

Ontologías, metadatos y agentes: recuperación “semántica” de la información

Peis Redondo, Eduardo, Hassan Montero, Yusef
Dept. Biblioteconomía y Documentación
Fac. Biblioteconomía y Documentación.
Campus de la Cartuja, Universidad de Granada – España.
epeis@ugr.es

Herrera Viedma, Enrique, Herrera, Juan Carlos
Dept. Ciencias de la Computación e Inteligencia Artificial
Fac. Biblioteconomía y Documentación
Campus de la Cartuja, Universidad de Granada – España.
eviedma@ugr.es

A pesar de que la Web permite el acceso a una impensable cantidad de información, no constituye, ni mucho menos, un mecanismo de localización perfecto: los motores de búsqueda realizan una labor útil, pero puede afirmarse que están llegando al límite de sus posibilidades; la sobrecarga informativa produce la sensación de colapso; está comprobado que las técnicas léxico-estadísticas, por sí solas, no pueden solucionar la problemática de la recuperación de la información en la Web...

En el presente trabajo describimos la propuesta denominada “Web Semántica” - extensión de la actual Web que pretende que no sólo los humanos, sino también las máquinas, sean capaces de “comprender” el significado de los hipertextos- con el fin de analizar su viabilidad y las ventajas que implicaría en la Recuperación de Información en la Web.

Palabras clave: Web Semántica; Ontologías; Agentes Inteligentes; Redes de Confianza;

1. INTRODUCCIÓN

A pesar de que el surgimiento de la Web ha supuesto un cambio radical en cuanto a la facilidad de acceso y difusión de la información electrónica, el gigantesco número de hipertextos y su crecimiento exponencial, sumados a su falta de estructuración lógica, está dando lugar a graves problemas en la recuperación de información en la Web.

Los actuales buscadores, basados en indexación humana en el caso de los índices de búsqueda, e indexación automática en el caso de los motores de búsqueda, son incapaces de ofrecer tasas de exhaustividad y precisión realmente aceptables en sus resultados.

El problema de la precisión en la recuperación de información puede ser visto como consecuencia de la falta de significado o semántica que para los ordenadores tienen los documentos Web, en su amplia mayoría formateados mediante HTML - lenguaje de

etiquetado que únicamente es capaz de expresar la forma de presentación (maquetación, colores, tipografía...) de los contenidos -.

En el presente trabajo describimos la propuesta denominada “Web Semántica” (**Berners-Lee, Hendler, Lassila**; 2001), con el fin de analizar su viabilidad y las ventajas que implicaría en la Recuperación de Información en la Web. La estructura seguida en el trabajo es: (2) Introducción al proyecto Web Semántica; (3) Descripción de las tecnologías y lenguajes implicados; (4) Discusión acerca de la viabilidad del proyecto; y (5) Conclusiones.

2. WEB SEMÁNTICA

La propuesta inicial de World Wide Web que **Berners-Lee** (1989) ideó, ya apuntaba a lo que posteriormente se ha denominado “Web Semántica”, proyecto que podemos definir como una extensión de la actual Web cuyo objetivo es que no sólo los humanos, sino

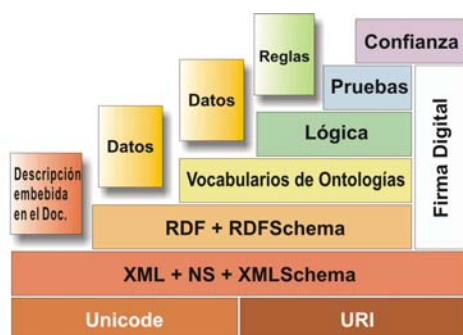
también las máquinas, sean capaces de “comprender” el contenido de los documentos.

Para que este objetivo sea posible es necesario proveer de mecanismos y tecnologías con las que poder definir la semántica de los hiperdocumentos, metainformación que posteriormente podrá ser utilizada por agentes y buscadores inteligentes con el fin de ofrecer resultados precisos y contextualizados.

