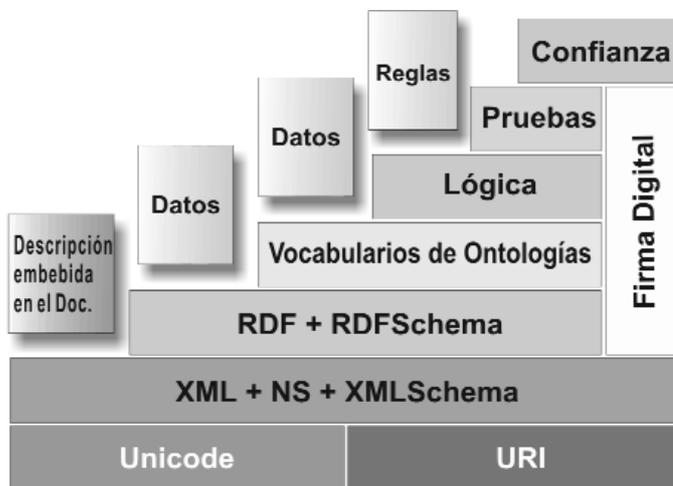
## 1. Introducción

La web, tal y como la conocemos, es un gigantesco repositorio de hiperdocumentos cuyo diseño únicamente permite su comprensión por humanos. Estos materiales están confeccionados usando lenguajes de etiquetado que expresan la forma en que los navegadores deben presentar su contenido (colores, maquetación, fuentes, etc.) y no su significado o semántica. Dado el gigantesco y creciente número de estos recursos, los actuales motores de búsqueda son incapaces de ofrecer tasas de precisión mínimamente adecuadas en sus resultados, evidenciando que las técnicas léxico-estadísticas no pueden solucionar por sí solas la problemática de la recuperación de información.

La propuesta inicial de world wide web que **Berners-Lee** (1989) ideó, ya apuntaba a lo que posteriormente se ha denominado web semántica (**Berners-Lee; Hendler; Lassila**, 2001), una extensión de la actual web cuyo objetivo es que no sólo los humanos, sino también las máquinas, sean capaces de "comprender" el contenido de los documentos. De esta forma, en el proceso de recuperación de información, el usuario interrogaría a un agente software que realizaría tareas complejas de asociación e inferencia de conocimiento, devolviendo al usuario resultados precisos y contextualizados. Para conseguirlo, sería necesario proporcionar semántica a la web y, esta tarea, podría realizarse mediante la elaboración de ontologías y etiquetados descriptivos.

Para cumplir los objetivos, **Berners-Lee** (1998) considera que son necesarias varias capas o niveles, entre las que destacamos: un modelo básico para establecer asertos (propiedades sobre los recursos), para el que se empleará rdf (*resource description framework*) (**Lassila; Swick**, 1999); un modelo para definir relaciones entre las fuentes a través de clases y objetos expresado mediante *rdf schema* (**Brickley; Guha**, 2002); una capa lógica que permita realizar consultas e inferir conocimiento, donde entrarían en juego las ontologías y los agentes software; y una capa de seguridad que permita asignar niveles de fiabilidad a determinados recursos de forma comprobable posteriormente por los agentes, para lo que se usarán firmas digitales y redes de "confianza". El diagrama de la figura 1, presentado por **Berners-Lee** en la *Xml conference* de 2000, nos puede servir como aproximación visual al conjunto de tecnologías que forman el esquema de capas mencionado.

Los siguientes apartados de este trabajo tienen por objeto ofrecer una visión general de las tecnologías, herramientas y grupos de trabajo implicados, así como del estado de la cuestión del proyecto de web semántica. Se procede a analizar el esquema de capas de la siguiente forma: en el punto 2 se estudian aquellas tecnologías y lenguajes que conforman las primeras capas del esquema, que tienen la función de describir, estructurar e identificar los recursos web y que poseen soporte tecnológico real en la actual web (uris, xml, rdf y *rdf schema*). En el apartado 3 se analizarán las capas que pretenden dotar de lógica e inteligencia a la web y que, por el momento, no tienen soporte en la actual (ontologías y agentes). Posteriormente se considerarán las capas necesarias para dotar de fiabilidad, confianza y seguridad de acceso a los recursos web y que actualmente se encuentran en discusión (firmas digitales y redes de confianza). Por último, mostraremos nuestra propia opinión sobre el proyecto de web semántica y su viabilidad.



Semantic Web "layer cake" (Berners-Lee. XML Conference, 2002)

Figura 1

## 2. Uris, xml, rdf y rdf schema

Mediante los uri (*uniform resource identifier*) (**Berners-Lee et al.**, 1998),

cuyo subconjunto más conocido son los url, podemos identificar de forma inequívoca cualquier recurso en la Red: artículos, imágenes, sonidos, etc. Es el fundamento básico sobre el que se asienta la web, aunque será con la web semántica cuando adquirirá verdaderamente su función de identificador de objetos. Cualquiera podrá ser identificado mediante un uri: nuestro microondas tendrá uno asociado, el de nuestra web personal o de nuestra dirección e-mail nos identificará a nosotros, la función que realizamos en nuestro trabajo se expresará mediante otro, etc. También es cierto que pueden surgir problemas, ya que uno mismo podría identificar dos cosas distintas. De hecho, existe un grupo de trabajo del W3C que trata de resolver estos problemas de identificación en la web semántica<sup>1</sup>.

