

**Universidad Central de Venezuela
Facultad de Ciencias
Escuela de Computación**

Lecturas en Ciencias de la Computación
ISSN 1316-6239

Notas sobre Web Semántica

Jaime Blanco.

RT 2007-09

Centro IOMMA
Caracas, Septiembre, 2007.

**Universidad Central de Venezuela
Facultad de Ciencias
Escuela de Computación**

Lecturas en Ciencias de la Computación
ISSN 1316-6239

Notas sobre Web Semántica

Jaime Blanco

RT 2007-03

Centro IOMMA
Caracas, Septiembre, 2007

Notas sobre Web Semántica

Jaime Blanco

jblanco@ciens.ucv.ve

Universidad Central de Venezuela. Facultad de Ciencias. Escuela de Computación.
Centro de Investigación de Operaciones y Modelos Matemáticos (CIOMMA)
Venezuela. Caracas
Apdo. 47002, 1041-A.

RT 2007-03

ABSTRACT

Tenemos nuevos modelos para posibilitar una mejor colaboración e interacción entre personas y computadoras conectadas a Internet. El objetivo fundamental de la Web semántica, es buscar nuevos métodos de estructurar la información con la finalidad de que las computadoras de alguna manera puedan "entender" el contenido de los documentos que alojan, de forma tal, que estructuren y manejen la información a partir de una valoración semántica de estos contenidos. La idea es transformar la manera en la que trabaja la WWW. Lo que se busca es dar formato a los datos, con la finalidad de que éstos sean entendidos por las máquinas. Se muestran algunos elementos necesarios que la componen: XML, RDF, PICS, ONTOLOGÍAS y AGENTES

Keywords: Web Semántica, XML, RDF, PICS, Ontologías, Agentes.

Septiembre, 2007

Tabla de contenido

1.	El Intercambio de Conocimientos en la Web Semántica.....	6
2.	Elementos básicos de la web semántica.....	7
2.1.	XML.....	7
2.2.	RDF.....	18
2.3.	PICS	20
2.4.	ONTOLOGÍAS.....	21
2.5.	AGENTES	22
2.6.	web de primera generación vs. web de segunda generación	23
3.	La recuperación de la información en la Web semántica	24
4.	Hacia una nueva política de la propiedad intelectual.....	25
5.	Grupos de investigación.....	25
6.	Conclusiones	26
7.	Referencias.....	26

Lista de figuras

Figura 1.	Relación entre HTML, SGML y XML.....	9
Figura 2.	Plataforma para la selección de contenido en internet	20

Web Semántica

Una segunda generación de Web se acerca, estará basada en un nuevo modelo que posibilite una mejor colaboración e interacción entre personas y computadoras conectadas a Internet. Esta es una visión de avanzada del mismo creador del World Wide Web y director del World Wide Web Consortium (W3C), Tim Berners-Lee, así como de James Hendler, profesor de la Universidad de Maryland e investigador en la Agencia de la Defensa para la Investigación de Proyectos Avanzados (DARPA).

Este proyecto, al que también se le ha denominado Web Inteligente, implica la configuración de información legible tanto para las máquinas, como para los usuarios. Actualmente, detrás de la anarquía en la que ha caído el diseño Web, donde predomina la apariencia sobre la funcionalidad, la información es "coherente" para las personas; mientras detrás, en la programación y la estructura interna de los documentos, todo es un perfecto caos. Esta situación impide aprovechar de manera óptima, todo el potencial que los sistemas informáticos, lenguajes de programación y protocolos de comunicación pudieran proveer.

Llegar al objetivo fundamental de la Web semántica, es buscar nuevos métodos de estructurar la información con la finalidad de que las computadoras de alguna manera puedan "entender" el contenido de los documentos que alojan, de forma tal, que estructuren y manejen la información a partir de una valoración semántica de estos contenidos. Para lograrlo, los datos se definirán de un modo específico para ser entendidos por las computadoras y no sólo presentados a los usuarios; también podrán ser almacenados, procesados y reutilizados automáticamente sin que las personas tengan que iniciar los procesos.

La idea es transformar la forma en la que trabaja la WWW; no intenta desarrollar una nueva telaraña de la información, por el contrario, busca perfeccionar la que actualmente se encuentra en operación y ampliar sus capacidades, a través de lenguajes y herramientas que ordenen la información almacenada en Internet, de manera similar al análisis semántico del lenguaje humano.

XML RDF (Resource Description Framework o Marco de Descripción de Recursos), es un modelo con el cual se da formato a los datos, con la finalidad de que éstos sean entendidos por las máquinas. A corto plazo la Web semántica deberá construirse a partir de esos datos.

En suma, la meta es crear un ambiente transparente en la fachada y también en los cimientos. Para transformar la red en algo muy parecido a un cerebro digital global.

La Web semántica es el futuro de la Web, una Web que facilitará la localización de recursos, la comunicación entre sistemas y programas, que nos ayudará a gestionar nuestro día a día, hasta llegar a niveles que hoy pueden ser considerados como de ciencia ficción.

La red semántica no es una red independiente de la actual, sino una ampliación; una red en la cual la información está dotada de significados bien definidos, con el fin de que la coordinación del trabajo entre humanos y ordenadores sea más completa. Se han dado ya los primeros pasos para incluir la red semántica en la estructura de la red ya existente. En un futuro no lejano, estos desarrollos introducirán prestaciones nuevas e importantes, al lograr que las máquinas multipliquen su capacidad de procesar y comprender los datos que hoy tan sólo se exhiben en pantalla. En su universalidad el hipertexto permite "conectar cualquier cosa con todo", por tanto la técnica no debe discriminar entre un borrador y la prestación más depurada, entre información comercial y académica, ni entre culturas, idiomas, etc. No generamos el mismo tipo de información si estamos pensando en una persona o en una máquina, así tenemos por un lado un video de 5 sec. a una poesía y por el otros están las bases de datos, programas hechos por friskis informáticos,...

Hasta ahora, la red estaba concebida más para proporcionar documentos a los humanos que para manipular datos y/o procesar información de manera automática, pero a partir de ahora tenemos que cambiar el chip pues nos encontramos frente a una potente herramienta facilitadora de la comunicación, que es la red semántica, además ésta no pierda una de las características esenciales que es ser descentralizada.

Otra carencia de la situación actual es que, con los estándares Web del momento, no se puede diferenciar entre información personal, académica, comercial, etc. Es decir, cuando un buscador Web realiza una consulta con algunas palabras clave, normalmente aparece información que no es útil porque no corresponde a lo que estamos buscando. Además no todas las páginas proporcionan igual cantidad de información, debido precisamente a que no existe un formato o convenio que nos diga qué contenido debemos añadir a las páginas Web.

Por otro lado, los agentes de búsqueda actuales no se diseñan para "comprender" la información que reside en la Web, precisamente porque es prácticamente imposible conocer la representación de los datos ubicados en las diferentes páginas.

Si, por ejemplo, queremos buscar en la red a un dentista, que pertenezca a mi compañía de seguro médico "Segurín", y que se encuentre en mi ciudad "Cáceres", una forma podría ser indicarle al buscador tres palabras clave: *dentista*, *Segurín*, *Cáceres*. La información que puede encontrar el buscador puede estar relacionada con la prensa, demandas judiciales, información que reúna compañías de seguros, ... Normalmente gastamos mucho tiempo en seleccionar la información que nos puede ser útil, y navegando por las referencias URL hasta encontrar, con suerte, lo que estamos buscando. En el ejemplo, no es seguro que los dentistas tengan páginas personales de su consulta, o que la compañía del seguro tenga un directorio de sus dentistas, y que la información de los dentistas esté relacionada con la ciudad que estoy buscando. Si por casualidad encontrásemos varias páginas personales de dentistas que tuvieran en su anotación HTML esas palabras clave, es bastante improbable que la información que nos proporcionen sea similar; algunos pueden tener datos informativos, otros publicaciones, especialidades, horarios, ...