La infraestructura de tecnologías y lenguajes necesaria para la implementación de la Web Semántica se puede esquematizar en varias capas o niveles (**Berners-Lee; 1998**):

- 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 los recursos, 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;
- 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.



Semantic Web “layer cake” (Berners-Lee, XML Conference, 2002)

Figura 1

3. TECNOLOGÍAS Y LENGUAJES

3.1. URI

Los URI -Uniform Resource Identifier- (**Berners-Lee et al.; 1998**), cuyo subconjunto más conocido son los URL -Uniform Resource Locators-, proporcionan el mecanismo para identificar de forma inequívoca cualquier recurso en la red: artículos, imágenes, sonidos, etc. Con la Web Semántica, los URIs cumplirán además con la función de identificadores de objetos del mundo real. Cualquier objeto podrá ser identificado mediante un URI: nuestro microondas tendrá un URI asociado, el URI de nuestra web personal o de nuestra dirección e-mail nos identificará a nosotros, la función que realizamos en nuestro trabajo se expresará mediante un URI...

Los problemas que pudieran surgir en la identificación de objetos – por ejemplo, un mismo URI podría identificar dos objetos distintas - están intentando ser resueltos por el grupo de trabajo que el W3C ha designado para tal fin¹.

3.2. XML

En un nivel superior nos encontramos con los documentos y su estructuración lógica. XML –eXtensible Markup Language-, base sintáctica de la Web Semántica sobre la que se sustentan el resto de capas, es un metalenguaje que nos 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 representa 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-...) 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 hiperdokumental a un espacio hiperconceptual gracias a tecnologías en desarrollo como Xlink –XML Linking and Addressing Languages- (**DeRose, Maler, Orchard; 2001**) (**Martelli, Signore; 2002**).

3.3. RDF y RDF Schema

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 forma externa a éstos o de forma embebida, siempre y cuando estos recursos tengan sintaxis XML.

Por otro lado, 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; y, restricciones de dominio y rango sobre las propiedades.

Mientras que la relación entre XML y XML Schema es una relación de control sintáctico, la relación entre RDF y RDFS es de control semántico: el esquema XML asegura que los elementos son usados correctamente, mientras que un esquema RDF asegura 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).

3.4. Ontologías

Las 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), una Ontología es “una especificación explícita y formal de una conceptualización”. Más concretamente, una Ontología estará formada por una taxonomía relacional de conceptos y por un conjunto de axiomas o reglas de inferencia 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 asertos sobre otros asertos-, dificultan el razonamiento inferencial (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 clases, 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. SHOE tiene varias implementaciones prácticas, como buscadores, APIs Java o editores de anotaciones².

OIL -Ontology Inference Layer- (Fensel et al.; 2000), desarrollado en el marco del proyecto On-To-Knowledge³ y el proyecto IBROW⁴ 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 por OIL 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).

OIL se encuentra estructurado en varias capas de sublenguajes. La capa base o núcleo de OIL coincide plenamente con RDFS, a excepción de la reificación, y cada una de las capas superiores añade funcionalidad y complejidad a su capa subyacente. 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 DAML+OIL es un lenguaje que unifica estos dos lenguajes, 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 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 lenguaje⁵ 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)⁶ 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 legible por humanos (puede que basada en un 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.

El borrador de OWL incluye también OWL Lite, versión reducida de OWL, cuyo objetivo es ser adoptada rápida y ampliamente por los desarrolladores, para lo que concentra las características más comúnmente usadas de 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.).

3.5. Agentes inteligentes

Un Agente es 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 debe tener un agente inteligente son:

- Comunicativo. El agente debe entender las necesidades, objetivos y preferencias del usuario, ya que de dicha comunicación o interacción usuario-agente depende que el agente pueda llevar a cabo su función de forma eficiente. Así mismo debe poder comunicarse con el entorno mediante representaciones compartidas de conocimiento (ontologías).
- Autónomo. El agente, 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. El agente debe ser capaz de aprender del entorno: usuarios (preferencias), fuentes de información y otros agentes.