**«La web semántica es una extensión de la actual web cuyo objetivo es que no sólo los humanos, sino también las máquinas, sean capaces de 'comprender' el contenido de los documentos»**

En un nivel superior encontramos los documentos y su estructuración lógica. Xml, base sintáctica de la web semántica sobre la que se sustenta el resto de capas, es un metalenguaje que permite definir nuestros propios lenguajes de etiquetado validándolos mediante dtds (*document type definitions*) o *xml schemas*. Si bien es cierto que un etiquetado xml por sí solo no implica un mecanismo con el que poder expresar la semántica de los documentos (Cover, 1998), una migración progresiva de html a lenguajes de etiquetado xml —*Xhtml (extensible html)*, *SVG (scalable vector graphic)*, etc.— supondría un gran avance, principalmente por la capacidad de xml para utilizar espacios de nombre (*xmlNamespaces*) (Bray; Hollander; Layman, 1999), para diferenciar entre contenido y presentación —*XSL (extensible stylesheet language)*— y, fundamentalmente, para que la web pase de ser un espacio hiperdocumental a otro hiperconceptual gracias a tecnologías en desarrollo como *XLink (xml linking and addressing languages)* (DeRose; Maler; Orchard, 2001), (Martelli; Signore, 2002).

Rdf es un lenguaje de etiquetado creado mediante sintaxis xml que define un modelo de datos para describir recursos (cualquier objeto identificable por un uri), mediante enunciados o asertos en forma de tripletas sujeto-predicado-objeto, donde sujeto y predicado son uris y objeto puede ser un uri o un valor literal. Ya que rdf es un vocabulario xml, puede describir recursos de manera externa a éstos o de forma embebida, siempre y cuando tengan sintaxis xml.

En esta misma capa, *rdf schema* es un vocabulario rdf que nos permite describir recursos mediante una orientación a objetos similar a la de muchos lenguajes de programación como Java. Para ello, proporciona un mecanismo para definir clases, objetos y propiedades; relaciones entre clases y propiedades; así como restricciones de dominio y rango sobre las propiedades.

Mientras que la relación entre xml y *xml schema* es de control sintáctico, entre rdf y *rdf schema* es de naturaleza semántica: el esquema xml asegura que los elementos son usados correctamente, mientras que un esquema rdf certifica que, por ejemplo, en un aserto donde el sujeto es “carné de conducir” y el predicado es “nombre”, el objeto es el nombre de una persona y no el modelo de un coche (Berners-Lee, 1998).

A estos lenguajes de etiquetado hay que sumar propuestas como la llevada a cabo por el *Dcmi (Dublin core metadata initiative)*<sup>2</sup>, que ha definido un conjunto básico de elementos para la descripción de metadatos (Rachel; Wagner, 2002), tales como *title*, *creator* o *subject*, que pueden ser incrustados en cualquier documento con sintaxis xml gracias a los espacios de nombre. Aunque los elementos *Dublin core* preceden en el tiempo a rdf, actualmente existe una serialización de éstos en sintaxis rdf (Miller; Beckett; Brickley, 2002).

Como indican algunos autores (Ossenbruggen; Hardman; Rutledge, 2002), (Palmer, 2001), esta primera capa de la web semántica tiene soporte tecnológico en la actual web. Xml y rdf han alcanzado el nivel de recomendación W3C (Jacobs, 2001) y *rdf schema* se encuentra próximo a alcanzarlo, a lo que debemos sumar la gran cantidad de software existente para su edición e implementación.

En este sentido, el grupo de trabajo *Swad-Europe (Semantic web advanced development)*<sup>3</sup> tiene como objetivo poner de manifiesto, mediante ejemplos prácticos, cómo el conjunto de tecnologías que conforman la web semántica suponen una ventaja real para la actual web, resolviendo problemas en áreas como: tesauros, clasificaciones, agendas, búsqueda de recursos, etc. (Matthews; Wilson; Brickley, 2002).

### 3. Ontologías y agentes

#### 1. Ontologías.

Han sido tradicionalmente usadas como modelo de representación de conocimiento en inteligencia artificial. Según la definición ofrecida por Gruber (1993) y posteriormente extendida por Studer, Benjamins y Fensel (1998), es “una especificación explícita y formal de una conceptualización”. Más concretamente, estará formada por una taxonomía relacional de conceptos y por un conjunto de axiomas o reglas de infe-

rencia mediante los cuales se podrá inferir nuevo conocimiento.

La web, sin embargo, es mucho más extensa y dinámica que una base de conocimiento (Horrocks, 2002), por lo que la implementación de un lenguaje de definición de ontologías estándar y la posterior consistencia de dichas ontologías (Hendler, 2002), se presentan como un reto complicado, principalmente por la naturaleza descentralizada de la web semántica.