Es indudable que las ventajas que ofrece Internet son enormes a la hora de buscar información, pero adolece de una manera de encontrar información de forma precisa y de poder realizar deducciones con la información existente. En la sección 2, mostraremos la visión de la Web semántica para solucionar estos problemas indicados y veremos qué ventajas puede aportar a la situación actual. En la sección 3 se explicará el concepto de ontología, pieza fundamental para soportar la representación del conocimiento que necesita la Web semántica. Y en la sección 4 se verá una perspectiva de los pasos necesarios que se deben abordar para hacer efectivo este adelanto en la comunicación del conocimiento.

1. EL INTERCAMBIO DE CONOCIMIENTOS EN LA WEB SEMÁNTICA

En los últimos años, muchos investigadores están diseñando modelos para transformar la red desde un espacio de información a un espacio de conocimientos. Recientemente, Tim Berners-Lee, uno de los inventores de la Web, defiende el desarrollo de la Web con conocimientos [Ber01], y organizaciones como SemanticWeb (www.semanticweb.org) se encargan de estandarizar lenguajes y herramientas para hacer efectiva la Web semántica. Pero, ¿qué se puede hacer en la Web semántica?

La idea es que los datos puedan ser utilizados y “comprendidos” por los computadores sin necesidad de supervisión humana, de forma que los agentes Web puedan ser diseñados para tratar la información situada en las páginas Web de manera semiautomática.

Se trata de convertir la información en conocimiento, referenciando datos dentro de las páginas Web a metadatos con un esquema común consensuado sobre algún dominio. Los metadatos no sólo especificarán el esquema de datos que debe aparecer en cada instancia, sino que además podrán tener información adicional de cómo hacer deducciones con ellos, es decir, axiomas que podrán aplicarse en los diferentes dominios que trate el conocimiento almacenado.

Con ello, se mejorará la búsqueda de información y se potenciará el desarrollo de aplicaciones de comercio electrónico, ya que las anotaciones de información seguirán un esquema común, y los buscadores Web compartirán con las anotaciones Web los mismos esquemas. Empresas que traten con clientes y proveedores, podrán intercambiar sus datos de productos siguiendo estos esquemas comunes consensuados.

Los agentes Web no sólo encontrarán la información de forma precisa, si no que podrán realizar inferencias automáticamente buscando información relacionada con la que se encuentra situada en las páginas, y con los requerimientos de la consulta indicada por el usuario.

2. ELEMENTOS BÁSICOS DE LA WEB SEMÁNTICA

Antes de explicar cómo funcionará la red semántica, describiremos, los elementos necesarios que la compondrán, así: XML, RDF, PICS, ONTOLOGÍAS y AGENTES

2.1. XML

HTML (Hypertext Markup Language) se ha convertido en un lenguaje de marcas de inmensa popularidad durante estos últimos años. También debemos anotar que nos hemos encontrado con sus propias limitaciones, que algunas de ellas se han querido subsanar con la incrustación de scripts, javascripts, Active X, HTML dinámico, hojas de estilo en cascada (CSS). Todo esto es insuficiente para crear una arquitectura abierta de tipo cliente/servidor, con lo que el W3C (World Wide Web Consortium), organismo que vela por el desarrollo de la World Wide Web, se ha replanteado crear un nuevo estándar llamado XML (eXtensible Markup Language), que parte de las amplias especificaciones del SGML (Standard Generalized Markup Language) es el estándar internacional para la definición de la estructura y el contenido de diferentes tipos de documentos electrónicos. XML fue desarrollándose por el Grupo de Trabajo XML desde 1996 (en estos primeros años llamado SGML Editorial Review Board). La especificación XML 1.0 ha sido ratificada por la W3C el 10 de febrero de 1998, e interpretado como "un sistema para definir, validar y compartir formatos de documentos en la Web".

XML es un lenguaje de metamarcado que ofrece un formato para la descripción de datos estructurados. Esto facilita unas declaraciones de contenido más precisas y unos resultados de búsquedas más significativos en varias plataformas. Además, XML habilitará una nueva generación de aplicaciones para ver y manipular datos basadas en el Web.

XML ofrece una representación estructural de los datos que se puede implementar ampliamente y es fácil de distribuir. XML es un subconjunto de SGML optimizado para el Web. Definido por el World Wide Web Consortium (W3C) (en inglés), XML garantiza que los datos estructurados sean uniformes e independientes de aplicaciones o fabricantes. La interoperabilidad resultante está creando rápidamente una nueva generación de aplicaciones de comercio electrónico en la Web. XML, que proporciona un estándar de datos que puede codificar el contenido, la semántica y los esquemas de una gran variedad de casos, desde los más simples a los más complejos, sirve para marcar lo siguiente:

- Un documento normal.
- Un registro estructurado, como un registro de citas o un pedido de compra.
- Un objeto con datos y métodos, como el formulario permanente de un objeto Java o de un control ActiveX.
- Un registro de datos, como el conjunto de resultados de una consulta.
- Metacontenido sobre un sitio Web, como el formato de definición de canal (CDF).
- Representaciones gráficas, como la interfaz de usuario de una aplicación.
- Entidades y tipos de esquema estándar.

- Todos los vínculos entre datos y personas que hay en el Web.

Cuando los datos llegan al escritorio del cliente, se pueden manipular, editar y presentar en varias vistas, sin tener que regresar al servidor. Ahora los servidores pueden ser más escalables, gracias a la reducción de las cargas de ancho de banda y computación. Además, dado que los datos se intercambian en el formato XML, se pueden combinar fácilmente desde distintas fuentes.

XML es muy valioso para Internet, así como para los entornos de intranets corporativas de gran tamaño, pues proporciona interoperabilidad mediante un formato basado en estándares flexible y abierto, con formas nuevas de acceso a las bases de datos existentes y de entregar datos a clientes de Web. Las aplicaciones se pueden generar más rápidamente, su mantenimiento es más sencillo y pueden ofrecer fácilmente varias vistas de los datos estructurados.

En un principio, no rivalizarán HTML y XML, estos se complementarán el uno al otro, anudándose ambas gramáticas. Este Lenguaje de marcas extensible (XML) es una versión abreviada del SGML (Standard Generalized Markup Language).

Algunos de los objetivos planteados por el Grupo de Trabajo XML y el W3C son:

- XML debe ser directamente utilizable sobre Internet.
- XML debe soportar una amplia variedad de aplicaciones.
- XML debe ser compatible con SGML.
- Debe ser fácil la escritura de programas que procesen documentos XML.
- El número de características opcionales en XML debe ser absolutamente mínimo, idealmente cero.
- Los documentos XML deben ser legibles por los usuarios de este lenguaje y razonablemente claros.
- El diseño de XML debe ser formal, conciso y preparado rápidamente.
- Los documentos XML deben ser fácilmente creables.
- La brevedad en las marcas XML es de mínima importancia.

A estos fines se le une unos estándares como el Unicode e ISO/IEC 10646 para caracteres, el Internet RCF 1766 para identificación de lenguajes, ISO 639 para códigos de nombres de lenguajes y también el ISO 3166 para códigos de nombres de países, para la normal comprensión de esta versión de XML.

Tampoco tenemos que equivocarnos y pensar que el XML es un HTML++. Tanto el XML como el HTML tienen su base en el SGML. Es decir, es un metalenguaje que nos permite definir lenguajes para definir la estructura y el contenido de nuestros documentos. La definición de la estructura y el contenido de un tipo de documento se realizan en una DTD. En ella definimos los elementos que conformarán ese tipo de documentos y como tienen que estar organizados para que sea correcto. Un ejemplo de DTD es por ejemplo la que define cómo tendrán que ser los documentos HTML. Por tanto, el HTML no es más que un tipo de documento SGML que se utiliza en la Web, y esto es importante, ya que aquí radica su principal diferencia con el XML.

