



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES ARAGÓN

LA WEB SEMÁNTICA

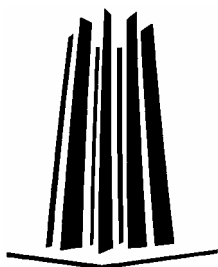
Trabajo monográfico en la modalidad de:
ALTO NIVEL ACADÉMICO

Que para obtener el título de:
INGENIERO EN COMPUTACIÓN

Presenta:
ARTURO RODRÍGUEZ GARCÍA

Asesor:
M. EN C. MARCELO PÉREZ MEDEL

MÉXICO, 2009





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimiento

A mi familia, por haberme apoyado a cumplir cada uno de mis sueños,
y por haber soportado mis locuras y desvelos.

Índice

Introducción	5
Capítulo 1 El camino hacia la Web semántica	7
1.1 Génesis de la World Wide Web	8
1.2 El gran iceberg de la información.	9
1.3 La Web sintáctica y sus limitaciones	11
1.4 ¿Qué es la Web semántica?	14
1.5 La relación entre Web semántica e Inteligencia Artificial	16
1.6 La evolución hacia la Web semántica	17
1.7 Las fases de la evolución	19
Capítulo 2 Arquitectura de la Web semántica	20
2.1 La capa no semántica	21
2.1.1 Unicode	21
2.1.2 URI	23
2.2 La capa XML	24
2.2.1 Metadatos	24
2.2.2 Representación del contenido Web semántico mediante el modelo de datos XML	27
2.2.3 XML Namespaces	31
2.2.4 XML Schema	33
2.2.5 XML y el consumo humano de la Web	37
2.3 La capa RDF	41
2.3.1 Descripción de recursos Web mediante RDF	41
2.3.2 Establecimiento de relaciones entre recursos Web mediante RDF	43
2.3.3 RDF Schema	44
2.4 La capa ontológica	48

2.4.1 Elementos de una ontología	49
2.4.2 OWL	51
2.5 Las capas aún no desarrolladas.	54
2.5.1 La capa lógica	54
2.5.2 La capa de evidencia.	55
2.5.3 La capa de confianza.	56
2.6 Concepto de firma digital	57
 Capítulo 3 Aplicaciones de la Web semántica	 58
3.1 Agentes inteligentes personales	59
3.2 La Web semántica social	62
3.3 Administración del conocimiento	66
3.4 Comercio Electrónico (E-commerce)	67
3.5 Herencia cultural	68
3.6 Relaciones interpersonales	68
 Conclusiones69
 Fuentes	 71
 Índice de figuras y tablas	 73

Introducción

En la última década del siglo XX la Web nació y se expandió rápidamente por todo el mundo, trayendo consigo una nueva forma de compartir y obtener información que la humanidad nunca antes había visto. Hoy en día, a menos de veinte años de su nacimiento, la Web se ha vuelto parte de la vida cotidiana de muchas personas, extendiéndose por todos los rincones del orbe y creando una Pangea virtual en donde todos los continentes se encuentran unidos. La Web no discrimina a nadie, y dota a cada ser humano con acceso a Internet de la posibilidad de utilizar los recursos y servicios que tiene, o de ser partícipe mediante la aportación de más información que contribuya a su crecimiento.

La Web es grandiosa como la conocemos hoy en día, pero podría ser mucho mejor. Mientras millones de personas la utilizan segundo a segundo alrededor del planeta, grupos de científicos están detrás de ella, dándole el soporte necesario para su funcionamiento y realizando arduas investigaciones en la búsqueda de ideas para mejorarla.

Actualmente, el interés primordial de esos científicos es llevar a la Web que conocemos a una nueva fase en su evolución. Esta nueva fase se llama Web semántica, y consiste en reestructurar el contenido de la Web de modo que pueda ser entendido tanto por los humanos como por las computadoras, a diferencia del contenido Web actual que simplemente es desplegado por las computadoras y cuya interpretación se delega a los usuarios que lo ven.

El hecho de dotar a las computadoras de la posibilidad de entender el significado de la información disponible en la Web es una idea que en principio parece futurista e imposible, pero en realidad es una posibilidad que se está gestando paulatinamente. De hecho, los avances con respecto a la implementación de esta nueva Web ya son significativos, y se puede considerar que una incipiente Web semántica ya está en funcionamiento.

El presente trabajo monográfico tiene como objetivo mostrar un panorama de la Web semántica, lo que motiva su construcción, su importancia, su estructura, los elementos que la conforman, así como las aplicaciones que se le pueden dar.

En el primer capítulo se plantea un breve esbozo histórico acerca del surgimiento de la Web, pasando después a una explicación acerca de su estructura y funcionamiento. Se abordarán algunas de las limitaciones que tiene la Web actual, y se mostrará a la Web semántica como una solución a esas problemáticas.

En el segundo capítulo se explica la forma en que se implementará esta nueva Web, mostrando en primera instancia su arquitectura y detallando a continuación los elementos que la conforman. En esta parte se señalarán los lenguajes y tecnologías que se utilizarán para hacer de la Web semántica una realidad, entre ellos XML, XML Schema, RDF, OWL, entre otros. La idea de este capítulo es justificar como ese lenguaje o tecnología está relacionado con la implementación de la nueva Web y qué papel jugará en su estructura, pero sin entrar en detalles específicos de cada uno de ellos, ya que para eso existen manuales al respecto que abordan profundamente la sintaxis y estructuras para cada lenguaje en particular.

En el tercer capítulo se mostrará la importancia de la Web semántica mediante la presentación de una serie de aplicaciones que se le pueden dar, algunas de las cuales ya están presentes en la actualidad y otras que, aunque no se han logrado implementar, las veremos funcionando en unos cuantos años.

El camino hacia la Web semántica

La World Wide Web (Red Global Mundial) se puede definir como un sistema o estrategia de comunicación a través de Internet, formado por un conjunto de documentos denominados páginas Web, que están almacenadas en diferentes computadoras alrededor del mundo (denominadas servidores). Dichas páginas están enlazadas entre sí mediante hipervínculos y son visualizadas mediante un navegador Web.

La Web se ha vuelto parte de la vida cotidiana de muchas personas, desde empresarios que crean un sitio Web para mostrar al mundo las ventajas de comprar sus productos y solicitar sus servicios, hasta los pequeños infantes que buscan imágenes sobre sus caricaturas favoritas, sin olvidar mencionar a las amas de casa en busca de un consejo de belleza en los portales femeninos, o a los adolescentes descargando música y videojuegos.

La Web se ha vuelto el medio para compartir información más importante de la humanidad. La gran libertad de subir cualquier tipo de información, el anonimato de las personas que participan en este medio, así como la facilidad de uso, son factores que han provocado el éxito de la Web, pero que también han ocasionado un crecimiento desmesurado de la información existente, a tal grado que navegar en la Web se ha vuelto una tarea verdaderamente compleja. En la Web hay mucha información, pero nada nos garantiza que encontraremos lo que realmente buscamos.

En este capítulo se abordarán las problemáticas actuales de la Web como la conocemos, y se planteará cómo la Web semántica busca mejorar el funcionamiento de esta gran telaraña de información que la humanidad ha entretejido en menos de veinte años.

1.1 Génesis de la World Wide Web

Fue en el año de 1980 cuando el científico inglés Tim Berners-Lee hizo la primera propuesta donde se podían ver en forma práctica las nociones básicas acerca de lo que hoy conocemos como la Web. Esta propuesta consistió en un programa de software denominado *Enquire*, cuyo nombre y esencia estaban basados en el libro de consejos que vio durante su infancia en la casa de sus padres, llamado *Enquire Within upon Everything* (que se traduce como “Preguntar en el interior acerca de cualquier cosa”), el cual “servía como portal a un mundo de información acerca de todo, desde como quitar manchas de la ropa, hasta consejos para invertir dinero”. (Berners-Lee, Fischetti y Dertouzos, 2000a: 1)

El sueño original de Tim Berners-Lee era crear un programa que funcionara como ese libro que vio cuando era un niño, de modo que el usuario pudiera preguntar cualquier cosa y el programa hiciera la búsqueda y entregara una respuesta concreta a la interrogante.

En 1989 Tim Berners-Lee, tras años de experimentar con *Enquire*, hizo una propuesta más formal y sólida acerca de su idea. De esta manera surgió la Web. Para finales de 1990, Tim Berners-Lee había instalado el primer servidor Web del mundo, escrito el primer navegador (llamado *WorldWideWeb*) y creado las primeras páginas Web en las que se describía su proyecto.

Aunque originalmente Tim Berners-Lee desarrolló esta estrategia de comunicación vía Internet para los investigadores de Física de alta energía de la Organización Europea para Estudios Nucleares (CERN)¹, en 1993 esta misma institución anunció que la Web sería gratuita para todo aquel que quisiera utilizarla, sin ningún tipo de honorarios².

La Web no tardó en popularizarse, en pocos años se convirtió en uno de los mecanismos proveedores de información más importantes de la humanidad, revolucionando

¹ EL CERN se encuentra ubicado en Ginebra, Suiza. Es una de las organizaciones europeas más importantes en investigación y necesita desarrollar herramientas computacionales poderosas que le permitan resolver problemas referentes a la investigación nuclear, destinando gran parte de sus recursos a la investigación de temas relacionados con computación.

² Si Tim Berners-Lee no hubiera propuesto que la Web fuera gratuita y en vez de ello hubiera decidido lucrar con su idea, sería hoy uno de los hombres más ricos del mundo.

la forma en que los usuarios de Internet compartían y accedían a la información. Muchas páginas y sitios Web se crearon rápidamente, así como los navegadores que permitían visualizar el contenido Web disponible.