Con *rdfs* podemos describir jerarquías de clases, tales como ontologías simples, sobre las que poder realizar consultas y razonamiento automático. Aún así, *rdfs* no es lo suficientemente expresivo para representar ontologías de la complejidad que necesita la web semántica, ya que el agente inteligente sólo podría realizar la inferencia sobre la herencia de propiedades, pues *rdfs* no permite declarar axiomas. A esto habría que añadir que algunas características de *rdfs*, como la *reificación* (posibilidad de tratar un aserto como parte de otro, permitiendo de esta forma declarar unos asertos sobre otros), dificultan el razonamiento inferencial

**Swets & Zeitlinger Publishers**, editora de esta revista, ha encargado la distribución de todas sus publicaciones a la siguiente empresa del grupo **Swets & Zeitlinger**:

**Extenza-Turpin**, Blackhorse Road,  
Letchworth, SG6 1HN, Herts,  
Reino Unido.

Tel.: +44-146 267 2555;

fax: 146 248 0947

[subscriptions@turpinltd.com](mailto:subscriptions@turpinltd.com)

Rogamos a nuestros suscriptores que para solventar cualquier asunto administrativo se dirijan siempre directamente a **Extenza-Turpin**. Sin embargo recordamos que continúan en funcionamiento los (nuevos) números de teléfono de atención al suscriptor en Barcelona:

Tel.: +34-932 081 970;

fax: 932 081 971

(Silva Muñoz, 2001) (Ossenbruggen; Hardman; Rutledge, 2002).

Por esta razón, coincidimos con Patel-Schneider y Fensel (2002) cuando al referirse al esquema de la figura 1, afirman que Berners-Lee usa el término “vocabularios de ontologías” queriendo referirse a “lenguajes de ontologías”, ya que *rdfs* no se diseñó como solución final para la representación de conocimiento, sino como un núcleo para ser extendido (Silva Muñoz, 2001).

*Shoe (simple html ontology extensions)* (Luke; Spector; Rager, 1996), diseñado antes de que se acuñara el término web semántica, fue el primer lenguaje de etiquetado para definir ontologías desarrollado para su uso en la web. Tanto las ontologías como las anotaciones se expresan mediante etiquetas incrustadas en archivos html, aunque actualmente también posee una serialización en xml. Permite definir clases y relaciones entre ellas, así como reglas de inferencia expresadas en forma de cláusulas de Horn (Luke, 1996). Su principal carencia es que no dispone de mecanismos para expresar negaciones o disyunciones. Ha tenido diversas implementaciones prácticas como buscadores, *APIs Java* o editores de anotaciones<sup>4</sup>.

*OIL (ontology inference layer)* (Fensel et al., 2000), elaborado en el marco de los proyectos *On-To-Knowledge*<sup>5</sup> e *Ibrow*<sup>6</sup>, y derivado en parte de *Shoe*, es el primer lenguaje de representación de ontologías basado en estándares *W3C*: tiene sintaxis xml y está definido como una extensión de *rdfs*. El modelo utilizado para la representación del conocimiento lo ha heredado, por una parte de la lógica descriptiva (declaración de axiomas o reglas) y, por otra, de los sistemas basados en frames (taxonomía de clases y atributos). Se encuentra estructurado en varias capas de sublenguajes: la base, o núcleo, coincide plenamente con *rdfs*, a excepción de la *reificación*, y cada una de las capas superiores añade funcionalidad y complejidad a las subyacentes. Esto posibilita, teóricamente, la reutilización de agentes diseñados para *rdfs*, que podrían procesar ontologías *OIL* (Fensel, 2002).

Entre las limitaciones de *OIL* (Horrocks et al., 2000) podíamos destacar las siguientes: no ofrece la posibilidad de sobrescritura de valores heredados de una superclase; presenta falta de expresividad en la declaración de axiomas (reglas); y no soporta dominios concretos (ej.: números enteros, cadenas de caracteres, etc.).

*Daml+OIL* es una propuesta de lenguaje de representación de ontologías más reciente, fruto de la cooperación entre los grupos de trabajo de *OIL* y *Darpa (US Defense Advanced Research Projects Agency)*, quienes anteriormente habían desarrollado el lenguaje

*Daml* (*Darpa's agent markup language*) con la finalidad de extender el nivel de expresividad de *rdfs*. Por tanto, es un lenguaje que unifica estos dos anteriores, aunando esfuerzos en el camino de la construcción de un lenguaje estándar para la definición de ontologías. Aunque *Daml+OIL* hereda muchas de las características de *OIL*, difiere en algunas (Horrocks, 2000). De hecho, se aleja del modelo basado en frames de *OIL*, potenciando la lógica descriptiva (Horrocks; Goble; Bechhofer, 2001).