El XML no es ningún tipo de documento SGML, sino que es una versión abreviada de SGML optimizada para su utilización en Internet. Esto significa que con él vamos a poder definir nuestros propios tipos de documentos (podremos definir nuestras propias etiquetas) y, por tanto, ya no dependeremos de un único e inflexible tipo de documento HTML.

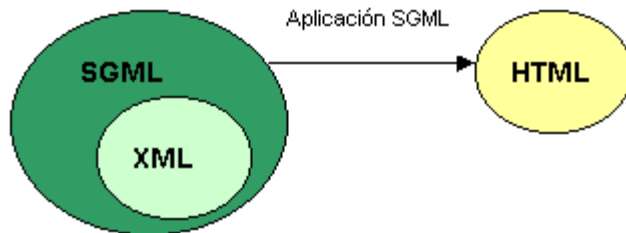


Figura 1. Relación entre HTML, SGML y XML

Diferencia entre el SGML, XML y HTML

El XML más que un HTML++ hay que considerarlo como un SGML-- optimizado para su utilización en Internet. Como escribió Richard Ligth en su libro Presentig XML, "XML ofrece el 80% de las ventajas del SGML con un 20% de su complejidad". Y es que los diseñadores de XML intentaron dejar fuera sólo aquellas partes que raramente se utilizan. Esta reducción resultó ser muy importante: la especificación XML ocupa aproximadamente 30 páginas, frente a las 500 del SGML.

Estructura de XML.

Un documento XML tiene dos estructuras, una lógica y otra física. Físicamente, el documento está compuesto por unidades llamadas entidades. Una entidad puede hacer referencia a otra entidad, causando que esta se incluya en el documento. Cada documento comienza con una entidad documento, también llamada raíz. Lógicamente, el documento está compuesto de declaraciones, elementos, comentarios, referencias a caracteres e instrucciones de procesamiento, todos los cuales están indicados por una marca explícita. Las estructuras lógica y física deben encajar de manera adecuada:

Los documentos XML se dividen en dos grupos, documentos bien formados y documentos válidos.

Un objeto textual o documento XML se dice que está bien formado si, considerándolo como conjunto, encaja con las especificaciones XML de producción, lo que implica:

- Contiene uno o más elementos.
- Hay exactamente un elemento, llamado raíz o elemento documento, de forma que ninguna parte del mismo aparece en el contenido de ningún otro elemento. Para todos los demás elementos, si la etiqueta inicial está en el contenido de otro elemento, la etiqueta final forma parte del contenido del mismo elemento. Dicho de forma más clara, los elementos, delimitados por una etiqueta de inicio y otra de final, se encuentran anidados correctamente.

- Cumple todas las restricciones que proporciona su especificación a través del DTD. Si no se utiliza DTD, el documento debe comenzar con una Declaración de Documento único (SDD, Standalone Document Declaration) que indique:
`<?XML version="1.0" standalone="yes"?>`
- Cada una de sus partes procesadas está bien formada.
- Todas las etiquetas deben estar balanceadas: esto es, todos los elementos que contengan datos de tipo carácter deben tener etiquetas de principio y fin (no está permitida la omisión excepto para los elementos vacíos.);
- Todos los valores de los atributos deben ir entrecomillados (el carácter comilla simple [el apóstrofe] puede utilizarse si el valor contiene caracteres comillas dobles, y viceversa): si necesitas ambos, utiliza `'` y `"`;
- Cualquier elemento VACÍO (p.e. aquellos que no tienen etiqueta final como ``, `<HR>`, y `
` y otros de HTML) deben terminar con `'/>'` o debes hacerlos no VACÍOS añadiéndoles una etiqueta de fin; Ejemplo: `
` se convertirá en `
` o en `
</BR>`.
- No debe haber etiquetas aisladas (`<` ó `&`) en el texto (p.e. debe darse como `<` y `&`); y la secuencia `]]>` debe darse como `]]>`; si no ocurre esto como final de una sección marcada como CDATA;
- Los elementos deben anidar dentro de sí sus propiedades (no se deben sobreponer etiquetas, como en el resto de SGML);
- Los Archivos bien-formados sin-DTD pueden utilizar atributos en sus elementos, pero éstos deben ser todos del tipo CDATA, por defecto.

De un documento bien formado, se dice además que es válido, si tiene DTD como el resto de aplicaciones SGML. Un archivo XML válido comienza como cualquier otro archivo SGML, con una Declaración de Tipo de Documento:

```
<?xml version="1.0"?>
<!DOCTYPE anuncio SYSTEM "http://www.foo.org/ad.dtd">
<anuncio>
<titulo>...<foto/>...</titulo>
<texto>...</texto>
</anuncio>
```

Dado que XML está diseñado para ser un subconjunto de SGML, cualquier documento XML válido debe ser también un documento SGML válido.

Estructuras lógicas.

Cada documento XML contiene uno o más elementos, cuyos límites están delimitados por etiquetas de comienzo y de final o, en el caso de elementos vacíos, por una etiqueta de elemento vacío.

Cada elemento tiene un tipo, identificado por un nombre, denominado identificador genérico, y puede tener un conjunto de especificaciones de atributos.

Cada especificación de atributo tiene un nombre y un valor. Estas especificaciones no restringen la semántica, el uso o (mas allá de la sintaxis) los nombres de los tipos de los elementos y los atributos, excepto de los nombres que comienzan por XML, que se reservan para estandarizar etiquetas o atributos en versiones posteriores del estándar.

Estructuras físicas.

Un documento XML puede consistir en una o más unidades de almacenamiento virtual, llamadas entidades. Todas estas unidades tienen contenido y todas ellas (excepto la entidad documento y el subconjunto externo del DTD) están identificadas por un nombre. Cada documento XML contiene una entidad, llamada entidad documento, que sirve como punto de comienzo para el procesador XML y puede contener el documento completo.

Como hemos comentado, las entidades pueden ser analizadas o sin analizar (también llamadas procesadas o sin procesar). El contenido de una entidad analizada se conoce también como texto de reemplazo, y es parte integrante del documento. Las entidades no analizadas son recursos (tales como enlaces) cuyo contenido puede o no ser texto, o en caso de que sea texto que no sea XML. Cada entidad no asociada tiene una notación asociada, identificada por un nombre. Aparte de obligar al procesador XML a haga accesible a la aplicación el nombre de esta notación y sus identificadores asociados, XML no proporciona ninguna otra restricción sobre el contenido de estas entidades. La forma de invocar ambos tipos de entidades es a través de su nombre, en el caso de las analizadas a través de su referencia a entidad y en el de las no analizadas a través de sus atributos de entidad.

Las entidades generales, son entidades analizadas que se usan en el interior del documento. Las entidades parametrizadas son entidades analizadas que se usan en el ámbito del DTD. Estos dos tipos de entidades usan distintos tipos de referencias y se reconocen en contextos distintos.

XML consta de cuatro especificaciones (el propio XML sienta las bases sintácticas y el alcance de su implementación):

- DTD (Document Type Definition): Definición del tipo de documento. Es, en general, un archivo(s) que encierra una definición formal de un tipo de documento y, a la vez, especifica la estructura lógica de cada documento. Define tanto los elementos de una página como sus atributos. El DTD del XML es opcional. En tareas sencillas no es necesario construir una DTD, entonces se trataría de un documento "bien formado" (well-formed) y si lleva DTD será un documento "validado" (valid).
- XSL (eXtensible Stylesheet Language): Define o implementa el lenguaje de estilo de los documentos escritos para XML. Desde el verano de 1997 varias empresas informáticas como Arbortext, Microsoft e Inso vienen trabajando en una propuesta de XSL (antes llamado "xml-style") que presentaron a W3C. Permite modificar el aspecto de un documento. Se puede lograr múltiple columnas, texto girado, orden de visualización de los datos de una tabla, múltiples tipos de letra con amplia variedad en los tamaños. Este estándar está basado en el lenguaje de semántica y especificación de estilo de documento (DSSSL, Document Style

Semantics and Specification Language, ISO/IEC 10179) y, por otro lado, se considera más potente que las hojas de estilo en cascada (CSS, Cascading Style Sheets), usado en un principio con el lenguaje DHTML. "Se espera que el CSS sea usado para visualizar simples estructuras de documentos XML (actualmente se ha conseguido mayor integración en XML con el protocolo CSS2 (Cascading Style Sheets, level 2) ofreciendo nuevas formas de composición y una más rápida visualización) y, por otra parte, XSL pueda ser utilizado donde se requiera más potencia de diseño como documentos XML que encierran datos estructurados (tablas, organigramas, etc.)(2)".