El papel del agente inteligente en el proceso de recuperación “semántica” de información no debe confundirse con la de un buscador inteligente. Un buscador inteligente se aprovechará del enriquecimiento semántico de los recursos web para mejorar (principalmente en la precisión) la recuperación de información, aunque su funcionamiento se basará, como los actuales buscadores, en la previa indización de todos aquellos recursos susceptibles de ser recuperados.

En cambio, un agente inteligente recorrerá la Web a través de los enlaces entre recursos (hiperdocumentos, ontologías, ...) en busca de aquella información que le sea solicitada, pudiendo además interactuar con el entorno para el cumplimiento de tareas encomendadas. Por ejemplo, un agente inteligente, ante una consulta dada, podría consultar autónomamente un buscador, y a partir de sus resultados, explorar la Web hasta encontrar la información solicitada, pudiendo finalmente llevar a cabo una acción sobre dicho recurso, como podría ser la reserva de una habitación en un hotel.

3.6. Firmas Digitales y “redes de confianza”

Como indican **Beners-Lee, Hendler y Lassila** (2001), una de las características fundamentales de la Web Semántica serán las firmas digitales. Estos autores definen las firmas digitales 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 un “resumen” del documento o mensaje. 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 clave pública solo puede ser descifrado a través de su correspondiente clave 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 la manipulación o alteración ilícita del recurso). La identificación y la integridad 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, Madsen, Adams**; 2001).

El grupo de trabajo XML Signature⁷, fruto del esfuerzo conjunto entre el W3C y el IETF (Internet Engineering Task Force)⁸, 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 de trabajo 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 las firmas digitales 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 la 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 dicha clave 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”. Este concepto 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 individuos, que a su vez harían lo mismo sobre otros, 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)⁹, un vocabulario RDF para describir información sobre los individuos y sus relaciones en una red social, el cual extienden con el objetivo de posibilitar 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.

4. DISCUSIÓN

Como indican algunos autores (**Ossenbruggen, Hardman, Rutledge**; 2002)(**Palmer**; 2001), las primeras capas de la Web Semántica (URIs, XML, RDFS) tienen soporte tecnológico en la actual Web. La identificación de recursos en la actual Web se basa en los URIs, XML y RDF han alcanzado el nivel de Recomendación W3C (**Jacobs**; 2001), y RDF Schema se encuentra próximo a alcanzar dicho nivel, 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)¹⁰ tiene como objetivo poner de manifiesto, mediante ejemplos prácticos, cómo este conjunto de tecnologías suponen una ventaja real para la actual Web, resolviendo problemas en áreas como: tesauros, clasificaciones, agendas, búsqueda de recursos, etc. (**Matthews, Wilson, Brickley**; 2002)

En cambio, aunque OWL puede posicionarse como el lenguaje estándar para la definición de ontologías, todavía se encuentra en versión borrador W3C, y el resto de propuestas, aunque demostrada su viabilidad técnica, como ya hemos comentado anteriormente, su complejidad conceptual las alejan de ser ampliamente aceptadas por la comunidad de desarrolladores web.

A un nivel más conceptual, otro problema al que debe enfrentarse el proyecto de Web Semántica es al propio esquema de capas ideado por Berners-Lee. En éste, cada capa 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 la semántica de XML, por lo que, como ya se ha mencionado en este trabajo, 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. Aunque según este esquema, era de suponer que la relación entre OWL y RDF(S) sería la misma que entre RDFS y RDF, como indican **Patel-Schneider y Fensel** (2002), OWL no puede mantener la sintaxis RDF y ser definido como extensión semántica de RDF(S) al mismo tiempo ya que

daría lugar a paradojas semánticas. En este sentido, es importante desarrollar relaciones sintáctico-semánticas alternativas para solucionar el problema de ubicación de OWL sobre RDF(S).

Para el funcionamiento de agentes inteligentes no solo sería imprescindible la consolidación de un lenguaje estandarizado y compartido de definición de ontologías, sino también, la creación de ontologías y la anotación de los recursos web en función de dichas ontologías.