A nivel práctico, y como indican Horrocks, Goble y Bechhofer (2001), *Daml+OIL* demuestra ser más útil como soporte para ontologías que, por ejemplo, *rdf schema*, sin embargo aún tiene ciertas carencias como formato de intercambio y modelado de ontologías. Si bien la complejidad técnica en *Daml+OIL* no es un problema, lo que se puede demostrar por el gran número de herramientas y aplicaciones desarrolladas con soporte para este lenguaje<sup>7</sup> en sus escasos dos años de vida, su complejidad conceptual (dificultad de uso y aprendizaje por parte del autor de ontologías) sí lo es (Harmelen, 2002).

En julio de 2002, el grupo de trabajo *WebOnt* (*Web-Ontology*)<sup>8</sup> del W3C publica el primer borrador de *OWL* (*web ontology language*) (Harmelen; McGuinness, 2002), lenguaje derivado de *Daml+OIL* y cimentado sobre *rdfs*. Para solucionar el problema de la complejidad conceptual en *OWL*, Harmelen y Horrocks (2002) proponen que tenga una doble sintaxis: una legible por máquinas (basada en *rdf*) y otra por humanos (puede que basada en una *dtd* de *xml*). Ejemplo de trabajos en este sentido es una nueva notación llamada *N3* (Berners-Lee, 2001), una sintaxis alternativa que facilita la lectura y elaboración de documentos *rdf*.

No obstante, era de suponer que la relación entre *OWL* y *rdfs*, sería la misma que entre *rdfs* y *rdf* pero, como indican Patel-Schneider y Fensel (2002), *OWL* no puede mantener la sintaxis *rdf* y ser definido como extensión semántica de *rdfs* al mismo tiempo ya que daría lugar a paradojas semánticas.

El borrador de *OWL* incluye también *OWL Lite*, su versión reducida, cuyo objetivo es ser adoptada rápida y ampliamente por los desarrolladores, para lo que concentra las características más comúnmente usadas en *OWL* y *Daml+OIL*. En este sentido, Clark y Uschold (2002) proponen que *OWL* se estructure en capas que, al igual que en *OIL*, difieran en complejidad, para adaptarse a las necesidades de nivel de expresividad requerido por las diferentes comunidades de usuarios (autores de ontologías, programadores, etc.) y los diferentes tipos de aplicaciones (agentes, motores de búsqueda, etc.).

## 2. Agentes.

Son una entidad software que recoge, filtra y procesa información contenida en la web, realiza inferencias sobre dicha información e interactúa con el entorno sin necesidad de supervisión o control constante por parte del usuario. Estas tareas son realizadas en representación del usuario o de otro agente. Para Hendler (1999) las cualidades que ha de tener un agente inteligente son:

—Comunicativo. Debe entender las necesidades, objetivos y preferencias del usuario, ya que de dicha comunicación o interacción usuario-agente depende que éste pueda llevar a cabo su función de forma eficiente. Así mismo, debe poder comunicarse con el entorno mediante representaciones compartidas de conocimiento (ontologías).

—Autónomo. Además de comunicarse, debe poder interactuar con el entorno, tomando decisiones y actuando por sí solo, limitando sus acciones según el nivel de autonomía permitida por el usuario.

—Adaptable. Debe ser capaz de aprender del entorno: usuarios (preferencias), fuentes de información y otros agentes.

En este escenario, las características propias de un entorno abierto como la web, y por ende de la web semántica, dificultan la implementación de agentes inteligentes. Por ello, llegados a este punto cabe preguntarse qué sería necesario para el funcionamiento de agentes inteligentes en la web semántica. Por un lado, sería imprescindible un lenguaje estandarizado y compartido de definición de ontologías (*OWL* podría ser la mejor opción) y, por otro, sería conveniente la creación de ontologías y la anotación de recursos web en función de las mismas.

**«Ya que *rdf* es un vocabulario *xml*, puede describir recursos de manera externa a éstos o de forma embebida, siempre y cuando tengan sintaxis *xml*»**

Hasta que esto no suceda no se podrán crear agentes inteligentes interoperativos y mientras tanto, como es lógico, un usuario podría preguntarse qué ventajas tiene anotar sus páginas web mediante ontologías si no existen agentes que justifiquen este esfuerzo. Este fenómeno, denominado por Hendler (2001) como “el problema del huevo y la gallina”, es similar al que se dio en los primeros días de la web: ¿para qué publicar si había tan pocos usuarios? o, ¿para qué desarrollar navegadores si no había páginas web? (Palmer, 2001). La solución podría alcanzarse llevando a cabo aplicaciones informáticas de publicación web, que anotaran

automáticamente las páginas creadas por el usuario en función de ontologías ya existentes en la web, con la menor intervención posible del usuario en el proceso.

### 3. Algunos ejemplos.

Entre las actualmente escasas muestras de implementación de esta capa de la web semántica, destacamos dos productos de la investigación académica:

—*Retsina*<sup>9</sup> *semantic web calendar agent*<sup>10</sup> (**Payne; Singh; Sycara**, 2002) es un agente inteligente que organiza y gestiona la agenda del usuario, sobre programas o calendarios de eventos descritos mediante rdf y a través de su integración con gestores de información personal tales como *Microsoft Outlook*.