- XLL (eXtensible Linking Language): Define el modo de enlace entre diferentes enlaces. Se considera que es un subconjunto de HyTime (Hipermedia/Timed-based structuring Language o Lenguaje de estructuración Hipermedia/basado en el tiempo, ISO 10744) y sigue algunas especificaciones del TEI (Text Encoding Initiative o Iniciativa de codificación de texto). Desde marzo de 1998 el W3C trabajo en los enlaces y direccionamientos del XML. Provisionalmente se le renombró como Xlink y a partir de junio se le nombra como XLL. Este lenguaje de enlaces extensible tiene dos importantes componentes: Xlink y el Xpointer. Va más allá de los enlaces simples que sólo soporta el HTML. Se podrá implementar con enlaces extendidos. Jon Bosak establece los siguientes mecanismos hipertextuales que soportará esta especificación:
 - Denominación independiente de la ubicación.
 - Enlaces que pueden ser también bidireccionales.
 - Enlaces que pueden especificarse y gestionarse desde fuera del documento a los que se apliquen (Esto permitirá crear en un entorno intranet/extranet un banco de datos de enlaces en los que se puede gestionar y actualizar automáticamente. No habrá más errores del tipo "404 Not Found").
 - Hiperenlaces múltiples (anillos, múltiples ventanas, etc.).
 - Enlaces agrupados (múltiples orígenes).
 - Transclusión (el documento destino al que apunta el enlace aparece como parte integrante del documento origen del enlace).
 - Se pueden aplicar atributos a los enlaces (tipos de enlaces).
- XUA (XML User Agent): Estandarización de navegadores XML. Todavía está en proceso de creación de borradores de trabajo. Se aplicará a los navegadores para que compartan todas las especificaciones XML.

Aplicaciones de XML.

Algunas de las aplicaciones de XML son:

- Ofrecer mecanismos más versátiles de mostrar datos. Actualmente, bajo el nombre de DOM (Document Object Model) se está desarrollando una API que sea soportada por todos los procesadores de XML y HTML. La idea detrás de esta API es que podamos representar (a través de javascripts o JavaApplets) documentos XML en los navegadores Web, pero de una forma más sofisticada que los documentos HTML, ya que XML no solo proporciona una sintaxis, sino también una semántica.
- Buscadores inteligentes. Debido a que la información en los documentos XML está etiquetada por su significado de forma precisa, podemos localizarla de forma

mucho más clara que en documentos HTML. Con DTDs estandarizados para distintas aplicaciones (librerías, tiendas de deporte, catálogos de componentes,...) podríamos programar buscadores Web que recuperasen información sobre un producto de cualquier website en el mundo sabiendo que todos tendrán el mismo formato de datos (gracias al DTD), aunque no tengan necesariamente la misma representación gráfica (gracias al XML/XSL).

- Intercambio de información entre sistemas heterogéneos. El fundamento es el mismo que para los buscadores inteligentes. Debido a que el DTD proporciona un formato estándar para representar la información de un tema específico, puede usarse para simplificar el intercambio de información entre distintas fuentes (actualmente existen ya dos DTD estandarizados uno para fabricantes de chips y otro para industrias químicas, llamado CML).

Ventajas de XML.

- Los autores y proveedores pueden diseñar sus propios tipos de documentos usando XML, en vez de limitarse a HTML. Los tipos de documentos pueden ser explícitamente 'hechos a la medida de una audiencia', por lo que las difíciles manipulaciones que debes hacer con HTML para conseguir efectos especiales serán cosa del pasado: autores y diseñadores serán libres de inventar sus propias etiquetas;
- La información contenida puede ser más 'rica' y fácil de usar, porque las habilidades hipertextuales de XML son mayores que las de HTML.
- XML puede dar más y mejores facilidades para la representación en los visualizadores.
- Elimina muchas de las complejidades de SGML, en favor de la flexibilidad del modelo, con lo que la escritura de programas para manejar XML será más sencilla que haciendo el mismo trabajo en SGML.
- La información será más accesible y reutilizable, porque la flexibilidad de las etiquetas de XML pueden utilizarse sin tener que amoldarse a reglas específicas de un fabricante, como es el caso de HTML.
- Los archivos XML válidos son válidos también en SGML, luego pueden utilizarse también fuera de la Web, en un entorno SGML (una vez la especificación sea estable y el software SGML la adopte).

Desarrollo de aplicaciones con XML

Existen cuatro tipos de aplicaciones que impulsarán el desarrollo del XML: Aplicaciones que exijan que el cliente Web medie entre dos o más bases de datos. Se hará posible la integración de bases de datos distribuidas en los navegadores que admitan XML, pudiéndose modificar el contenido y la estructura de esta. Actualmente implantado en amplias redes nacionales, sin embargo, se limitan las posibilidades al establecerse una intranet/extranet y con amplias bases de datos que sólo permiten la visualización de los datos en el navegador. XML establecerá una arquitectura de 3 niveles (three-tier) que está representada de la siguiente manera:

- Se irá pareciendo cada vez más a una arquitectura cliente-servidor.
- Aplicaciones que intentan transferir una parte significativa de la carga del proceso del servidor al cliente Web. Funcionará con un subprograma Java que se insertará

en el PC del cliente. Esta carga hará que muchas de las funciones de modificación puedan desarrollarse desde el mismo navegador Web del cliente. Por ejemplo, dentro de una intranet con una aplicación de gestión bibliotecaria sobre los documentos disponibles en el centro bibliotecario. Se establecerán categorías específicas como los datos de la ficha catalográfica, fecha de préstamo, fecha de devolución, nombre del usuario, número de registro, etc. Todos estos ítems se pueden agrupar por categorías y estarán descritos externamente desde los "esquemas" (schemas) o los RDF del XML. Permite actualizarse en el acto y con la posibilidad interactiva por parte del usuario en el pedido, gestión y modificación de documentos desde su lejano navegador. El lado más negativo es que se necesitará mayor ancho de banda y mayor potencia del procesador del equipo para poder soportar esta arquitectura de tres-niveles.

- Aplicaciones que precisen que el cliente Web presente diferentes versiones de los mismos datos a diferentes usuarios. Se podría aplicar un manual de diferentes grados (iniciación, intermedio y avanzado) con sus diferentes idiomas, etcétera. Esto hará que este manual se pueda personalizar por los usuarios y extraer la información requerida de un capítulo determinado, con una ordenación y formatos concretos.
- Aplicaciones en las que agentes Web inteligentes intentan adaptar la búsqueda de información a las necesidades de los usuarios individuales. Habrá una interacción entre la información requerida y las preferencias del usuario de la aplicación. Con el XML vendrá una segunda generación de robots que permitirá una mayor precisión de la búsqueda requerida. Actualmente podemos encontrar aplicaciones de medios de comunicación como los periódicos personalizados. Por otro lado, el impulso de los motores de búsqueda con XML y la creación de buscadores de lenguaje natural.
- Otras aplicaciones que se desarrollarán son las operaciones para comercio electrónico con la normativa EDI. Se puede hacer ya comparación de precios entre los distintos tiendas virtuales.