Hoy en día, la Web es usada para una infinidad de propósitos: correo electrónico, mensajería instantánea, revisar catálogos en línea, compras por Internet, conocer nuevas personas, buscar información para tareas escolares, compartir fotos, descargar música, etc.

1.2 El gran iceberg de la información

El primer paso para entender a la Web es darnos una idea general de su estructura y tamaño actual.

De acuerdo con los autores Vipul Kashyap, Christoph Bussler y Matthew Moran (2008: 23-24), existen dos tipos de contenido Web, uno al que llamaremos la “superficie de la Web” (surface web) que consiste en páginas a las que se puede acceder y ver su contenido mediante el uso del navegador. La superficie de la Web consiste en aproximadamente 2.5 billones de documentos, con una tasa de crecimiento de 7.3 millones de páginas por día. La mayoría de estos documentos son páginas HTML estáticas³ que incluyen imágenes, audio y video.

El otro tipo de contenido es la “Web profunda” (deep web), que consiste en bases de datos especializadas para la Web y sitios Web dinámicos⁴, que no son conocidas por los usuarios comunes. La información en la Web profunda es de 400 a 550 veces más grande que la que se encuentra en la superficie.

Gracias a las nuevas tecnologías, muchas páginas de la Web profunda pueden ser visualizadas desde el navegador. Páginas en formato distinto a HTML (como PDF, Word,

³ Un documento Web estático es aquel que no cambia su contenido (el contenido está preestablecido y se modifica únicamente cuando el creador lo hace manualmente).

⁴ Un documento Web dinámico es aquel que sufre cambios frecuentes en su contenido. Cuando un usuario accede a dicho documento, éste se genera automáticamente mediante un software.

Excel, Corel Suite, etc.) son traducidas a HTML y son desplegadas en la lista de resultados de los motores de búsqueda.

El contenido almacenado en bases de datos de la Web profunda es devuelto por los servidores Web y convertido en páginas HTML dinámicas. En la Web profunda también están páginas que sólo pueden ser visualizadas por determinados usuarios (accediendo mediante una contraseña).

De acuerdo con la clasificación anterior podríamos imaginar a la Web como un iceberg de tamaño colosal, en donde lo único que vemos es la pequeña punta que sale del agua, como se ilustra en el siguiente esquema:

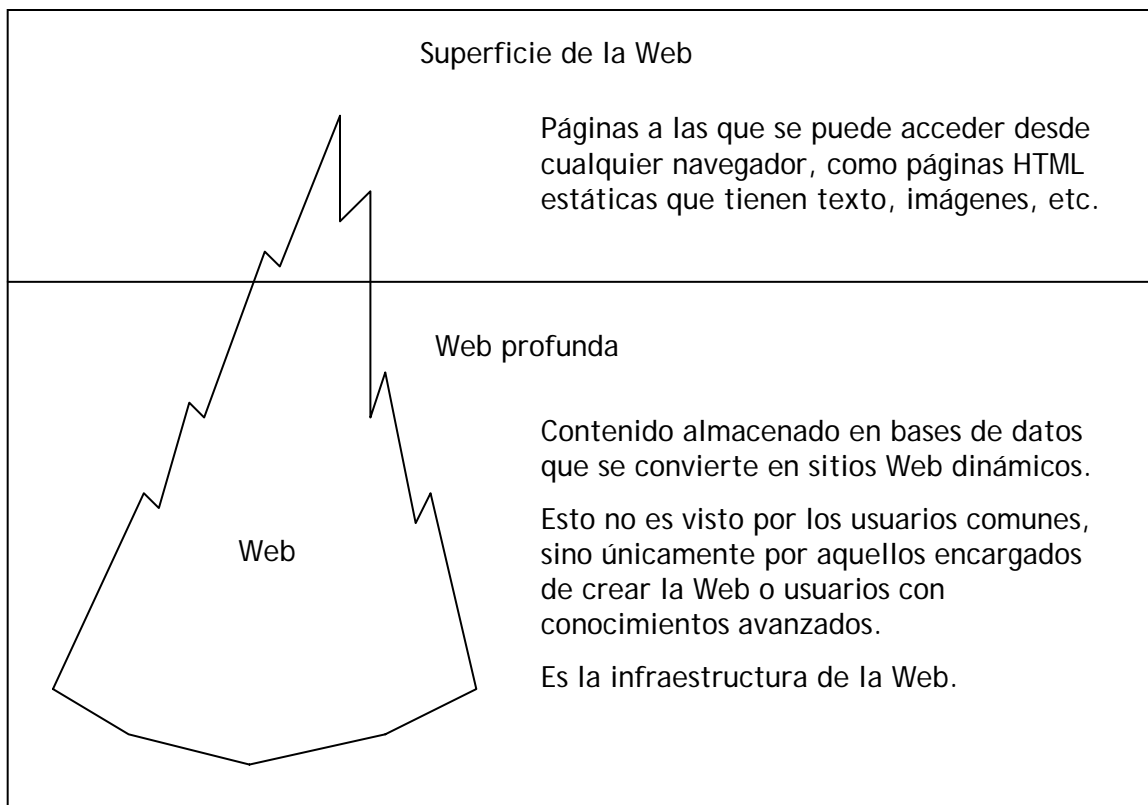


Figura 1.1 La Web vista como un gran iceberg de información

Por ejemplo, cuando utilizamos un servicio de correo electrónico (como Yahoo o Hotmail) e ingresamos a nuestra cuenta y consultamos uno de nuestros mensajes, estamos viendo únicamente la superficie de la Web. Nuestro mensaje es traído de la Web profunda,

en donde están millones de mensajes para otros usuarios, y están almacenados otros elementos que permiten el funcionamiento de la Web, como las cuentas de todos los usuarios, sus contraseñas, etc.

Cabe agregar que la Web se caracteriza por sus propiedades de universalidad y descentralización. Es universal porque no discrimina a nadie, está hecha para todos y cualquiera puede participar en ella, sin hacer distinciones entre culturas, ideologías, edades, etc. Es descentralizada porque no hay un punto rector que guíe su crecimiento. (Berners-Lee, Hendler y Lassila, 2001)

A pesar de la gran ventaja que ofrece la Web para compartir y obtener información, la forma en que está construida no es la más adecuada y presenta ciertos problemas, este aspecto lo analizaremos cuidadosamente a continuación.

1.3 La Web sintáctica y sus limitaciones

La Web que utilizamos hoy en día está conformada por una gran cantidad de páginas, la gran mayoría de ellas escritas en HTML, que es típicamente “el lenguaje usado para codificar información acerca de renderización (tamaño de fuente, color, posición en la pantalla, etc.) e hipervínculos a otras páginas Web o recursos en la Web (archivos multimedia, texto, direcciones de e-mail, etc.)”. (Breitman, Casanova y Truskowski, 2007: 3)

Sin embargo, HTML lo único que hace es dar un formato de presentación visual a lo que se desplegará en el navegador, sin entender en ningún momento el significado del mensaje mostrado. La interpretación del contenido del mensaje es una tarea del ser humano que lo leerá.

Podemos notar que el poder de las computadoras está siendo subutilizado, ya que “la mayoría de las páginas Web están todavía diseñadas para consumo humano y no pueden ser procesadas por las máquinas. Las computadoras son usadas solamente para mostrar la

información, esto es, para decodificar el esquema de color, encabezados y links codificados en las páginas Web". (*Ibidem*: 3)

Cabe mencionar que "incluso el contenido Web que es generado automáticamente de bases de datos es usualmente presentado sin la estructura de información original encontrada en las bases de datos". (Antoniou y Van Harmelen, 2004: 1).

Por estas razones, "la Web de hoy en día podría ser definida como la Web sintáctica, donde la presentación de la información es proveída por las computadoras, y la interpretación e identificación de la información relevante es delegada a los seres humanos". (Breitman *et. al.*, 2007: 4)

Es claro que el ser humano por naturaleza tiene ventajas sobre las computadoras para analizar y discernir entre información útil y no útil, pero el problema se vuelve complejo cuando la cantidad de información es tan grande que es imposible revisarla toda. Esta problemática se presenta a cualquier persona que hace uso de un motor de búsqueda en la Web.

Una de las herramientas principales para búsqueda de información en la Web son los motores de búsqueda (por ejemplo AltaVista, Yahoo, Google, etc.), los cuales solicitan al usuario una o varias palabras claves, y realizan una búsqueda en la Web para encontrar documentos en los cuales estén presentes esas palabras. El resultado por lo general es que el usuario recibe cientos de miles de entradas (páginas Web) que contienen el término o la combinación de los términos solicitados.

Para entender la problemática de la Web actual, ilustremos con un ejemplo. Imaginemos que un usuario quiere investigar todo lo referente a la ofrenda de Día de Muertos que se coloca en la Facultad de Estudios Superiores Aragón (FES Aragón) año tras año. Este usuario ingresa a un motor de búsqueda, en este caso Google, e introduce como palabras claves *Ofrenda de día de muertos de la FES Aragón*. Imaginemos que este usuario es novato en el manejo de la Web y no conoce trucos de búsqueda (como entrecomillar las palabras para buscar una frase exacta o utilizar operadores lógicos para lograr una búsqueda más precisa).