—*Smore* (*Semantic markup, ontology, and rdf editor*) (**Hendler et al.**, 2002) es un entorno software integrado para creación de contenidos web etiquetados semánticamente. Las descripciones semánticas son realizadas automáticamente mediante anotaciones rdf y su integración con ontologías. El entorno está compuesto por diferentes módulos, uno de los cuales, *PhotoSmore*, permite incluso la anotación semántica de áreas concretas de una imagen.

### 4. Firmas digitales y redes de “confianza”

Como indican **Berners-Lee, Hendler y Lassila** (2001), una de las características fundamentales de la web semántica serán las firmas digitales. Estos autores las definen como “bloques de datos cifrados que las computadoras y agentes podrán usar para verificar que la información adjunta ha sido proporcionada por una fuente fiable”. Al aplicar un algoritmo de “hashing” sobre un documento o mensaje a firmar, obtenemos un bloque de datos, llamado “hash” o “digest”, que representa su “resumen”. La firma digital se obtiene al cifrar dicho “hash” mediante un sistema de criptografía de clave pública, que se basa en el empleo de dos tipos de claves: pública y privada (todo lo que se cifre mediante una de carácter público sólo puede ser descifrado a través de su correspondiente privada y viceversa).

Las ventajas que confiere la firma digital a los recursos, documentos o mensajes de la web son principalmente: identificación (posibilidad de determinar la identidad del emisor o autor del recurso) e integridad (facilidad para detectar su manipulación o alteración ilícita). Ambas se comprueban aplicando el algoritmo de “hashing” sobre el recurso y comparando su resultado con el “hash” obtenido al descifrar la firma digital a través de la clave pública del autor (**Simon; Madson; Adams**, 2001).

El grupo de trabajo *Xml Signature*<sup>11</sup>, fruto del esfuerzo conjunto entre el W3C y el *Ietf* (*Internet Engi-*

*neering Task Force*)<sup>12</sup>, ha desarrollado una especificación con sintaxis xml para la representación de firmas digitales en recursos web (**Eastlake; Reagle; Solo**, 2002), que alcanzó el grado de recomendación W3C en febrero de 2002. Este grupo, ofrece desde su web enlaces a multitud de herramientas y bibliotecas de programación para el procesado y verificación de firmas digitales definidas bajo esta especificación.

Sin embargo, aunque resultan de utilidad para describir la autenticidad de las relaciones de autoría (predicados) entre individuos y recursos, no describen la confianza de los usuarios y agentes inteligentes sobre dichos autores y, por lo tanto, sobre el contenido de sus recursos (**Hendler; Golbeck; Parsia**, 2002). A esto hay que añadir que la infraestructura de clave pública (PKI) utilizada en las firmas digitales tiene un carácter centralizado. De hecho, la única forma de asegurar la validez del vínculo entre el titular de la clave pública y ella misma es utilizar un certificado emitido por una autoridad de certificación. En este sentido, no debemos olvidar que la web semántica pretende tener un carácter completamente descentralizado. La solución podría ser la denominada “web of trust”, que basa su funcionamiento en el concepto de confianza o fiabilidad, el cual es definido por **Reagle** (2002) como “el grado en que un agente considera un aserto como verdadero para un contexto dado”.

Esta capa, ubicada en la zona más alta del esquema de la figura 1, es una aplicación de la investigación sobre redes sociales a las redes de recursos (uris), expresados en forma de tripletas rdf de la web semántica.

Los individuos, identificados por medio de uris, expresarían su grado de confianza (o desconfianza) sobre otros, que a su vez harían lo mismo sobre los demás, lo que daría como resultado extensas e interoperables redes de confianza procesables por agentes inteligentes.

**Hendler, Golbeck y Parsia** (2002) muestran de forma práctica cómo las redes sociales pueden ser aplicadas para construir la llamada “web of trust”. Toman como base *Foaf* (*Friend-of-a-friend*)<sup>13</sup>, un vocabulario rdf para describir información sobre los individuos y sus relaciones en una red social, el cual extienden con el objetivo de poder expresar el nivel de confianza entre los individuos de la Red para, finalmente, mostrar su implementación práctica en áreas como el correo electrónico.

El uso conjunto de firmas digitales y redes de confianza, aplicadas sobre reglas de inferencia de la capa lógica, puede proporcionar mecanismos complejos de autenticación de usuarios y validación de permisos de acceso a determinados recursos por parte de dichos usuarios.

En opinión de **Reagle** (2002), el problema de la centralización de la *PKI* puede ser resuelto mediante el uso generalizado de redes de confianza y bloques de datos “hash” que harían innecesario el empleo de criptografía en la mayoría de los casos y, consecuentemente, de las autoridades de certificación.

## 5. Conclusiones

El proyecto de web semántica, en su organización y funcionamiento, tiene bastantes similitudes con la organización de una biblioteca, donde la semántica de las obras (recursos) se describe mediante metadatos (*rdf*, *rdfs*, *DC*, etc.), los usuarios pueden ayudarse de bibliotecarios (agentes inteligentes) así como del conocimiento y capacidad de razonamiento de éstos (ontologías) para localizar y recuperar la información que necesitan.