Además de esto es importante destacar que XML aporta mucha potencia y flexibilidad a las aplicaciones basadas en la Web, proporcionando numerosas ventajas a los programadores y usuarios, como lo son:

- Búsquedas con más significado: Los datos se pueden etiquetar de forma exclusiva con XML, lo que permite que un cliente especifique libros escritos por Winston Churchill, en lugar de sobre Winston Churchill. Las búsquedas que utilizan los métodos actuales, por el contrario, probablemente mezclarían ambos tipos de libros. Sin XML, es necesario que la aplicación de búsqueda comprenda el esquema de cada base de datos, en el que se describe cómo se ha generado. Esto es prácticamente imposible, pues cada base de datos describe sus datos de una forma distinta. Sin embargo, con XML los libros se pueden clasificar fácilmente en categorías estándar por autor, título, ISBN u otros criterios. De esta forma, los agentes podrían buscar libros sobre Winston Churchill en estos sitios de librerías de una forma coherente.
- Programación de aplicaciones Web flexibles: Una vez encontrados los datos, el XML se puede distribuir a otras aplicaciones, objetos y servidores de nivel medio

para continuar su procesamiento, o bien se puede entregar al escritorio para su visualización en un explorador. XML, junto con HTML para la presentación, las secuencias de comandos para la lógica y un modelo de objeto común para la interacción con los datos y la presentación, ofrece todas las tecnologías necesarias para el desarrollo de flexibles aplicaciones Web de tres niveles.

- Integración de datos procedentes de fuentes dispares: La capacidad de buscar en varias bases de datos no compatibles entre sí es, hoy en día, prácticamente imposible. XML permite combinar fácilmente los datos estructurados procedentes de fuentes distintas. Se pueden utilizar agentes de software para integrar los datos en un servidor de nivel medio desde bases de datos de fondo y otras aplicaciones. A continuación, dichos datos se pueden entregar a clientes u otros servidores para su agregación, procesamiento y distribución. La capacidad de ampliación y la flexibilidad de XML le permiten describir los datos contenidos en una gran variedad de aplicaciones muy diversas, desde las recopilaciones descriptivas de páginas Web hasta los registros de datos. Además, dado que los datos basados en XML son autodescriptivos, se pueden intercambiar y procesar sin necesidad de una descripción incorporada de los datos entrantes.
- Computación y manipulación locales: Después de entregarlos al cliente, los datos en formato XML se pueden analizar, editar y manipular de forma local, siendo las aplicaciones clientes quienes realizan los cálculos. Los usuarios pueden manipular los datos de diversas formas, y no limitarse a presentarlos. El Modelo de objeto de documento (DOM) de XML también permite manipular datos con secuencias de comandos u otros lenguajes de programación. Los cálculos relativos a los datos se pueden realizar sin volver al servidor. La separación entre la interfaz de usuario que ve los datos y los propios datos permite crear, de forma natural, potentes aplicaciones para el Web que antes sólo se encontraban en bases de datos avanzadas, todo con un formato simple, flexible y abierto.
- Varias vistas de los datos: Una vez entregados los datos al escritorio, se pueden ver de varias formas. Al describir los datos estructurados de una forma simple, abierta y extensible, XML sirve de complemento para el HTML, que se utiliza ampliamente para describir las interfaces de usuario. Una vez más, mientras que el lenguaje HTML describe el aspecto de los datos, XML describe los propios datos. Dado que la presentación está separada de los datos, la definición de dichos datos en XML permite especificar varias vistas, lo que significa que los datos se pueden representar de la forma adecuada. Los datos locales se pueden presentar de una forma dinámica determinada por la configuración del cliente, las preferencias del usuario u otros criterios. CSS y XSL proporcionan mecanismos declarativos para describir una vista de los datos en particular.
- Actualizaciones granulares: Los datos se pueden actualizar de forma granular con XML, por lo que no es necesario volver a enviar un conjunto completo de datos estructurados cada vez que cambia parte de dichos datos. Sólo es preciso enviar el elemento modificado del servidor al cliente, y los datos modificados se pueden presentar sin necesidad de actualizar toda la interfaz de usuario. En la actualidad, aunque sólo cambie un elemento de los datos, es preciso volver a construir toda la página, incluso si la vista permanece igual, lo que reduce drásticamente la escalabilidad del servidor.

Asimismo, XML permite agregar otros datos, como las temperaturas máximas y mínimas previstas, la pluviosidad prevista y su probabilidad (expresada en porcentaje). Esta información adicional se puede trasladar a la vista del usuario sin que el explorador tenga que enviar una vista nueva. Si se solicitan datos adicionales, como la presión atmosférica, se pueden enviar sin volver a generar la vista.

Ejemplos

Ejemplo de libro en Amazon

Podemos observar como Amazon presenta en su Web información sobre los libros

El código en HTML es el siguiente:

```
<p>
<dt>
<b>
<a href="/exec/obidos/ASIN/0764531999/qid=919015337">
Xml : Extensible Markup Language</a></b> ~
<NOBR><font color=#990033>Usually ships in 24 hours</font></NOBR>
<dd> Elliotte Rusty Harold / Paperback / Published 1998
<br>
Our Price: $31.99 ~
<NOBR><font color=#990033>You Save: $8.00 (20%)</font></NOBR>
<br>
<a href="/exec/obidos/ASIN/0764531999/qid=919015337">
<i>Read more about this title...</i></a>
```

y en XML lo podríamos escribir de la siguiente manera:

```
<?xml version="1.0"?>
<libro>
  <titulo>Xml: Extensible Markup Language</titulo>
  <disponible tiempo="24" unidad="hours"/>
  <autor>Elliotte Rusty Harold</autor>
  <formato>Paperback</formato>
  <publicacion>1998</publicacion>
  <precio cantidad="31.99" moneda="dolar"/>
  <descuento cantidad="20"/>
  <enlacelibro href="/exec/obidos/ASIN/0764531999/qid=919015337"/>
</libro>
```

Es evidente que no hay que ser ningún programador experto para entender que cualquier programa informático podrá trabajar de forma más eficiente sobre el segundo ejemplo que sobre el primero.

Esto permitirá, por ejemplo, realizar motores de búsqueda mucho más eficaces, lo que nos permitirá un acceso más rápido y eficiente a la información. Nos permitirá acceder a nuestras páginas favoritas desde nuestro teléfono móvil, o desde la radio de nuestro

coche, en el momento en el que los programas de reconocimiento de voz trabajen con XML. Facilitará el intercambio de información y la cooperación entre las empresas facilitando el comercio electrónico, etc. Y es que el XML busca precisamente crear la capacidad de hacerlo todo en la Web.

La potencia de esta forma de trabajar radica en que estamos etiquetando e identificando el contenido, olvidándonos en un principio por la forma de presentarlo. ¡Ya nos encargaremos luego de darle un formato de salida!. El W3c está trabajando actualmente en el desarrollo de un lenguaje de hojas de estilo que nos lo permitan, denominado XSL (Extensible Style Language). Mediante una XSL podremos transformar un document XML en otro XML (por ejemplo en HTML) o convertirlo a un formato de impresión: RTF, PDF, etc.

Comentarios

En su calidad de estándar del sector para expresar datos estructurados, XML ofrece muchas ventajas a las organizaciones, desarrolladores de software, sitios Web y usuarios finales. Las oportunidades aumentarán cuantos más formatos de datos de mercado vertical se creen para mercados claves, como el mercado de búsqueda avanzada en bases de datos, banca en línea, médico, legal, comercio electrónico, etc. Cuando los sitios ofrezcan datos, en lugar de limitarse a las vistas de datos, las oportunidades serán extraordinarias.