El motor de búsqueda realiza su trabajo y en tan sólo 0.13 segundos le regresa al usuario una lista de 3 450 documentos que contienen las palabras "Ofrenda", "de", "día", "de", "muertos", "de", "la", "FES", "Aragón". Esta lista no ofrece ningún tipo de respuesta al usuario, simplemente le da una serie de sugerencias sobre sitios que podrían tener la información que él requiere. Ahora, le toca al usuario verificar cuál de todos estos sitios le podría servir. Es lógico que no revisará la lista completa, simplemente verificará algunos de esos sitios para ver si contienen algo que le sea útil.

Como se pudo ver en el ejemplo anterior, los motores de búsqueda por palabras clave sufren varias limitaciones, las cuales se enlistan a continuación:

- ✓ Los resultados de la búsqueda pueden tener un gran número de páginas Web que contengan los términos o la combinación de términos solicitados, lo cual no es útil para el usuario, ya que es imposible que revise cada uno de los documentos para clasificar entre los útiles y los no útiles.
- ✓ Muchas de las entradas hacen referencia a páginas Web que contienen los términos que el usuario estaba buscando, pero dichas palabras están dispersas por el documento sin tener conexión entre ellas.
- ✓ Es común que una misma página Web aparezca listada varias veces en los resultados de la búsqueda.
- ✓ Si la solución a nuestra búsqueda está dispersa entre varias páginas Web, el motor de búsqueda debería entregar el conjunto de entradas que le permitan al usuario tener lo que está buscando, sin embargo siempre muestra los resultados en forma de páginas individuales.
- ✓ Los motores de búsqueda no entregan soluciones a lo que buscamos, sino sugerencias de documentos en donde podríamos hallar la respuesta.

A pesar de sus limitaciones, la Web depende de estos motores de búsqueda que son utilizados día tras día por millones de usuarios alrededor del mundo. Se han implementado algoritmos para mejorar el funcionamiento, por ejemplo, en algunos motores de búsqueda se colocan en las primeras entradas las páginas Web que son más visitadas. Sin embargo, el hecho de que una página Web sea visitada por muchas personas no implica que sea la mejor respuesta para un usuario en particular, y en cambio, una página poco popular pero con la respuesta correcta podría quedar hasta el final de la lista.

A final de cuentas, el usuario es quien tiene la responsabilidad de seleccionar las páginas que le sean útiles. “Por supuesto, el proceso de interpretación es muy demandante y requiere gran esfuerzo para evaluar, clasificar y seleccionar información relevante” (*Ibidem*: 4). Esto dista mucho del sueño original del creador de la Web, en donde el usuario simplemente pregunta y las computadoras se encargan de buscar la respuesta.

Debido a que el volumen de los datos digitales disponibles está creciendo en una razón exponencial, es imposible para los seres humanos administrar la complejidad y volumen de la información disponible. Este fenómeno es conocido como “sobrecarga de información”. (*Ibidem*: 4)

Es en este punto donde surge la necesidad de un cambio en la Web, haciendo que las computadoras no sólo desplieguen la información en pantalla, sino que ayuden al proceso de interpretación y selección de la información de una manera más efectiva.

Sin embargo, la estructura actual de la Web sintáctica no permite esto, ya que “las páginas Web no contienen información acerca de ellas mismas, esto es, acerca de sus contenidos y las materias a las cuales se refieren” (*Ibidem*: 4). Una página Web lo único que “sabe” es la forma en que mostrará el contenido en pantalla, pero ignora por completo el significado de lo que está mostrando.

La solución a estos problemas podría ser modificar la estructura de la Web actual, de modo que las páginas Web tengan información sobre ellas mismas y entiendan aunque sea un poco el significado de su contenido. Es aquí donde entra la Web semántica.

1.4 ¿Qué es la Web semántica?

Para solucionar los problemas expuestos anteriormente, una alternativa es “representar el contenido de la Web en una forma que sea más fácilmente procesable por la máquina y use técnicas inteligentes para tomar ventajas de estas representaciones. Nos

referiremos a este plan de revolucionar la Web como la iniciativa de la Web Semántica". (Antoniou *et. al.*, 2004: 3)

Una definición clara de la Web semántica podría expresarse como una Web de datos inteligentes procesables por máquinas, donde los datos inteligentes son independientes de las aplicaciones, compuestos, clasificados, y parte de un gran ecosistema de información. (Daconta, Obrst y Smith, 2003: 4)

Esta nueva Web posee motores de búsqueda semánticos, los cuales reemplazarán a los motores de búsqueda por palabras clave, y serán capaces de responder preguntas como "¿qué vuelos me llevan el fin de semana a Canadá?", en vez de sugerir sitios Web asociados con las palabras "vuelos", "fin de semana" y "Canadá". Estos motores de búsqueda serán capaces de examinar a la Web como un todo (en vez de verlo como un conjunto de documentos independientes), y formarán la respuesta a la pregunta extrayendo información de varios sitios, apoyándose en el entendimiento que tendrán de la información dispuesta en la Web. (Benson, 2008: 4)

La clave de la Web semántica está en darle más importancia a los datos que la que tienen en la estructura actual de la Web. Tradicionalmente, los datos han estado en segundo plano con respecto al procesamiento de los mismos. El software útil es dependiente totalmente de los datos⁵. La Web semántica propone un cambio de paradigma, en donde el poder no esté en las aplicaciones sino en los datos (Daconta *et. al.*, 2003: 2). De esta manera, los datos deberán estar estructurados de una manera independiente de las aplicaciones que los vayan a utilizar. Pero, ¿cómo se podría lograr esto?.

Para organizar el contenido Web los investigadores de Inteligencia Artificial han propuesto una serie de modelos conceptuales. "La idea central es categorizar la información en un modo estándar, facilitando su acceso. Esta idea es similar a la solución usada para clasificar seres vivos. Los biólogos usan una taxonomía bien definida, la taxonomía de Linneo, adoptada y compartida por la mayoría de la comunidad científica

⁵ Esta actitud incorrecta ha generado la expresión GIGO que significa "Garbage in, Garbage out" (es decir, si introduces datos incorrectos o basura, tendrás como resultado basura). GIGO básicamente revela el defecto en el argumento original estableciendo la dependencia entre procesamiento y datos. (Daconta *et. al.*, 2003: 2)

alrededor del mundo. De la misma manera, los científicos de la computación están buscando un modelo similar para ayudar a estructurar el contenido Web". (Breitman *et. al.*, 2007: 6)

Sin embargo, esta estructuración de la información ha causado gran controversia, ya que se considera que el gran éxito de la Web actual está relacionado con la libertad que provee. "En el mismo ambiente, podemos encontrar sitios Web muy sofisticados, diseñados por especialistas, y páginas Web personales, creados por individuos con poca o ninguna experiencia en computación. No hay censura en la calidad de la información de las páginas Web" (*Ibidem*: 6). Categorizar la información bajo un estándar restringiría la libertad actual de la Web.

La estructura de la nueva Web semántica tiene que permitir que las propiedades de universalidad y descentralización de la Web no se pierdan, y al mismo tiempo, permitir que las máquinas entiendan el significado de los documentos Web que procesan.

1.5 La relación entre Web semántica e Inteligencia Artificial

Una de las controversias más importantes es el papel que jugará la Inteligencia Artificial en la Web semántica. A pesar de que se habla que tomarán gran importancia los datos y técnicas inteligentes, en realidad los algoritmos de búsqueda de soluciones de la Web semántica serán bastante sencillos y no tendrán nada que ver con los algoritmos complejos y poderosos que se plantean en el campo de la Inteligencia Artificial. Básicamente, el poder de la Web Semántica se centrará en la forma en que están almacenados los datos, y no en la forma en que se buscan.

De hecho, el mismo creador de la Web, Tim Berners-Lee, afirma que "la Web semántica no es Inteligencia Artificial. El concepto de documentos entendibles por las máquinas no implica alguna inteligencia artificial mágica que permita a las máquinas comprender balbuceos humanos. Sólo indica la habilidad de la máquina para resolver un problema bien definido haciendo operaciones bien definidas sobre datos existentes bien

definidos. En vez de pedir a las máquinas que entiendan el lenguaje de la gente, esto involucra pedirle a las personas hacer un esfuerzo extra", dicho esfuerzo consistirá en estructurar los datos de una manera más ordenada y estandarizada. (Berners-Lee, 1998a)

Podemos concluir que la Web semántica no utilizará los algoritmos de búsqueda de soluciones que provee la Inteligencia Artificial, ya que éstos a pesar de ser muy poderosos necesitan mucho tiempo de procesamiento, sobre todo cuando se aplican a una enorme cantidad de información (hay que recordar que la Web contiene billones de documentos). Sin embargo, la nueva Web se apoyará en conceptos básicos de la Inteligencia Artificial, para lograr una estructuración óptima de la información.

A pesar de la controversia, los investigadores de Inteligencia Artificial siguen investigando al respecto. De hecho, el tema de la Web semántica es parte de libros, conferencias y seminarios de científicos dedicados a esta área de las Ciencias de la Computación. Los conocimientos emanados de estas investigaciones colaborarán de una u otra forma al perfeccionamiento de la nueva Web.