La web, vista de esta forma, sería la más extensa y heterogénea biblioteca existente y justificaría que la única forma eficiente de recuperar información fuera mediante agentes computacionales y no humanos. Sin embargo, el problema de estos agentes es doble: todo el conocimiento que poseen (ontologías) debe ser definido por usuarios y autores y, además, con ese conocimiento únicamente podrían ofrecer un servicio de consulta, ya que no son capaces de describir recursos por sí solos, por lo que éstos también tendrán que ser descritos por usuarios y autores. Esta descripción de fuentes y el conocimiento de los agentes, realizados de forma descentralizada y cooperativa, es la que previsiblemente provocará mayores problemas en la futura implementación de la web semántica. No obstante, no es posible negar que además de en la recuperación de información, con este nuevo modelo se abren prometedores horizontes en los servicios web, en las bibliotecas digitales o en la computación ubicua.

Por otra parte, es necesario apuntar los problemas de recuperación de información a través de metadatos en un entorno abierto y descentralizado como la web semántica. La falta de normalización en la descripción de recursos podrá llevar a que cada usuario, por ejemplo, describa el nombre de un autor de forma totalmente diferente al resto. Hay que recordar también el fracaso de uno de los primeros intentos de describir metadatos sobre los recursos en la web a través de etiquetas *<meta>* incrustadas en los ficheros html. Muy pronto los motores de búsqueda dejaron de utilizarlas debido principalmente al uso interesado e ilícito que muchos websites hicieron de ellas como técnica para posicionarse en sus resultados por encima de los recursos de la competencia.

En un nivel más técnico, otro problema al que se debe de enfrentar es al propio esquema de capas ideado por **Berners-Lee**. En éste, cada nivel posee sintaxis

y, al menos la mayoría, semántica. El problema es que no todas extienden de la misma forma la semántica, ni mantienen la sintaxis de su capa subyacente. *Rdf* es un subconjunto sintáctico de *xml* pero no mantiene su semántica, por lo que, como ya se ha mencionado, *xml* es considerado la base sintáctica de la web semántica. *Rdfs*, en cambio, no sólo mantiene la sintaxis de *rdf* sino que además extiende su semántica. En este sentido, es importante desarrollar relaciones sintáctico-semánticas alternativas para solucionar el problema de ubicación de *OWL* sobre *rdfs*.

Por otra parte, es totalmente necesario intentar mantener un equilibrio entre facilidad de uso y capacidad expresiva, ya que no podemos olvidar que la simplicidad y accesibilidad de *html* fueron las características clave que contribuyeron al surgimiento de la web o, que fue esta armonía entre sencillez y funcionalidad la que hizo que *xml* se impusiera entre los desarrolladores, algo que en su día no consiguió *sgml* debido a su complejidad conceptual.

**«Un usuario podría preguntarse qué ventajas tiene anotar sus páginas web mediante ontologías si no existen agentes que justifiquen este esfuerzo»**

A pesar de los problemas a los que se enfrenta la web semántica, no dudamos de la viabilidad del proyecto, avalada por la utilidad real que están demostrando a día de hoy el conjunto de lenguajes, tecnologías y herramientas afines sobre la actual web y por el extenso y creciente número de investigadores, grupos de trabajo y grandes empresas que se encuentran implicados en este proyecto.

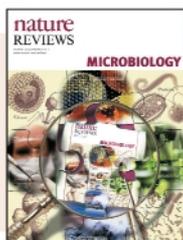
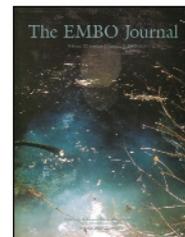
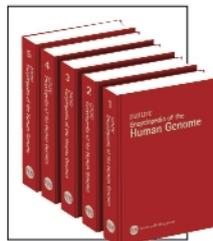
## Notas

1. <http://www.w3.org/2001/03/identification-problem/>
2. <http://www.dublincore.org>
3. <http://www.w3.org/2001/sw/Europe/>
4. <http://www.cs.umd.edu/projects/plus/SHOE/>
5. <http://www.ontoknowledge.org>
6. <http://www.swi.psy.uva.nl/projects/ibrow/home.html>
7. <http://www.daml.org/tools>
8. <http://www.w3.org/2001/sw/WebOnt/>
9. *Reusable environment for task-structured intelligent networked agents.*
10. <http://www.daml.ri.cmu.edu/Call>
11. <http://www.w3.org/Signature/>
12. <http://www.ietf.org>
13. <http://rdfweb.org/foaf/>

# Nature Publishing Group

Nature Publishing Group (NPG) es una de las más importantes editoriales de información científica. Se puede acceder a nuestro portfolio de revistas y obras de referencia de alta calidad por medio de suscripciones en papel u online.