Hoy en día, los servicios de atención al cliente están migrando de centros de llamadas y sedes físicas a sitios Web, por lo que se beneficiarán de la sólida funcionalidad de XML. Dado que la mayoría de estas aplicaciones empresariales implican la manipulación o la transferencia de datos y registros de base de datos, como pedidos de compra, facturas, información del cliente, citas, mapas, etc., XML va a revolucionar las posibilidades de los usuarios finales en Internet, permitiendo la implementación de un amplio abanico de aplicaciones empresariales. Además, la información que ya esté almacenada en sitios Web, tanto si está guardada en documentos como en bases de datos, puede marcarse con vocabularios basados en XML y orientados a intranets. Dichos vocabularios también ayudan a las pequeñas y medianas empresas que necesiten intercambiar información entre clientes y proveedores.

Un mercado vital y todavía por descubrir es el de las herramientas de desarrollo que simplifican a los usuarios finales la creación de sus propios sitios Web cooperativos, lo que incluye las herramientas para generar datos XML heredados de información de bases de datos e interfaces de usuario ya existentes. Además, se pueden desarrollar esquemas estándar para describir portafolios u otros datos que podrían utilizar, por ejemplo, el diseño, los gráficos y otras funciones de Excel o de otras hojas de cálculo existentes. Las herramientas declarativas y visuales para describir XML generadas a partir de bases de datos heredadas constituyen una oportunidad muy potente. Las herramientas personalizadas para ver datos XML se pueden escribir en el sistema de programación de Visual Basic®, en Java y en C++.

XML va a necesitar herramientas nuevas y potentes para la presentación de datos XML ricos y complejos dentro de un documento. Esto se consigue asignando una capa de presentación fácil de usar por encima de un conjunto complejo de datos jerárquicos que pueden cambiar de forma dinámica. Entre los diseños que se podrán utilizar para los datos XML se incluyen los esquemas contraíbles, las vistas dinámicas de tablas dinámicas y una sencilla hoja para cada portafolio.

Los sitios Web pueden ofrecer cotizaciones de bolsa, artículos de prensa o datos sobre el tráfico en tiempo real, los cuales se pueden conseguir a través de un filtro de difusiones del Web o mediante un sondeo inteligente de un árbol de servidores que repliquen dichos sitios. Con XML se puede evitar la sobrecarga de información con sólo redactar normas personalizadas sobre el envejecimiento de la información, igual que en el correo electrónico. Las herramientas basadas en XML para que los usuarios creen dichas normas, así como el software de cliente y servidor para ejecutarlas, constituyen una gran oportunidad. Un modelo de objeto estándar no podría habilitar dichas funciones, normalmente escritas en secuencias de comandos, para filtrar los mensajes entrantes, examinar los mensajes almacenados, crear mensajes salientes, obtener acceso a bases de datos, etc. Dichos agentes se pueden redactar de forma que se ejecuten en cualquier lugar de forma automática.

2.2. RDF

RDF son las siglas definen Resource Description Framework (algo así como marco de descripción de recursos) Como su nombre indica el área en la que está enmarcado es la descripción de recursos de la red, entendiendo por recurso todo lo que nos dé de sí la imaginación en tanto que a definir cualquier cosa, páginas, personas, dispositivos... RDF permite que las condiciones que se quieren "preguntar" sobre un recurso sean definidas como un conjunto de propiedades que componen el esquema.

RDF ofrece una estructura semántica *inambigua* (por el uso de los URI, Uniform Resource Identifier) que permite codificación, intercambio y procesamiento automático de los metadatos normalizados.

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

RDF usa la sintaxis del lenguaje XML para el intercambio y procesamiento de metadatos, las condiciones se recogen en los *rdf: Descripción de los elementos XML*.

Para poder explotar la Web semántica, se necesitan lenguajes de marcado apropiados que representen el conocimiento de las ontologías. Actualmente, mediante anotaciones RDF-RDF Schema se pueden representar algunas facetas sobre conceptos de un dominio y permite, mediante relaciones taxonómicas, crear una jerarquía de conceptos. Pero se necesitan lenguajes de marcado (basados en RDF) con mayor expresividad y capacidad de razonamiento para representar los conocimientos que contienen las ontologías

[Cor00]. De esta forma, existen ya disponibles herramientas como Protégé (<http://protege.semanticweb.org>), OntoEdit (<http://ontoserver.aifb.uni-karlsruhe.de/ontoedit/>), o WebOnto (<http://kmi.open.ac.uk/projects/webonto/>) para realizar anotaciones en páginas Web con lenguajes de marcado propios. El lenguaje con gran capacidad expresiva que está emergiendo como un estándar para realizar anotaciones de ontologías en Web es DAML - Darpa Agent Markup Language (<http://www.daml.org>), aunque en este momento no tiene sus formatos totalmente definidos.

Por otro lado, se necesitan agentes y aplicaciones Web que exploten este conocimiento anotado en las páginas Web. Estos agentes de conocimientos Web serán capaces de interpretar los esquemas ontológicos y axiomas de diferentes dominios, mantendrán la consistencia de las instancias que se inserten en las páginas Web siguiendo los esquemas ontológicos definidos, realizarán una búsqueda con inferencias utilizando los axiomas situados en los esquemas, y podrán realizar ligaduras de los árboles taxonómicos de varias ontologías.

Para potenciar el uso de ontologías en la Web, se necesitan aplicaciones específicas de búsqueda de ontologías, como (Onto)Agent (<http://delicias.dia.fi.upm.es/OntoAgent>) [Arp00], que indiquen a los usuarios las ontologías existentes y sus características para poder utilizarlas en su sistema.

La Web semántica es un espacio conceptual de información en el cual los recursos identificados por URIs (Identificador Universal de Recursos) se pueden procesar por máquinas. Funciona mediante los principios de 'comprensión parcial' y de 'inferencia' (pudiendo deducir nuevos conocimientos a partir de datos ya entendidos), y por lo tanto, de la evolución y la transformación. Dado que el URIs se está utilizando para representar los recursos, los sistemas pueden crecer sobre una base global descentralizada, similar a los sistemas de documentación del hipertexto en los primeros tiempos de la WWW.

Una vez que se asigna un URI a los datos, pueden entonces ser referidos por cualquiera, y como tales, complejas e intrincadas relaciones pueden ser construidas, cuestionadas y procesadas. En la base de este plan está la esperanza que la gente comenzará a publicar sus datos en RDF (Resource Description Framework o Marco de Descripción de Recursos)... En general, XML RDF es el formato elegido para la Web semántica... Parece haber una pequeña confusión entre la gente sobre la diferencia entre RDF y la Web Semántica. RDF es simplemente un modelo y formato de datos que permite a la gente crear datos legibles por la máquina. La Web Semántica será construida por encima de estos datos. Por lo tanto, cuando se publica algo en RDF, no se está creando necesariamente una parte de la Web semántica, sino que se está poniendo los datos a disposición los procesadores semánticos de la Web, si eso significa algo... El principio de la Web semántica es realmente bastante básico - datos legibles por la máquina, base global.

El esquema RDF y el DAML, el Darpa Agent Markup Language (+OIL. Ontology Interchange Language), son dos lenguajes RDF de nivel básico muy importantes. Entre

los dos, permiten a la gente definir nuevas aplicaciones sobre el RDF de una manera estructurada e interoperable... los lenguajes ontológicos e inferentes son un paso más allá, y proporcionan aún más potencia. Se pueden crear términos inversos, términos transitivos, equivalencias, tipos de datos, uniones, intersecciones, y así sucesivamente...

2.3. PICS

La tecnología de filtros "Platform for Internet Content Selection" (PICS) (Plataforma para la selección de contenido en Internet), desarrollada por el "World Wide Web Consortium", creado en 1994, provee un tipo de infraestructura de etiquetado de contenido para Internet (meta datos, códigos, privacidad), que posibilita que el usuario elija contenidos de manera selectiva. La PICS confía en el etiquetado y valoración de contenidos de los "Internet Content Providers" (Proveedores de contenido de Internet), terceras partes como las Agencias Gubernamentales, o directamente de los informes elaborados por los usuarios (<http://www.w3.org/PICS>).