1.6 La evolución hacia la Web semántica

La iniciativa de la Web semántica es una propuesta realmente interesante, sin embargo el camino hacia ella no será tan fácil ni instantáneo, sobre todo por el gran terreno que ya tiene ganado la estructura actual de la Web. No se puede desaparecer la Web actual y hacer que al otro día mágicamente esté instaurada la nueva Web. Tampoco se pueden tener dos Webs (una sintáctica y una semántica) por separado funcionando al mismo tiempo. Por esta razón "es importante entender que la Web semántica no será una nueva autopista de la información paralela a la existente World Wide Web; en lugar de eso evolucionará gradualmente de la Web existente". (Antoniou *et. al.*, 2004: 3)

Este proceso de transición será lento, pero ya se está suscitando. Actualmente la Web semántica es impulsada por la W3C (World Wide Web Consortium), un cuerpo de estandarización internacional para la Web. El mismo creador de la Web, Tim Berners-Lee,

es quien está dirigiendo la evolución hacia la Web Semántica, esperando de esta iniciativa “la realización de su visión original de la Web, una visión donde el significado de la información juegue un rol mucho más importante que como lo hace en la Web de hoy en día”. (*Ibidem*: 3)

El gobierno de Estados Unidos está muy interesado en la Web semántica, sabiendo todas las aplicaciones y ventajas que tendrá una vez lista. Por esa razón está realizando fuertes investigaciones sobre el tema y ha creado el Proyecto DAML (DARPA⁶ Agent Markup Language), que se encarga de estudiar todos los temas relacionados a la Web semántica. (*Ibidem*: 3)

Surge entonces la pregunta de si la Web semántica es únicamente una idea o será una realidad. La respuesta es que “mientras la implementación de la Web semántica en Internet es todavía una visión, los bloques de construcción para la Web semántica están siendo desarrollados en fragmentos pequeños y prototipos. De este modo, las piezas están cayendo en su lugar haciendo de la promesa una realidad”. (Daconta *et. al.*, 2003: 24)

Las piezas que formarán parte de la Web semántica se están creando y están siendo ya utilizadas. Poco a poco, y una vez que su funcionamiento sea verificado, estos elementos se irán integrando como si fueran piezas de un rompecabezas hasta lograr armar un todo, que será la Web semántica.

Hoy podemos afirmar que los bloques de construcción están listos. “Las tecnologías y programas que soporten la Web Semántica están siendo desarrolladas, y las compañías están invirtiendo más dinero en llevar a sus organizaciones a un nivel en el que puedan utilizar estas tecnologías para una ventaja competitiva y monetaria”. (*Ibidem*: 25)

⁶ DARPA (Agencia de Investigación de Proyectos Avanzados de Defensa), es una agencia del Departamento de Defensa de los Estados Unidos encargada del desarrollo de nuevas tecnologías para uso militar. Su principal logro fue la creación de ARPANET, red de computadoras que se convirtió en lo que hoy conocemos como Internet.

1.7 Las fases de la evolución

Es difícil precisar en que momento la Web semántica estará lista para la humanidad. En los últimos años se ha hablado del concepto “versión de la Web”, para señalar qué tanto avance se tiene en el camino hacia la Web de los datos inteligentes. El autor Edward Benson (2008: 3-4) explica claramente estas fases o versiones de la Web.

La Web 1.0 es un sistema en el cual un pequeño grupo produce el contenido (creando páginas o sitios Web), mientras que la mayoría de los usuarios lo consumen pasivamente (simplemente consultan lo que otros crearon). Históricamente, corresponde a la última década del siglo XX.

La Web 2.0 tiene aplicaciones que proveen ambientes en los cuales los usuarios pueden crear y publicar su propio contenido sin tener que crear y mantener sitios Web por sí mismos. Por ejemplo, Blogger, Flickr, Wikipedia, Youtube, Facebook, etc. Ésta es la que se considera la fase actual de la Web, a pesar de que la mayoría de los sitios Web actuales todavía forman parte de la fase anterior. Corresponde a la primera década del siglo XXI.

La Web 3.0 es la implementación de la Web semántica, con documentos que pueden ser procesados y entendidos por las máquinas, junto con sus respectivos motores de búsqueda semánticos, brindando un esquema de razonamiento automatizado a la Web. Su periodo de auge será la segunda década del siglo XXI.

Aunque todavía se considera como un rumor, ya se está hablando de la Web 4.0, en la cual surgirá el concepto de WebOS (Web Operating System, o Sistema Operativo Web), pero todavía no hay mucha información al respecto, aunque se estima que cerca del año 2020 estará dando inicio esta nueva fase.

Sin duda alguna, el futuro de la Web es impredecible, ya que hay muchas propuestas para mejorarla. Pero algo es seguro, año tras año irá mejorando y convirtiéndose en una herramienta más útil para la humanidad.

Arquitectura de la Web semántica

La idea de una Web en donde las máquinas puedan entender el significado de su contenido suena bastante interesante y además sería de gran utilidad para los seres humanos. Pero, ¿cómo se podría construir en forma real esta nueva Web?

Para entender la idea de su construcción se puede hacer una analogía con un pastel de varias capas. La base del pastel es la Web que ya existe, y sobre ella se irán poniendo sucesivamente nuevas capas. En este proceso de crecimiento vertical, cada nueva capa implementada sobre la anterior permitirá un acercamiento mayor a la Web semántica. El diagrama de capas de pastel que se muestra a continuación fue propuesto por Tim Berners-Lee (2000b), y en él se muestran cuáles son las capas que se deben de ir colocando hasta llegar a la meta.

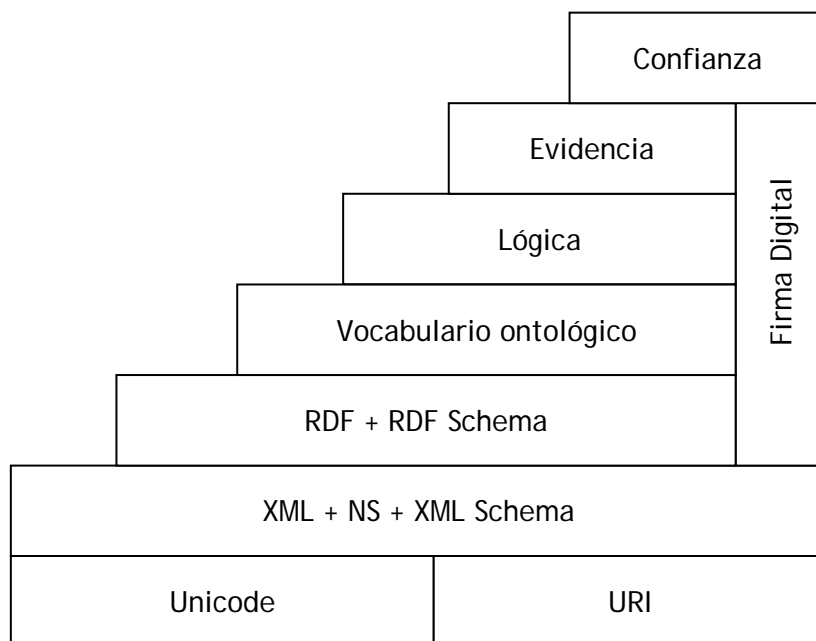


Figura 2.1 "Layer cake" de la Web semántica propuesto por Berners-Lee

Se puede considerar que las capas inferiores constituyen la forma más primitiva pero funcional de la Web semántica (con excepción de la primera), las cuales ya se están implementando en la actualidad. La capa de vocabulario ontológico es el punto en el que se encuentra el desarrollo de la Web semántica actualmente. Las capas superiores equivalen a la Web semántica en su plenitud, aunque éstas todavía no se desarrollan. Cuando el pastel esté completo, entonces tendremos a la Web semántica totalmente lista y al servicio de la humanidad.

En este capítulo, se hablará brevemente de cada una de esas capas, de abajo hacia arriba, para entender el funcionamiento de la Web semántica, y cómo se está gestando el desarrollo de la misma.

2.1 La capa no semántica

La primera de las capas no aporta ninguna funcionalidad semántica a la Web, pero sienta las bases para el funcionamiento de las siguientes capas. Está constituida por dos elementos que detallaremos a continuación.

2.1.1 *Unicode*

La Web semántica debe estar diseñada para permitir su uso por cualquier persona sin importar su origen, borrando cualquier frontera entre naciones y culturas. Por eso, desde su base, debe contemplar todas las posibilidades en un mundo con múltiples idiomas.

Unicode es un estándar de codificación de caracteres, que incluye todos los caracteres de uso común en la actualidad, cada uno de ellos asociado a un identificador numérico único. La versión 5.1 de Unicode incluye un total de 100 713 caracteres, incluyendo alfabetos de todas partes del mundo (europeos, africanos, asiáticos, etc.) tanto modernos como antiguos.

En la siguiente tabla, podemos ver un fragmento muy pequeño de la enorme tabla de Unicode¹. Para formar el identificador de un carácter se toma el número de la columna seguido por el número de renglón. Por ejemplo, al símbolo '<' le corresponde el identificador 003C.

	002	003	004	005	006	007
0	SP	0	@	P	`	p
1	!	1	A	Q	a	q
2	"	2	B	R	b	r
3	#	3	C	S	c	s
4	\$	4	D	T	d	t
5	%	5	E	U	e	u
6	&	6	F	V	f	v
7	'	7	G	W	g	w
8	(8	H	X	h	x
9)	9	I	Y	i	y
A	*	:	J	Z	j	z
B	+	;	K	[k	{
C	,	<	L	\	l	
D	-	=	M]	m	}
E	.	>	N	^	N	~
F	/	?	O	_	o	DEL

Tabla 2.1 Fragmento de la tabla Unicode

¹ El estándar completo de Unicode 5.1 se encuentra en la página <http://www.unicode.org/> en donde es posible ver todos los caracteres que contempla.

2.1.2 URI

La Web está formada por una gran cantidad de recursos (como documentos electrónicos, sitios Web, imágenes, etc.). A cada uno de esos recursos se asocia un identificador llamado URI (Uniform Resource Identifier, o Identificador Uniforme de Recurso), el cual consiste en una serie de caracteres que lo identifican.