Infórmese de los nuevos títulos para añadir en su colección:



**Nature Reviews  
Microbiology**  
Octubre 2003

ISSN 1740-1526  
Volumen 1  
12 ediciones por año

**Nature Encyclopedia  
of the Human Genome**  
Junio 2003

ISSN 0-333-80386-8  
5 volúmenes edición  
impresa

**EMBO Journal**  
Enero 2004  
Impacto 13.999\*

ISSN 0261-4189  
Volumen 23  
24 ediciones por año

Para información sobre suscripciones de revistas y acceso online por medio de licencias visite la Librarian Gateway:

[npg.nature.com/libraries](http://npg.nature.com/libraries)

Para más información contacte con Lucy Dunkerton.

Tel: +44 (0) 20 7843 4621  
Fax: +44 (0) 20 7843 4921  
E-mail: [l.dunkerton@nature.com](mailto:l.dunkerton@nature.com)  
Web: [npg.nature.com/libraries](http://npg.nature.com/libraries)

\* ISI, PA, USA

nature publishing group



## Bibliografía

- Berners-Lee, T.** *Information management: a proposal*, 1989. Consultado en: 10-11-02.  
<http://www.w3.org/History/1989/proposal.html>
- Berners-Lee, T.** *Semantic web road map*, 1998. Consultado en: 11-11-02.  
<http://www.w3.org/DesignIssues/Semantic.html>
- Berners-Lee, T.**, et al. *Request for comments: 2396, uniform resource identifiers*, 1998. Consultado en: 12-11-02.  
<http://www.ietf.org/rfc/rfc2396.txt>
- Berners-Lee, T.** *Notation 3*, 2001. Consultado en: 26-11-02.  
<http://www.w3.org/DesignIssues/Notation3>
- Berners-Lee, T.; Hendler, J.; Lassila, O.** "The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities". En: *Scientific American*, 2001, mayo.
- Bray, T.; Hollander, D.; Layman, A.** *Namespaces in xml*, 1999. Consultado en: 16-11-02.  
<http://www.w3.org/TR/REC-xml-names>
- Brickley, D.; Guha, R. V.** *Resource description framework (rdf) schema specification 1.0: W3C working draft*, 2002. Consultado en: 13-11-02.  
<http://www.w3.org/TR/rdf-schema>
- Brooks, T. A.** "The semantic web, universalist ambition and some lessons from librarianship". En: *Information research*, 2002, abril, v. 7, n. 4.
- Clark, P.; Uschold, M.** "The many faces of the semantic web". En: *Ieee intelligent systems*, 2002, marzo-abril, pp. 72-73.
- Cover, Robin.** *Xml and semantic transparency*, 1998. Consultado en: 03-11-02.  
<http://xml.coverpages.org/xmlAndSemantics.html>
- DeRose, S; Maler, E.; Orchard, D.** *Xml linking language (Xlink): W3C recommendation*, 2001. Consultado en: 25-11-02.  
<http://www.w3.org/TR/xlink/>
- Eastlake, D.; Reagle, J.; Solo, D.** *Xml-signature syntax and processing: W3C recommendation*, 2002. Consultado en: 29-11-02.  
<http://www.w3.org/TR/xmlsig-core/>
- Fensel, D.** *Language standardization for the semantic web: the long way from OIL to OWL*, 2002. Consultado en: 12-11-02.  
<http://informatik.uibk.ac.at/users/c703851ftp/paper/dwc.pdf>
- Fensel, D.**, et al. *OIL in a Nutshell*, 2000. Consultado en: 17-11-02.  
<http://www.cs.vu.nl/~ontoknow/oil/download/oilnutshell.pdf>
- Gruber, T. R.** "Toward principles for the design of ontologies used for knowledge sharing". En: *Formal ontology in conceptual analysis and knowledge representation*. The Netherlands: Kluwer Academic Publishers, 1993.
- Harmelen, Frank van.** "The complexity of the web ontology language". En: *Ieee intelligent systems*, 2002, marzo-abril, pp. 71-72.
- Harmelen, F.; Horrocks, I.** *OWL Language features: general line of attack & a first sketch*, 2002. Consultado en: 24-11-02.  
<http://www.cs.vu.nl/~frankh/spool/OWL-first-sketch.html>
- Harmelen, F.; McGuinness, D.** *Feature synopsis for OWL lite and OWL: W3C working draft*, 2002. Consultado en: 21-11-02.  
<http://www.w3.org/TR/owl-features/>
- Heflin, J.; Volz, R.; Dale, J.** *Requirements for a web ontology language: W3C working draft*, 2002. Consultado en: 17-11-02.  
<http://www.w3.org/TR/webont-req/>

**Hendler, James.** *Is there an intelligent agent in your future?*, 1999. Consultado en: 23-11-02.

<http://www.nature.com/nature/webmatters/agents/agents.html>

**Hendler, James.** "Agents and the semantic web". En: *Ieee intelligent systems*, 2001, marzo/abril, pp. 30-37.

**Hendler, James.** "Ontologies on the semantic web". En: *Ieee intelligent systems*, 2002, marzo-abril, pp. 73-74

**Hendler, J., et al.** *Smore – semantic markup, ontology, and rdf editor*, 2002. Consultado en: 26-11-02.

<http://www.mindswap.org/papers/SMORE.pdf>

**Hendler, J.; Golbeck, J.; Parsia, B.** *Trust networks on the semantic web*, 2002. Consultado en: 28-11-02.

<http://www.mindswap.org/papers/Trust.pdf>

**Horrocks, I.** *Remaining differences between Daml-OIL and OIL*, 2000. Consultado en: 25-11-02.

<http://www.cs.man.ac.uk/~horrocks/DAML-OIL/differences.html>

**Horrocks, I.** "An ontology language for the semantic web". En: *Ieee intelligent systems*, 2002, marzo-abril, pp. 74-75.