La PICS provee una solución basada en la tecnología para un problema inducido tecnológicamente. Esta situación es, compatible con la cultura norteamericana de preferir la autorregulación a la regulación gubernamental y es también compatible con la cultura de Internet, ya expuesta, y su fuerte dependencia de la autodeterminación del individuo que tiene el suficiente conocimiento experto para protegerse a sí mismo del contenido no deseado de la red.

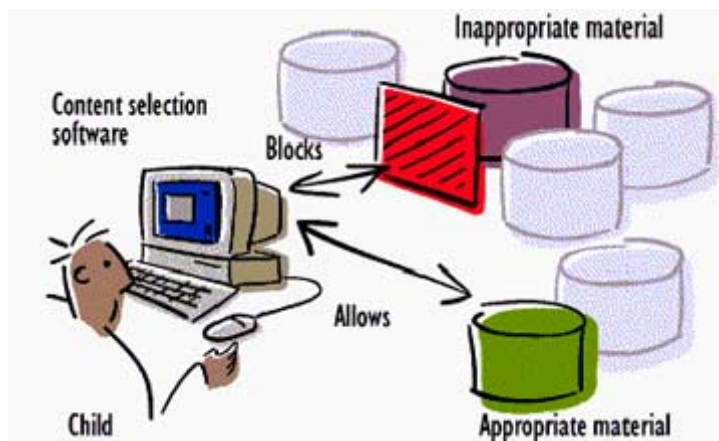


Figura 2. Plataforma para la selección de contenido en internet

Como las tecnologías de filtrado generalmente no eliminan el contenido de Internet, sino que sólo controlan el acceso a o la difusión del contenido, en principio, son compatibles con el pluralismo. La tecnología, en este caso, aparece neutral con respecto a los desarrollos culturales.

Los PICS nos indican lo adecuado o conveniente de determinados archivos de datos según la comunidad en la que se encuentre el usuario. Es una infraestructura para asociar las etiquetas con los contenidos de Internet. Aunque en un principio estaba destinado al control del acceso de los niños a Internet, su uso se puede extender a otras etiquetas que incluyan privacidad, licencias, etc. PICS es una plataforma sobre la cual se han

construido otros servicios de clasificación que no sólo define una manera de construir etiquetas sino que es un mecanismo para realizar las valoraciones. Este mecanismo esta formado, al menos por lo siguiente:

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

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

2.4. ONTOLOGÍAS

Las ontologías son colecciones de enunciados redactados en un lenguaje, como el RDF, que define las relaciones entre conceptos y especifica reglas lógicas para razonar con ellos. Los ordenadores "comprenderán" el significado de los datos semánticos de una página de la red siguiendo vínculos con ontologías especificadas.

Para que se pueda buscar la información de una manera precisa, se necesita que el conocimiento de la Web esté representado de forma que sea legible por los ordenadores, esté consensuado, y sea reutilizable. Las ontologías proporcionan la vía para representar este conocimiento.

El término ontología proviene de la filosofía; pero en IA, tiene diferentes connotaciones. La definición declarativa más consolidada es la propuesta por Gruber [Gru93a] y extendida por Studer y colegas [Stu98] que la describe como "una especificación explícita y formal sobre una conceptualización compartida". La interpretación de esta definición es que las ontologías definen conceptos y relaciones de algún dominio, de forma compartida y consensuada; y que esta conceptualización debe ser representada de una manera formal, legible y utilizable por los ordenadores.

Las ontologías tienen los siguientes componentes que servirán para representar el conocimiento de algún dominio [Gru93b]:

- **Conceptos:** son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.
- **Relaciones:** representan la interacción y enlace entre los conceptos del dominio. Suelen formar la taxonomía del dominio. Por ejemplo: *subclase-de*, *parte-de*, *parte-exhaustiva-de*, *conectado-a*, etc.

- **Funciones:** son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como *categorizar-clase*, *asignar-fecha*, etc.
- **Instancias:** se utilizan para representar objetos determinados de un concepto.
- **Axiomas:** son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: “*Si A y B son de la clase C, entonces A no es subclase de B*”, “*Para todo A que cumpla la condición C1, A es B*”, etc.

Estos últimos componentes, los axiomas, permiten junto con la herencia de conceptos, inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos.

Por ejemplo, con el conocimiento anotado en las páginas Web mediante ontologías, podremos utilizar un agente Web al que podamos preguntar sobre los dentistas que se encuentren a una cierta distancia de mi casa. Una de las posibles respuestas que me podría ofrecer el agente sería:

* Dra. Macías – Dentista – ortodoncia
 Consulta Particular (85€)- c/Rosa 4 - a 2 km
 Hospital Clínico “La Paz”-c/Principal 17 - a 5 km.

Incluso si la página Web de la Dra. Macías no tuviera especificada la dirección del hospital, el agente Web de búsqueda podía utilizar un atributo de los dentistas (*Trabaja_en_Hospital*), y con el valor de este atributo encontrar el atributo *Dirección* del concepto *Hospital* mostrar esta información del hospital concreto de esa instancia de *Dentista*.

Además de utilizar estas propiedades de relaciones y herencia de los conceptos especificados, mediante los axiomas se tendría mayor capacidad expresiva del dominio almacenado. Por ejemplo, si en este dominio tenemos declarado el axioma:

“*Si el médico trabaja en un hospital de mi seguro no tendré que pagar minuta.*”

El agente podría, utilizando el conocimiento representado en los conceptos, sus relaciones y utilizando el axioma, aconsejarnos sobre los dentistas que cumplieran este requisito.

2.5. AGENTES

El concepto de tecnología de agentes está aún en desarrollo, pero una definición que podemos manejar de agente es la siguiente: un agente es una entidad de software que funciona continua y autónomamente en un medio particular a menudo habitado por otros agentes y procesos, sin requerir de guía constante o intervención humana. En otras palabras, un agente es un asistente personal que está dentro de la computadora y que cumple varios roles en representación de una función específica o de un usuario.

En la Web semántica serán los encargados de realizar la búsqueda de servicios, para ello, la semántica facultará a los agentes para describir unos a otros la función exacta que realizan, y qué datos han de recibir para ello.

2.6.WEB DE PRIMERA GENERACIÓN VS. WEB DE SEGUNDA GENERACIÓN

Característica	Primera Generación	Segunda Generación
lenguaje principal	HTML	<i>XML</i>
forma y estructura	documentos no estructurados	Documentos estructurados
Semántica	semántica implícita	Etiquetado explícito (<i>metadatos</i> , Web semántica)
relación entre contenido y forma	HTML = fusión de forma y contenido	estructura en capas de forma y contenido: XML + transformación (p.e., <i>XSL</i>) a HTML, WML, PDF, u otros formatos
Editabilidad	documentos estáticos	Documentos dinámicos
descomponibilidad y recomponibilidad	sitios web monolíticos, independientes	bricolaje (agregación), sindicación, reasignación de contenido
interactividad	medio de difusión unidireccional	<i>web editable</i> , bidireccional
Audiencias	para consumo humano	para humanos y ordenadores (p.e., <i>servicios Web</i>)
control de producción	centralizado	descentralizado (<i>peer-to-peer P2P</i>)

3. LA RECUPERACIÓN DE LA INFORMACIÓN EN LA WEB SEMÁNTICA

Los actuales buscadores de Internet, como [google](#) se basan en un sistema en el que los usuarios meten manualmente los enlaces y al hacer una búsqueda se aplica un algoritmo de emparejamiento de patrones, que tiene en cuenta el número de veces que se hace referencia a cada url candidata a ser resultado de búsqueda.