Un URI puede funcionar de distinta manera, pudiendo ser:

- ✓ Un URL (Uniform Resource Locator, o Localizador Uniforme de Recurso) que asigna una dirección única a cada uno de los recursos de información. Los navegadores Web utilizan la URL para localizar la dirección del recurso y poder mostrarlo. Por ejemplo, el portal del sitio Web de la UNAM tiene como URL la secuencia de caracteres `http://www.unam.mx/` que nos permite su acceso desde el navegador.
- ✓ Un URN (Uniform Resource Name, o Nombre de Recurso Uniforme) que únicamente identifica a un recurso en la Web (simplemente lo nombra pero no indica donde está localizado). El URN asegura la identificación del recurso aún cuando deje de existir o no se encuentre disponible. (De esta manera, se evita que se asigne a otro recurso el mismo URI).
- ✓ Un URL y un URN al mismo tiempo (combinando ambas funciones).

La diferencia entre URL y URN puede quedar más clara si recordamos que la Web está formada por dos tipos de contenido, el que se encuentra en la superficie, al cual podemos acceder por medio del navegador y por lo tanto tiene un URL, y el contenido profundo, formado por una serie de recursos que no necesariamente podemos visualizar en el navegador, pero que mediante su URN podemos saber de su existencia. (Berners-Lee, Fielding, Irvine y Masinter, 1998b).

Los identificadores URI son la forma en que podemos referirnos a un recurso en la Web, y por ello se establecen como una base para la construcción de la Web semántica.

2.2 La capa XML

XML (eXtensible Markup Language, o Lenguaje de Marcas Extensible) es un metalenguaje, es decir, un lenguaje usado para crear otros lenguajes. Provee una estructura básica y un conjunto de reglas para desarrollar cualquier lenguaje de marcas o etiquetas. Por ejemplo, XML fue utilizado para la creación de XHTML², que es una versión reformulada de HTML. (Alesso y Smith, 2004: 65)

De acuerdo con los autores Vipul Kashyap, Christoph Bussler y Matthew Moran (2008: 23-40), el contenido de la Web semántica está formado por dos partes: los datos y los metadatos. XML es el encargado de representar el contenido de la Web semántica, como veremos a continuación.

2.2.1 Metadatos

La premisa básica de la Web semántica es dar a cada pieza de datos un significado bien definido. Para ello se necesita que los datos se describan a si mismos.

Los metadatos son datos o información acerca de los datos. Por ejemplo, en la Web los metadatos son utilizados para:

- ✓ Describir el esquema en que están almacenados los datos.
- ✓ Almacenar propiedades.
- ✓ Describir o ser un resumen del contenido de la información de los datos.
- ✓ Representar relaciones entre objetos de diferente tipo.

Imaginemos un sitio Web dedicado a cuestiones médicas que almacena una gran colección de imágenes. Una de sus fotografías corresponde a un tumor cerebral, pero el nombre del archivo es "imagen765.jpg". Se podría añadir un metadato que hable sobre el significado del dato (que es en este caso la fotografía), el cual funcionaría como un

² XHTML (eXtensible Hypertext Markup Language, o Lenguaje de Marcas de Hipertexto Extensible) es una aplicación XML. En la versión 1.0, XHTML funciona exactamente igual que HTML. Las diferencias en las siguientes versiones son pequeñas, por ejemplo, en XHTML todas las etiquetas se deben abrir y cerrar, y el nombre de las etiquetas se escribe en minúsculas.

descriptor de nuestra información y contendría la idea “esto es una imagen de un tumor cerebral”.

Los metadatos permiten representar el dominio de información al cual los datos pertenecen. Por ejemplo, la imagen del tumor cerebral pertenece al campo de medicina, ciencias de la salud, etc. Este conocimiento puede ser usado para hacer inferencias acerca de los datos subyacentes, determinar la relevancia o identificar relaciones entre datos almacenados en diferentes sitios Web.

Pero, ¿cómo puede XML ayudarnos en la creación de metadatos?. Imaginemos la siguiente situación. Un sitio Web especializado en eventos culturales cuenta con la siguiente información:

```
“La conferencia Robótica Evolutiva se impartirá en la FES Aragón el 28 de agosto a las 13:00 horas.”
```

Para una computadora lo anterior significa solamente una serie de caracteres válidos, pero no entiende el significado de su contenido. Cobrarán sentido hasta que un ser humano los lea. Pero podríamos agregar algunos metadatos con el objetivo de que una computadora pueda entender a qué se refiere cada uno de los elementos que conforman la oración. Para ello utilizamos etiquetas XML.

En primera instancia, trataremos de que la computadora pueda reconocer que el lugar en el que se realiza el evento es la FES Aragón. Por medio de etiquetas XML podemos asociar al dato “FES Aragón” un metadato que indique que se trata de un lugar.

Las etiquetas en XML están formadas por dos partes: apertura y cierre. La apertura se define entre los símbolos “<” y “>”, y la palabra que está entre esos símbolos es el nombre de la etiqueta. La etiqueta se cierra cuando la volvemos a escribir poniendo una diagonal “/” después del símbolo “<”. Todo lo que está entre una etiqueta de apertura y su respectiva etiqueta de cierre es el dato asociado a la etiqueta. Por ejemplo:

```
<lugar> FES Aragón </lugar>
```

En este caso:

- ✓ El nombre de la etiqueta es "lugar".
- ✓ La etiqueta de apertura es "<lugar>".
- ✓ La etiqueta de cierre es "</lugar>".
- ✓ El dato entre las etiquetas de apertura y cierre es "FES Aragón".
- ✓ El metadato es "lugar".
- ✓ El metadato "lugar" está asociado al dato "FES Aragón", y por lo tanto una computadora puede saber que "FES Aragón" es un "lugar".

Ahora volvamos al ejemplo con la oración completa. Podemos agregar tantas etiquetas como deseemos:

```
<Evento> La <tipo> conferencia </tipo> <nombre> Robótica
Evolutiva </nombre> se impartirá en la <lugar> FES Aragón </lugar>
el <fecha> 28 de agosto </fecha> a las <hora> 13:00 </hora> horas.
</Evento>
```

De esta manera, una computadora podría interpretar la información que recibe, obteniendo la siguiente información:

- ✓ El tipo de evento es una conferencia
- ✓ El nombre del evento es *Robótica evolutiva*
- ✓ El lugar del evento es la FES Aragón
- ✓ La fecha del evento es el 28 de agosto
- ✓ La hora del evento es a las 13:00

Esto no significa que la computadora entienda los conceptos "evento", "lugar", "fecha", "hora", "conferencia", "28 de agosto", etc. Simplemente entiende que "FES Aragón" es un "lugar", o que "28 de agosto" es una "fecha". Pero esto representa el primer paso para que las computadoras entiendan el significado de los datos que procesan.

Para darle mayor orden a la codificación de la información, el ejemplo anterior en XML se debe escribir de la siguiente forma:

```
<?xml version="1.0"?>
<evento>
  <tipo> conferencia </tipo>
  <nombre> Robótica Evolutiva </nombre>
  <lugar> FES Aragón </lugar>
  <fecha> 28 de agosto </fecha>
  <hora> 13:00 </hora>
</evento>
```

La primera línea indica que se trata de un documento escrito en XML y es obligatorio ponerla. Como se puede ver, dentro de una etiqueta se pueden anidar más etiquetas. Todos los datos y metadatos dentro de una etiqueta le pertenecen. Por ejemplo: la etiqueta "lugar" está asociada a la etiqueta "evento" por estar dentro de ella.

2.2.2 Representación del contenido Web semántico mediante el modelo de datos XML

XML permite crear un modelo de datos donde el esquema coexiste con los datos. Esto es muy diferente a los modelos tradicionales de datos utilizados en bases de datos o en programación orientada a objetos, donde el esquema de datos está separado y es independiente de los datos.

Por ejemplo, en bases de datos cuando utilizamos el modelo relacional, un esquema podría definir la relación persona con los atributos nombre y edad. Una instancia de ese esquema podría contener la tupla ("Arturo", "23"). Otra instancia de ese esquema puede contener la tupla ("Cristian", "22"). La relación y los nombres de los atributos están separados de los datos.

Pero en XML, el esquema de la información es almacenado junto con los datos. Los valores estructurados se denominan elementos. Los atributos se llaman etiquetas.

Por ejemplo:

```
<?xml version="1.0"?>
<persona>
  <nombre> Arturo </nombre>
  <edad> 23 </edad>
</persona>
```

En el ejemplo anterior:

- ✓ Las etiquetas son "persona", "nombre" y "edad".
- ✓ Los elementos son "Arturo" y "23".
- ✓ "Arturo" es el "nombre" de una "persona".
- ✓ "23" es la "edad" de una "persona".