**Horrocks, I., et al.** *The ontology inference layer OIL*, 2000. Consultado en: 20-11-02.

<http://www.ontoknowledge.org/oil/TR/oil.long.html>

**Horrocks, I.; Goble, C.; Bechhofer, S.** "Daml+OIL is not enough". En: *International semantic web working symposium (Swws)*, 2001.

**Jacobs, I.** *World wide web consortium process document*, 2001. Consultado en: 05-11-02.

<http://www.w3.org/Consortium/Process-20010719/>

**Lassila, O.; Swick, R.** *Resource description framework (rdf) model and syntax specification: W3C recommendation*, 1999. Consultado en: 13-11-02.

<http://www.w3.org/TR/REC-rdf-syntax>

**Lu, S.; Dong, M.; Fotouhi, F.** "The semantic web: opportunities and challenges for next-generation web applications". En: *Information research*, 2002, abril, v. 7, n. 4.

**Luke, Sean.** *Inferences in Shoe ontologies*, 1996. Consultado en: 18-11-02.

<http://www.cs.umd.edu/projects/plus/SHOE/inferences.html>

**Luke, S.; Spector, L.; Rager, D.** *Ontology-based knowledge discovery on the world-wide web*, 1996. Consultado en: 18-11-02.

<http://www.cs.umd.edu/projects/plus/SHOE/pubs/aaai-paper.html>

**Martelli, S.; Signore, O.** "Semantic characterisation of links and documents". En: *Ercim news*, 2002, octubre, n. 51, pp. 29-30.

**Matthews, B.; Wilson, M.; Brickley, D.** "Semantic web advanced development in Europe". En: *Ercim news*, 2002, octubre, n. 51, pp. 15-16.

**Miller, E.; Beckett, D.; Brickley, D.** *Expressing simple Dublin core in rdf/xml*, 2002. Consultado en: 26-11-02.

<http://dublincore.org/documents/dcmes-xml/>

**Ossenbruggen, J.; Hardman, L.; Rutledge, L.** "Hypermedia and the semantic web: a research agenda". En: *Journal of digital information*, 2002, mayo, v. 3, n. 1.

**Palmer, Sean B.** *The semantic web: an introduction*, 2001. Consultado en: 16-11-02.

<http://infomesh.net/2001/swintro/>

**Patel-Schneider, P.; Fensel, D.** "Layering the semantic web: problems and directions". En: *First international semantic web conference (Iswc)*, 2002, pp. 16-29.

**Payne, T. R.; Singh, R.; Sycara, K.** "Calendar agents on the semantic web". En: *Ieee intelligent systems*, 2002, mayo/junio, v. 17, n. 3, pp. 84-86.

**Rachel, H.; Wagner, H.** "A metadata registry for the semantic web". En: *D-Lib magazine*, 2002, mayo, v. 8, n. 5.

**Reagle, Joseph M.** *Key free trust in the semantic web: finding Bacon's key*, 2002. Consultado en: 23-11-02.

<http://www.w3.org/2002/03/key-free-trust>

**Silva Muñoz, Lidia.** *Representación de ontologías en la web semántica*. Consultado en: 19-11-02.

[http://www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo\\_lydia.pdf](http://www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo_lydia.pdf)

**Simon, E.; Madsen, P.; Adams, C.** *An introduction to xml digital signatures*, 2001. Consultado en: 22-11-02.

<http://www.xml.com/lpt/a/2001/08/08/xmlsig.html>

**Studer, R.; Benjamins, R.; Fensel, D.** "Knowledge engineering: principles and methods". En: *Data and knowledge engineering*, 1998, v. 25, n. 1-2, pp. 161-197.

**E. Peis; Y. Hassan**, Universidad de Granada, Facultad de Biblioteconomía y Documentación, Departamento de Biblioteconomía y Documentación  
[epeis@ugr.es](mailto:epeis@ugr.es)  
[yusef@nosolousabilidad.com](mailto:yusef@nosolousabilidad.com)

**E. Herrera-Viedma; J. C. Herrera**, Universidad de Granada, Departamento de Ciencias de la Computación e I. A.  
[viedma@decsai.ugr.es](mailto:viedma@decsai.ugr.es)  
[jcht@correo.de](mailto:jcht@correo.de)

## Próximos temas especiales

Noviembre 2003

**Gestión de documentación multimedia**

Enero 2004

**Minería de la web (web mining)**

Marzo 2004

**Bibliotecas digitales**

Los interesados pueden remitir notas, artículos, propuestas, publicidad, comentarios, etc., sobre estos temas a:

[epi@sarenet.es](mailto:epi@sarenet.es)