Si bien hemos aprendido a convivir con este sistema de búsqueda, la única información que recuperamos con él son conceptos descontextualizados, es decir, si en un buscador ponemos la palabra "flor" recuperaremos una editorial que se llamen flor, una página Web para expresar sentimientos en Internet y, si tenemos suerte alguna página especializada en margaritas.

La Web semántica nos permitirá hacer búsquedas precisas del tipo quiero el viaje más barato que hay entre Madrid y Barcelona, teniendo en cuenta que me gusta ir en ventanilla y en no fumadores.

En el estado actual de Internet esto es ahora posible gracias a los sistemas multiagente. Donde un agente es una entidad de software que funciona continua y autónomamente en un medio particular a menudo habitado por otros agentes y procesos, sin requerir de guía constante o intervención humana. También debe poseer ciertas habilidades sociales, reactividad ante el mundo que le rodea y comportamiento basado en deseos que pertenecen a cada agente en particular. Los sistemas multiagentes se han propuesto como la mejor herramienta para realizar aprendizaje automático en Internet. Mediante esta herramienta es posible hacer recomendaciones acertadas del tipo "haz click en este sitio Web, pues me parece que te interesa", "esta persona tiene intereses muy similares a los tuyos, te interesaría contactar con ella" y esto lo hará software puro y duro, sin más intervención humana que la de la propia acción del usuario y acertarán. También será posible clasificar automáticamente sitios ó documentos de una manera acertada y un montón de más cosas que ni siquiera nos imaginamos.

El problema con el que ahora se topa la tecnología de sistemas multiagente es la volatilidad y desestructuración de la información base para realizar las inferencias. Volviendo al ejemplo del sistema multiagente para encontrar el viaje de tus sueños el único lugar desde el que puede recuperar la información son páginas html en forma de tablas; éstas páginas pueden cambiar su estructura cada mes, lo que requiere cambios de programación. Esto no es un problema si la información está almacenada y puede ser consultada desde bases de datos y/o archivos xml/rdf.

Hoy en día podemos ver cómo se popularizan foros al estilo de [slashdot](#) que dejan un archivo rdf (por ejemplo <http://slashdot.rdf>), dónde quedan las noticias de manera estructurada. La idea es que éste u otros métodos de compartir la información dinámica se popularice. Mientras tanto esta tecnología quedará restringida a poderosas bases de datos centralizadas como por ejemplo las de Amazon, que hagan recomendaciones de compra en base a lo que ya se ha comprado.

4. HACIA UNA NUEVA POLÍTICA DE LA PROPIEDAD INTELECTUAL

Una razón por la que es necesaria una nueva política de la propiedad intelectual es que el copyright afecta ahora a todos. Los avances en tecnología de información y redes digitales permiten que cada uno de nosotros se convierta en editor... Debido a que la infracción individual del copyright es tan difícil de vigilar en un entorno de red distribuida, los propietarios de copyright persiguen cada vez las tecnologías que permiten su infracción... La Hollings' Consumer Broadband y la Digital Television Promotion Act proibirían los reproductores digitales de uso común y de código abierto... La innovación y la competencia serían sofocadas si sistemas asignados por mandato se convirtieran en la ley... James Boyle ha polemizado sobre una nueva política de la propiedad intelectual en su ensayo ['Environmentalism for the Net'](#). Este ensayo precisa que en los años 50 no existía el concepto de 'medio ambiente'. Las compañías madereras y de explotación minera pensaban que la legislación referente a cuestiones sobre recursos naturales sólo les afectaban a ellos y empujaron a políticas que en ocasiones produjeron erosión y contaminación que arruinaron corrientes y lagos, y que exterminaron especies acuáticas y terrestres. Hubo de pasar un tiempo para que los ornitólogos y los cazadores (así como sociedad en general) se dieran cuenta de que tenían un interés común en la preservación de la naturaleza. Juntos inventaron el concepto de medio ambiente, y este concepto permitió que un movimiento político de gran alcance lo protegiera. Lo que se necesita es un movimiento similar para proteger los intereses intangibles que todos tenemos en un entorno de información abierto, en un dominio público robusto, y en una ley equilibrada de propiedad intelectual".

5. GRUPOS DE INVESTIGACIÓN

El World Wide Web Consortium ([W3C](#)), dirigido por Tim Berners-Lee, cuenta con un [grupo de trabajo](#) dedicado por entero a la Web semántica. En la página de este grupo se afirma que "la Web semántica es una visión: la idea de tener datos en la Web definidos y enlazados de forma que puedan ser utilizados por máquinas no sólo con el fin de presentarlos sino también para la automatización, integración y reutilización de la información a través de varias aplicaciones".

Asimismo, también se afirma en esta página que "la Web sólo alcanzará su pleno potencial si se convierte en un lugar donde la información puede ser compartida y procesada tanto por herramientas automáticas como por la gente".

Varios grupos de investigación interesados en el avance de la Web semántica mantienen operativo el nodo [SemanticWeb.org](#), donde se pueden seguir los avances en esta materia. Tres de estos grupos ([Onto-Agents](#), [Scalable Knowledge Composition](#) y [Protégé](#)) se encuentran en la Universidad de Stanford, en California, mientras que otro ([Ontobroker](#)) pertenece a la Universidad de Karlsruhe, en Alemania.

El objetivo de SemanticWeb.org es demostrar las posibilidades que ofrece el concepto de Web semántica, con la idea de que el nodo se convierta en un "portal automático de comunidad" que sirva de escaparate de esta tecnología.

James Hendler da clases de informática en la Universidad de Maryland, donde ha estado investigando sobre representación del conocimiento en la Web. Junto a se equipo de investigación desarrolló [SHOE](#), el primer lenguaje de representación del conocimiento basado en Web. Al mismo tiempo, es el responsable de investigación informática basada en agentes de la Defense Advanced Research Projects Agency ([DARPA](#)).

Por su parte, Ora Lassila es investigadora del Nokia Research Center de Boston y miembro del consejo asesor del W3C. Lassila fue una de las creadoras del RDF.

6. CONCLUSIONES

La Web Semántica proporcionará un salto cualitativo sobre el potencial de la Web. Las principales ventajas de esta nueva revolución en Internet serán el desarrollo de aplicaciones con esquemas de datos comunes, fomento de las transacciones entre empresas por comercio electrónico y búsqueda de información con inferencias.

Para poder lograr estos objetivos se necesita unificar los contenidos semánticos por medio de ontologías que formalicen este conocimiento de forma consensuada y reutilizable.

Se necesita un lenguaje común basado en Web, con suficiente capacidad expresiva y de razonamiento para representar la semántica de las ontologías; este hecho parece que lo veremos en muy poco tiempo. Los futuros agentes se diseñarán para explotar el conocimiento de la Web.

7. REFERENCIAS

[Arp00] Arpirez J., Gómez-Pérez A., Lozano Tello A. and Pinto S. "Reference Ontology and (Onto) Agent: The Ontology Yellow Pages", Knowledge and Information Systems, An International Journal, Springer-Verlag, 2 (2000) 4, 387-412. Mar. 2000.

[Ber01] Berners-Lee T., Hendler J. and Lassila O, "The Semantic Web", Scientific American, Volume 284, Number 5 (May, 2001), pages 34-43.

[Cor00] Corcho O., Gómez-Pérez A. "A Roadmap to Ontology Specification Languages". EKAW 2000 pp. 80-96

[Gru93a] Gruber T. ,"Toward Principles for the Design of Ontologies Used for Knowledge Sharing" Technical Report KSL-93-04, Knowledge Systems Laboratory, Stanford University, CA, 1993.

[Gru93b] Gruber T., "A Translation Approach to Portable Ontology Specifications", Knowledge Acquisition, 5 (2), 199-220, 1993b.

[*Stu98*] Studer S, Benjamins R., and Fensel D., “Knowledge Engineering: Principles and Methods”, *Data and Knowledge Engineering*, 25, 161-197, 1998.

<http://www.w3.org/XML/>

<http://www.xml.com/xml/pub>