De esta manera los datos en XML se describen a sí mismos y se pueden modelar irregularidades que no podrían ser modeladas con datos relacionales u orientados a objetos, o que si se hicieran en esos modelos se consideraría erróneo. Por ejemplo, los ítems pueden tener elementos perdidos; puede haber múltiples ocurrencias de un mismo elemento; los elementos pueden tener valores atómicos en algunos casos o valores estructurados en otros. Las colecciones de elementos pueden tener estructuras heterogéneas. Veamos un ejemplo que contenga irregularidades:

```
<?xml version="1.0"?>
<evento>
  <tipo> Obra teatral </tipo>
  <nombre> Autopsia a un copo de nieve </nombre>
  <lugar> Ciudad Universitaria </lugar>
  <fecha> 12 de agosto </fecha>
  <fecha> 13 de agosto </fecha>
  <fecha> 14/ago </fecha>
  <fecha>
    <dia> 15 </dia>
    <mes> agosto </mes>
  </fecha>
```

```
<fecha>
  <mes> agosto </mes>
  <dia> 16 </dia>
</fecha>
<hora> </hora>
</evento>
```

En el ejemplo anterior hay varias irregularidades:

- ✓ El elemento que corresponde a la etiqueta hora no está.
- ✓ Existen varios elementos que corresponden a una misma etiqueta (fecha).
- ✓ Una misma etiqueta (fecha) contiene en algunos casos datos atómicos (como "14/ago"), y en otros casos datos estructurados (formado por un dato compuesto que incluye a las etiquetas "día" y "mes").

Este tipo de cosas no se ven en los modelos de bases de datos relacionales o en programación orientada a objetos, e incluso se les considera "malas prácticas". Pero en XML están permitidas todas estas irregularidades.

Este tipo de datos se han denominado "datos semiestructurados", y se caracterizan por ser extremadamente flexibles. Su uso está motivado por los problemas de integrar recursos de datos heterogéneos, por ello es la mejor opción para modelar recursos de datos en la Web.

En el ejemplo anterior también podemos notar que el orden de los elementos dentro de un mismo nivel jerárquico no importa. En el caso de las fechas que contienen datos estructurados podemos ver que en un caso se muestra primero el día y después el mes, y en el siguiente caso se muestra en forma invertida.

El aspecto más importante de XML es tener cuidado con el orden jerárquico de los elementos. Un documento XML puede ser esquematizado como un árbol, en donde las etiquetas se representan mediante un círculo, y los datos se representan con un rectángulo. Por ejemplo, a continuación se muestra un documento XML y su respectivo árbol:

```
<?xml version="1.0"?>
<evento>
  <tipo> Muestra fotográfica </tipo>
  <nombre> Rostros del olvido </nombre>
  <lugar> FES Aragón </lugar>
  <fecha>
    <dia> 10 </dia>
    <mes> marzo </mes>
  </fecha>
</evento>
```

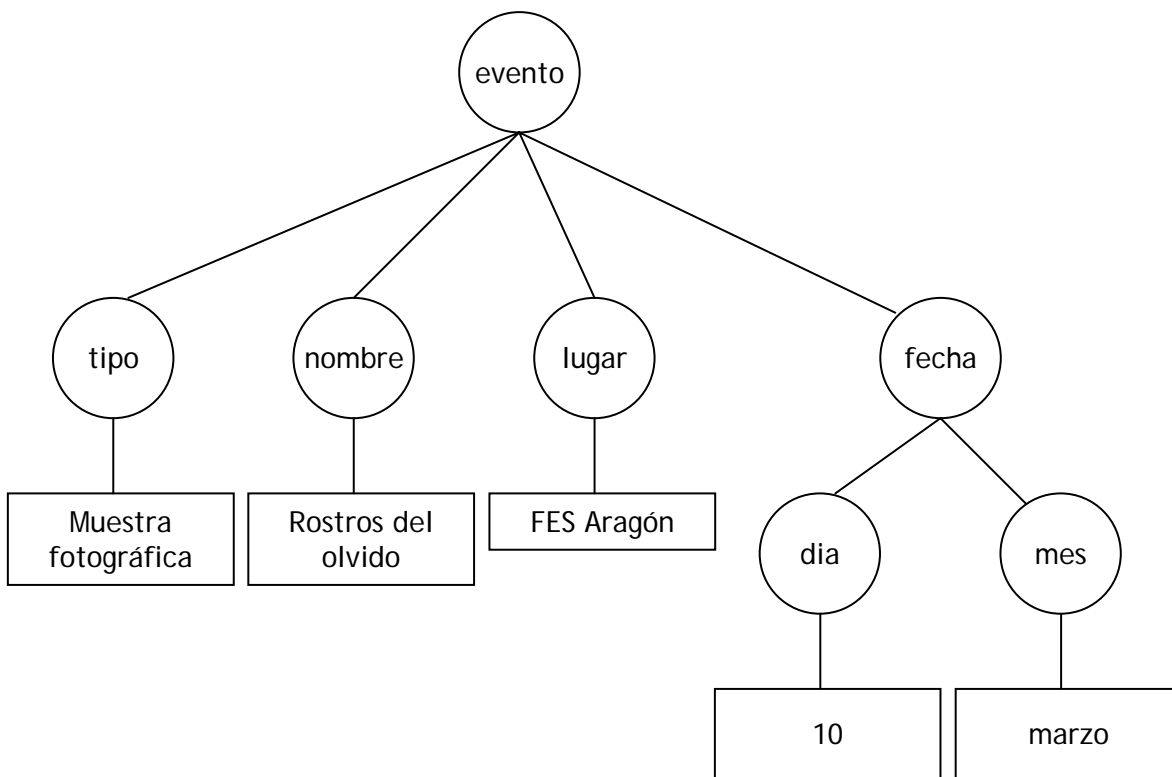


Figura 2.2 Ejemplo de un árbol XML

2.2.3 XML Namespaces

En la Web semántica queremos catalogar un enorme número de recursos distribuidos en todo el mundo, codificado en diferentes lenguajes, por diferentes grupos (Breitman *et. al.*, 2007: 7). Pero el hecho de la participación sin restricciones de cualquier persona en la conformación de la Web puede originar ciertos problemas.

Las etiquetas en XML son definidas por el desarrollador. Cada quien asigna el nombre de etiqueta que quiere, pero esto puede ser conflictivo ante la presencia de una gran cantidad de información donde el mismo nombre de etiqueta puede ser implementado por dos personas diferentes. Imaginemos la siguiente situación para ejemplificar lo anterior:

Un sitio Web de una revista de moda llamada "*Tendency*" permite a los usuarios consultar algunos de los artículos publicados en ejemplares anteriores. El creador del sitio Web utilizó el modelo de datos XML, y a continuación veremos cómo es el documento XML en el que se almacena información sobre uno de los artículos:

```
<?xml version="1.0"?>
<articulo>
  <titulo> Moda para la temporada de invierno </titulo>
  <autor> Gadiel Carmona </autor>
  <pagina> 26 </pagina>
</articulo>
```

Por otra parte, un sitio Web dedicado a la venta de juguetes llamado "ToyWorld" también fue modelado con XML y a continuación se muestra el documento XML que contiene información acerca de uno de sus artículos en venta:

```
<?xml version="1.0"?>
<articulo>
  <nombre> oso de peluche gigante </nombre>
  <precio> 1500 </precio>
</articulo>
```


Los dos sitios Web anteriores fueron creados por diferentes personas. Pero en ambos casos se utilizó la etiqueta "artículo". El significado de la etiqueta "artículo" es distinto de acuerdo al contexto del sitio Web al que pertenece. En el caso de la revista de moda, el "artículo" se refiere a una nota o reportaje contenido en la revista. En el caso de la tienda virtual de juguetes "artículo" corresponde a un producto en venta. Esta diferencia la entiende rápidamente un ser humano, pero ¿cómo lograríamos hacer que las computadoras distingan la diferencia entre una misma etiqueta que significa diferentes cosas dependiendo del contexto?.

La solución es el uso de XML Namespaces, que consiste en almacenar junto con el nombre de la etiqueta una referencia que nos indique a qué contexto pertenece esa etiqueta. El contexto lo señalamos indicando a qué sitio Web o recurso en la Web está asociado dicha etiqueta, y esto se lo indicamos mediante el URI del recurso que brinda el contexto adecuado. El objetivo de XML Namespaces es evitar conflictos entre los nombres asignados en las etiquetas.

El uso de Namespaces es un método para crear nombres de etiquetas únicos a nivel global para evitar colisiones entre elementos que usan el mismo nombre pero con diferentes significados. (Geroimenko, 2004: 99)

La declaración de un Namespace se hace dentro de la etiqueta que queremos asociar a un contexto, y seguimos la siguiente sintaxis:

```
xmlns="URI"
```

Por ejemplo, para el caso de la revista de moda, modificaríamos el código de la siguiente forma:

```
<articulo xmlns="http://www.tendency.com/">  
  <titulo> Moda para la temporada de invierno </titulo>  
  <autor> Gadiel Carmona </autor>  
  <pagina> 26 </pagina>  
</articulo>
```

Y en el otro caso:

```
<articulo xmlns="http://www.toyworld.com/" >
  <nombre> oso de peluche gigante </nombre>
  <precio> 1500 </precio>
</articulo>
```

De esta manera podemos garantizar que aunque en ambos casos se utilice la misma etiqueta ("artículo") no significa lo mismo. Cabe señalar que el Namespace afecta a todas las etiquetas anidadas. Por ejemplo, en el último caso, "nombre" y "precio" también están dentro del contexto del recurso cuyo URI es "http://www.toyworld.com/".

2.2.4 XML Schema

La Web semántica estará formada por un gran conjunto de documentos XML, en el que existirán subconjuntos de documentos que tengan características similares. Por ejemplo, un sitio Web de una librería puede trabajar con documentos XML en el que se almacene información sobre los libros. Uno de esos documentos se muestra a continuación:

```
<?xml version="1.0"?>
<libro>
  <titulo> Harry Potter y la piedra filosofal </titulo>
  <autor> J. K. Rowling </autor>
  <editorial> Salamandra </editorial>
  <edicion> 13 </edicion>
  <precio> 139.50 </precio>
</libro>
```

Pero es obvio que esta librería trabaja con miles de libros, y por lo tanto, todos los documentos XML tienen una estructura similar, que contiene las mismas etiquetas. Lo único que cambia son los valores de los datos almacenados en el documento. Veamos otro ejemplo:

```
<?xml version="1.0"?>
<libro>
  <titulo> El Código Da Vinci </titulo>
  <autor> Dan Brown </autor>
  <editorial> Umbriel </editorial>
  <edicion> 2 </edicion>
  <precio> 345.00 </precio>
</libro>
```

Un conjunto de documentos similares se puede procesar más fácilmente si se estandariza la forma en que debe estar estructurado.