Hasta que esto no suceda no se podrán desarrollar agentes inteligentes interoperativos, y mientras tanto, como es lógico, un desarrollador 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 desarrollando 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. Así mismo, la utilización de técnicas de Web Mining, - como indican **Berent, Hotho** y **Stumme** (2002) – podría disminuir considerablemente el esfuerzo necesario para la elaboración de ontologías, semiautomatizando el proceso.

Por otro lado, otro de los problemas a resolver es la fiabilidad en la descripción de los recursos, que será realizada de forma descentralizada y cooperativa. No podemos olvidar el fracaso de uno de los primeros intentos de describir metadatos de esta forma sobre los recursos en la Web, a través de etiquetas <meta> incrustadas en los ficheros HTML. Muy pronto los motores de búsqueda dejaron de considerar estas etiquetas en la indexación de los recursos web 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.

Quizá la solución a este problema podrían ser las comentadas “redes de confianza”, donde serían los mismos usuarios los que penalizarían a aquellos sitios web que “engañaran” en la descripción semántica de sus contenidos.

5. CONCLUSIONES

Una vez analizada la propuesta “Web Semántica” cabe cuestionarse qué mejoras supondría este modelo de Web en la recuperación de información. Queda claro que cuanto mayor sea la descripción semántica y estructuración lógica de los hipertextos, mayor será la precisión obtenida en la recuperación de información.

Igualmente, la exhaustividad se vería mejorada gracias a que los axiomas definidos en las ontologías, posibilitarán a agentes, así como buscadores inteligentes, inferir nuevo conocimiento no explicitado previamente, recuperando un mayor número de recursos relevantes para una consulta dada.

Además, los actuales buscadores Web tienen una cobertura que se estima bastante baja, ya que no tienen indexados todos los recursos existentes en la red. Un agente inteligente, ya que realiza la búsqueda de información en tiempo real explorando e interactuando con el entorno, puede llegar a recuperar recursos no indexados a priori, lo que supondría un avance en la cobertura de los recursos posibles a recuperar, aunque no debemos olvidar que el tiempo necesario para la tarea sería mayor.

A pesar de los problemas a los que se enfrenta la Web Semántica, discutidos en el anterior apartado, 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 a la Web Semántica, sobre la actual Web; por el extenso y creciente número de investigadores, grupos de trabajo, y grandes empresas, que se encuentran implicados en este proyecto; y por las ventajas y beneficios que supondrá la “Web Semántica”, entre otras, en áreas como la recuperación de información.

BIBLIOGRAFÍA

Berendt, B.; Hotho, A.; Stumme, G. “Towards Semantic Web Mining”. En: *First International Semantic Web Conference (ISWC)*, 2002, pp.264-278.

Berners-Lee, Tim. (1989). Information Management: A Proposal. Consultado en: 10-11-2002.
<http://www.w3.org/History/1989/proposal.html>

Berners-Lee, Tim. (1998). Semantic Web Road map. Consultado en: 11-11-2002.
<http://www.w3.org/DesignIssues/Semantic.html>

Berners-Lee, T. et al. (1998). Request for Comments: 2396, Uniform Resource Identifiers. Consultado en: 12-11-2002.
<http://www.ietf.org/rfc/rfc2396.txt>

Berners-Lee, Tim. (2001). Notation 3. Consultado en: 26-11-2002.
<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.

Berners-Lee, T.; Miller, E. “The Semantic Web lifts off”. En: *ERCIM News*, 2002, Octubre, n.51, pp.9-11.

Bray, T.; Hollander, D.; Layman, A.(1999). Namespaces in XML. Consultado en: 16-11-2002.
<http://www.w3.org/TR/REC-xml-names>

Brickley, D.; Guha, R.V. (2002). “Resource Description Framework (RDF) Schema Specification 1.0: W3C Working Draft”. Consultado en: 13-11-2002.
<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. (1998). XML and Semantic Transparency. Consultado en: 03-11-2002.
<http://xml.coverpages.org/xmlAndSemantics.html>

DeRose, S; Maler, E.; Orchard, D.(2001). XML Linking Language (XLink): W3C Recommendation. Consultado en: 25-11-2002.
<http://www.w3.org/TR/xlink/>