De acuerdo con el autor Eric Van der Vlist (2002) un esquema XML es una formalización de las restricciones, expresadas como reglas o un modelo de estructura, que aplica a una clase de documentos XML. Los esquemas XML tienen las funciones de validar y documentar.

La validación de documentos XML es la función principal de los esquemas. Cuando una aplicación Web recibe información en formato XML, el esquema XML permite verificar que el documento XML sea correcto. Comparar un documento contra un esquema permite asegurar que el contenido del documento se ajuste a lo que se espera, simplificando el código necesario para procesar la información contenida en el documento. La validación de documentos reduce el riesgo de procesar documentos XML que sean incorrectos y que puedan causar problemas al momento de ser procesados por una aplicación. Existen dos tipos de validación:

- ✓ La validación estructural señala qué elementos y atributos debe contener un documento, así como su jerarquía.
- ✓ La validación de datos verifica el contenido de la estructura, asegurándose que los datos contenidos sean de determinado tipo (entero, flotante, cadena, etc.).

Documentar vocabularios XML es otra función de los esquemas XML, ya que indican cuáles son las etiquetas que puede contener un documento. De esta manera limitamos y

hacemos uniforme el vocabulario que pueden tener los documentos que se asocian a un determinado esquema.

Una de las ventajas de los esquemas XML es que están escritos en lenguaje XML, por lo que no es necesario aprender un lenguaje distinto.

En los ejemplos de documentos XML de la librería podemos identificar que para cada libro se necesita de cinco datos importantes: título, autor, editorial, edición y precio. Los tres primero corresponden a un tipo de dato cadena, la edición únicamente puede tomar números enteros, pero el precio puede tomar números con punto decimal.

El esquema XML que corresponde a los documentos utilizados en la librería virtual es el siguiente:

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
<xs:element name="libro">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="titulo" type="xs:string"/>
      <xs:element name="autor" type="xs:string"/>
      <xs:element name="editorial" type="xs:string"/>
      <xs:element name="edicion" type="xs:integer"/>
      <xs:element name="precio" type="xs:decimal"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
</xs:schema>
```

Se necesita una aplicación que pueda comparar un documento XML con un esquema para verificar si es válido. Si el documento es válido, entonces es aceptado y se somete a procesamiento. Si no es válido, el documento se desecha y se devuelve un mensaje de error. Esto se ilustra en el siguiente esquema:

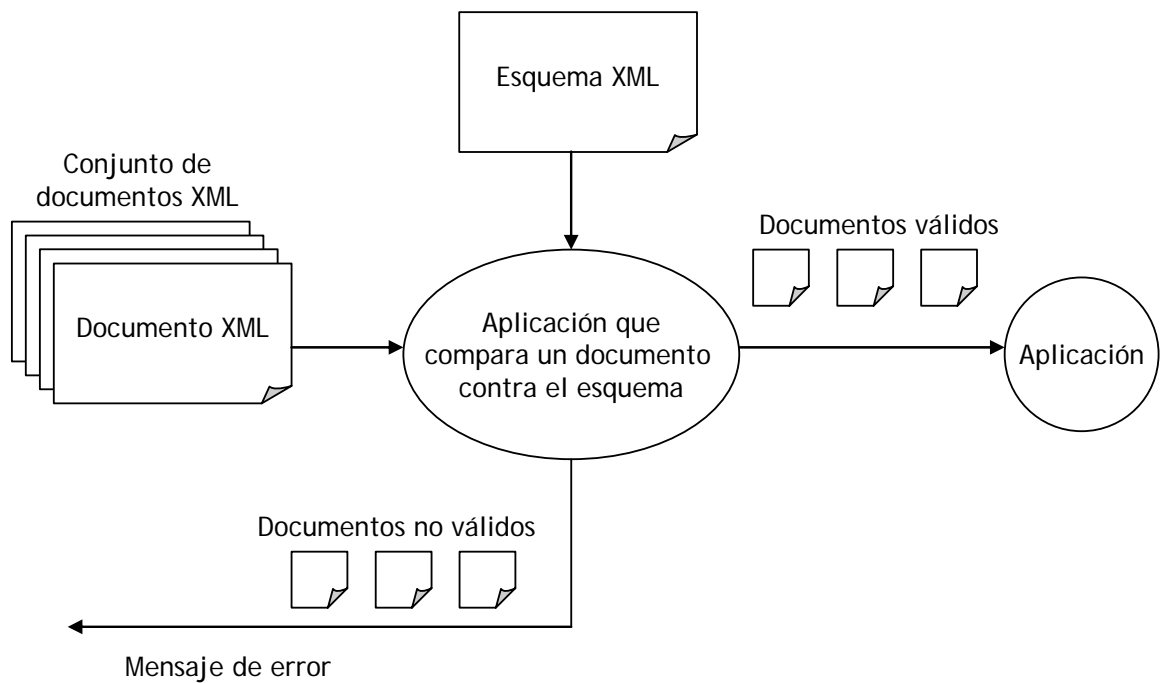


Figura 2.3 Proceso de validación de documentos XML

2.2.5 XML y el consumo humano de la Web

En este punto hemos resuelto un problema y creado otro. XML permite el almacenamiento, estructuración e intercambio de la información, tanto de datos como de metadatos. Los documentos XML están diseñados para ser procesados y entendidos por las máquinas, lo que constituye un primer paso hacia la Web semántica. Pero los documentos XML no están hechos para el consumo humano.

Imaginemos que un niño está investigando información acerca del horario de sus caricaturas favoritas, y entra a un sitio Web donde recibe un documento XML, cuyo contenido es desplegado en el navegador:

```
<?xml version="1.0"?>
<programa>
  <titulo> Bob esponja </titulo>
  <hora> 3:00 </hora>
</programa>
```

Alguien con conocimientos básicos en XML puede extraer la información, pero un usuario común como el niño pensará que la página no sirve e irá a otro sitio para buscar información. Y esto sucede en todos los casos en que se intente visualizar un documento XML desde el navegador. Lo único que veremos es su código, tal y como está escrito.

Afortunadamente ya se han creado una serie de tecnologías que permiten la transformación de un documento XML a un formato adecuado para los humanos. Por ejemplo: Flash, Java, HTML/XHTML, XSLT, SVG, etc. (Geroimenko y Chen, 2003: 12)

Una de esas tecnologías es XSL (eXtensible StyleSheet Language, o Lenguaje de Hojas de Estilo Extensible). XSL es usado en conjunción con HTML para describir cómo los datos en XML serán mostrados en un navegador Web. (Alesso *et. al.*, 2004: 65)

Veamos un ejemplo sencillo de un documento XML que está diseñado para ser entendido tanto por las máquinas como por los humanos. En esta ocasión el documento contiene información acerca de un disco de música, describiendo el contenido (nombre de las canciones incluidas junto con su respectiva duración):

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="programa.xsl"?>
<cd>
  <cancion>
    <titulo> Tus ojos </titulo>
    <duracion> 2:58 </duracion>
  </cancion>

  <cancion>
    <titulo> Por esta vez </titulo>
    <duracion> 3:42 </duracion>
  </cancion>

  <cancion>
    <titulo> Escena final </titulo>
    <duracion> 2:34 </duracion>
  </cancion>

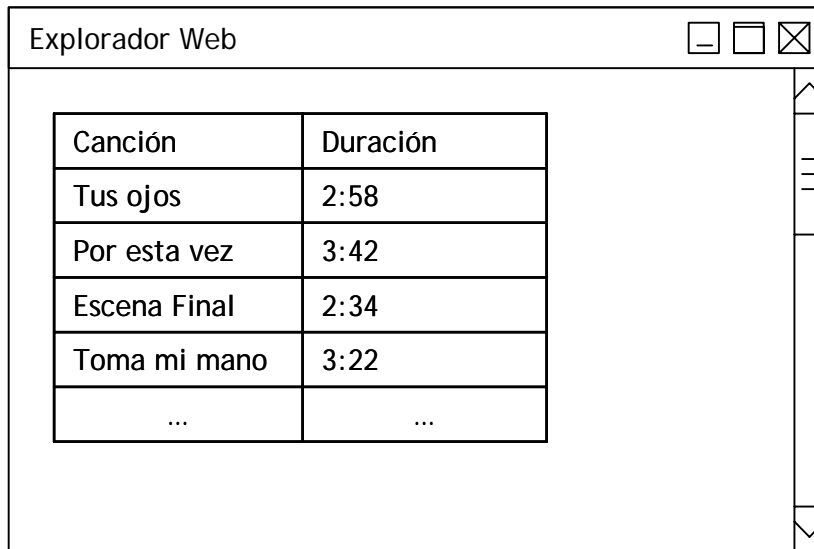
  <cancion>
    <titulo> Toma mi mano </titulo>
    <duracion> 3:22 </duracion>
  </cancion>

  ...
</cd>
```

En la segunda línea de código se hace una declaración que indica que este documento XML está asociado a una hoja de estilo XSL, la cual se encargará de dar un formato al contenido de modo que pueda ser desplegado en el navegador Web y pueda ser entendido por un humano. La hoja de estilo es el documento *programa.xsl*, cuyo contenido se muestra a continuación:

```
<?xml version="1.0"?>
<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:template match="/">
  <html>
  <body>
    <table>
    <tr>
      <th>Canción</th>
      <th>Duración</th>
    </tr>
    <xsl:for-each select="cd/cancion">
    <tr>
      <td><xsl:value-of select="titulo"/></td>
      <td><xsl:value-of select="duracion"/></td>
    </tr>
    </xsl:for-each>
    </table>
  </body>
</html>
</xsl:template>
</xsl:stylesheet>
```


Para entender el ejemplo anterior se requieren conocimientos básicos de HTML. Pero la explicación de lo que está haciendo es sencilla. El contenido del documento XML es mostrado en el navegador en forma de una tabla. El usuario vería esto en el navegador:



Canción	Duración
Tus ojos	2:58
Por esta vez	3:42
Escena Final	2:34
Toma mi mano	3:22
...	...