Eastlake, D.; Reagle, J.; Solo, D. (2002). XML-Signature Syntax and Processing: W3C Recommendation. Consultado en: 29-11-2002.
<http://www.w3.org/TR/xmlsig-core/>

Fensel, D. (2002). Language standardization for the Semantic Web: The long way from OIL to OWL. Consultado en: 12-11-2002.
<http://informatik.uibk.ac.at/users/c70385/ftp/paper/dwc.pdf>

Fensel, D. et al. (2000). OIL in a Nutshell. Consultado en: 17-11-2002.
www.cs.vu.nl/~ontoknow/oil/down/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. (2002). OWL Language features: general line of attack & a first sketch. Consultado en: 24-11-2002.
<http://www.cs.vu.nl/~frankh/spool/OWL-first-sketch.html>

Harmelen, F.; McGuinness, D. (2002). Feature Synopsis for OWL Lite and OWL: W3C Working Draft. Consultado en: 21-11-2002.
<http://www.w3.org/TR/owl-features/>

Heflin, J; Volz, R; Dale, J. (2002). Requirements for a Web Ontology Language: W3C Working Draft. Consultado en: 17-11-2002.
<http://www.w3.org/TR/webont-req/>

Hendler, James. (1999). Is There an Intelligent Agent in Your Future?. Consultado en: 23-11-2002.

<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. (2002). SMORE – Semantic Markup, Ontology, and RDF Editor. Consultado en: 26-11-2002.

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

Hendler, J.; Golbeck, J.; Parsia, B. (2002). Trust Networks on the Semantic Web. Consultado en: 28-11-2002.

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

Horrocks, Ian. (2000). Remaining differences between DAML-OIL and OIL. Consultado en: 25-11-2002.

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

Horrocks, Ian. “An Ontology language for the Semantic Web”. En: *IEEE Intelligent Systems*, 2002, marzo-abril, pp. 74-75.

Horrocks, I. et al. (2000). The Ontology Inference Layer OIL. Consultado en: 20-11-2002.

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, Ian. (2001). World Wide Web Consortium Process Document. Consultado en: 05-11-2002.

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

Lassila, O.; Swick, R. (1999). “Resource Description Framework (RDF) Model and Syntax Specification: W3C Recommendation”. Consultado en: 13-11-2002.

<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. (1996) Inferences in SHOE Ontologies. Consultado en: 18-11-2002.

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

Luke, S.; Spector, L.; Rager, D. (1996). Ontology-Based Knowledge Discovery on the World-Wide Web. Consultado en: 18-11-2002.

<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. (2002). Expressing Simple Dublin Core in RDF/XML. Consultado en: 26-11-2002.

<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. (2001). The Semantic Web: An introduction. 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.

Patel-Schneider, P; Siméon, J. “Building the Semantic Web on XML”. En: *First International Semantic Web Conference (ISWC)*, 2002, pp.147-161.

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 Metada Registry for the Semantic Web”. En: *D-Lib Magazine*, 2002, Mayo, v.8, n.5.

Reagle, Joseph M. (2002). Key Free Trust in the Semantic Web: Finding Bacon's Key. Consultado en: 23-11-2002.
<http://www.w3.org/2002/03/key-free-trust>

Silva Muñoz, Lidia.(2001). Representación de ontologías en la Web Semántica. Consultado en: 19-11-2002.
www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo_lydia.pdf

Simon, E.; Madsen, P.; Adams, C. (2001). An Introduction to XML Digital Signatures. Consultado en: 22-11-2002.
<http://www.xml.com/lpt/a/2001/08/08/xmldsig.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.

¹ www.w3.org/2001/03/identification-problem/

² www.cs.umd.edu/projects/plus/SHOE/

³ www.ontoknowledge.org

⁴ www.swi.psy.uva.nl/projects/ibrow/home.html

⁵ www.daml.org/tools

⁶ www.w3.org/2001/sw/WebOnt/

⁷ www.w3.org/Signature/

⁸ www.ietf.org

⁹ <http://rdfweb.org/foaf/>

¹⁰ www.w3.org/2001/sw/Europe/