Figura 2.4 Despliegue en el navegador utilizando XSL

Usando XSL se pueden crear páginas HTML visualmente muy atractivas con base en información contenida en documentos XML. Además, una misma hoja de estilo XSL se puede usar para muchos documentos que tengan un esquema similar.

En la nueva Web, XML y HTML representarán el dúo perfecto para la creación de documentos entendibles tanto por humanos como por computadoras.

2.3 La capa RDF

Con XML es posible describir y estructurar los datos que forman parte de los diversos recursos existentes en la Web. Pero, ¿cómo podemos crear información que describa los recursos de la Web y sus relaciones?. La solución está en la implementación de la capa RDF de la Web semántica.

RDF (Resource Description Framework, o Marco de Descripción de Recursos) es un estándar para representar información acerca de los recursos en la Web. Los documentos RDF están escritos en XML, y a este lenguaje se le denomina RDF/XML. Su objetivo es describir la información acerca de recursos Web de un modo estandarizado, de modo que pueda ser leído y entendido por las máquinas. RDF utiliza el URI de un recurso para poder referirse a él. (Powers, 2003: 1-13)

2.3.1 Descripción de recursos Web mediante RDF

Veamos un ejemplo sencillo de un documento RDF/XML que describe a una página Web cuyo URI es *http://www.mipagina.com* :

```
<?xml version="1.0"?>
<RDF>
  <datos about="http://www.mipagina.com">
    <autor> Arturo Rodríguez </autor>
    <fecha_de_creacion> 30/jun/09 </fecha_de_creacion>
  </datos>
</RDF>
```

El documento anterior nos brinda dos datos que describen al recurso Web: su autor, y la fecha de creación.

De acuerdo con Shelley Powers (2003: 14-28) una oración RDF está formada por tres partes:

- ✓ Un sujeto, que puede ser cualquier recurso Web, y se indica mediante su URI. En el código anterior, el sujeto es *http://www.mipagina.com*.
- ✓ Un predicado, que es una característica específica o relación que describe a un recurso. En el ejemplo anterior "autor" y "fecha_de_creacion" son predicados.
- ✓ Un objeto, que es otro elemento con quien el sujeto establece una relación mediante el predicado y que puede ser un dato o un recurso Web. En el ejemplo hay dos objetos, "Arturo Rodríguez" y "30/jun/09", los cuales son datos.

En el ejemplo de documento RDF podemos identificar dos oraciones, que si tratáramos de expresar en lenguaje natural las leeríamos de la siguiente forma:

- ✓ La página *http://www.mipagina.com* tiene como *autor* a *Arturo Rodríguez*.
- ✓ La página *http://www.mipagina.com* tiene como *fecha de creación* el *30/jun/09*.

Formalmente, una oración RDF se expresa como una 3-tupla de la forma (sujeto,predicado,objeto). Los predicados se deben anteceder del sujeto al que están asociados y un símbolo '#'. Por ejemplo, las oraciones anteriores se expresan formalmente como(*http://www.mipagina.com*, *http://www.mipagina.com#autor*, Arturo Rodríguez) y (*http://www.mipagina.com*, *http://www.mipagina.com#fecha_de_creacion*, 30/jun/09).

Un documento RDF se puede representar mediante un grafo RDF, en donde los sujetos y objetos se representan como nodos, y los arcos representan los predicados. El grafo correspondiente al documento RDF del ejemplo es:

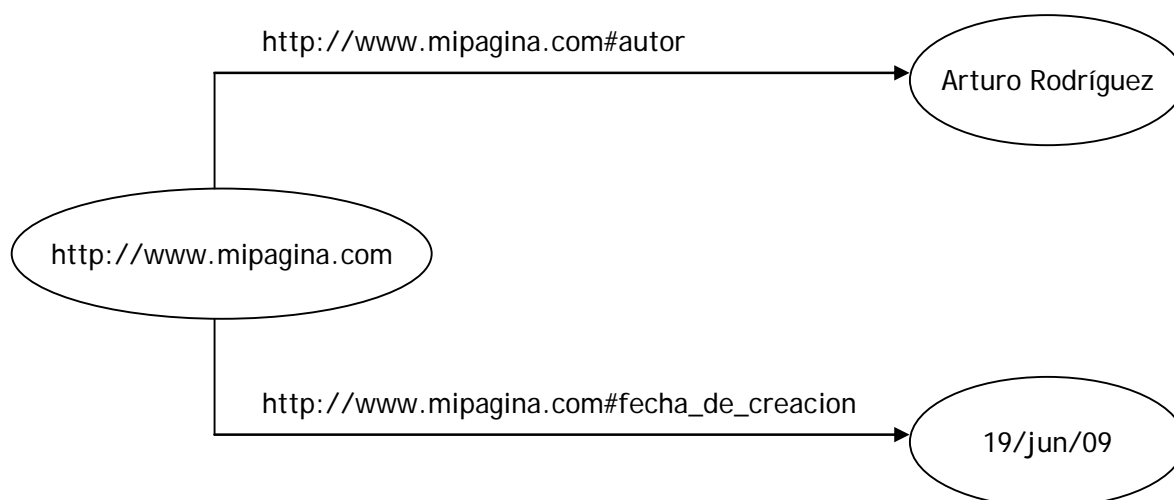


Figura 2.5 Ejemplo de grafo RDF

2.3.2 Establecimiento de relaciones entre recursos Web mediante RDF

Cuando se describieron las partes de una oración RDF se mencionó que un sujeto se asocia a un objeto mediante un predicado. Pero los objetos pueden ser tanto un dato como cualquier recurso en la Web. De esta manera, RDF además de permitir la descripción de un recurso, permite establecer relaciones entre los recursos Web.

Imaginemos la siguiente situación. Una página Web contiene información acerca del clásico videojuego *Mario Bros 64*, su URI es `http://mariobros64.com`. En otro sitio Web totalmente independiente en su creación al anterior se describe la consola *Nintendo 64*, y su URI es `http://nintendo64.com`. Estos dos sitios que fueron desarrollados en forma independiente pueden establecer una conexión mediante un documento RDF. En este ejemplo, el documento RDF del sitio Web del videojuego y su respectivo grafo son:

```
<?xml version="1.0"?>
<RDF>
  <info about="http://mariobros64.com">
    <consola> http://nintendo64.com </consola>
  </info>
</RDF>
```

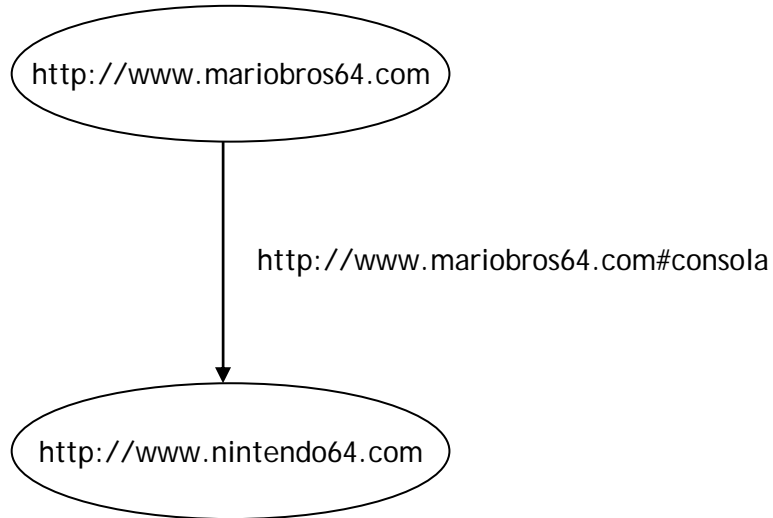


Figura 2.6 Grafo RDF en el que se relacionan dos recursos Web

RDF es una herramienta sumamente poderosa para la Web semántica, ya que permitirá establecer relaciones entre recursos Web de modo que las máquinas puedan entenderlas.

2.3.3 RDF Schema

La Web es una jungla virtual llena de todo tipo de recursos e información. Pero dentro de esta gran cantidad de documentos y sitios Web, podemos distinguir grupos de recursos que comparten características similares.

RDF Schema es un mecanismo para describir grupos de recursos Web afines, así como las relaciones entre estos recursos. Se le considera un lenguaje de descripción de

vocabulario. Este vocabulario ayuda a las aplicaciones a entender cómo se deben interpretar las oraciones RDF (Lacy, 2005: 111-112). Además, los esquemas RDF cumplen con la función de restringir el uso de un documento RDF a un lenguaje limitado intencionalmente. (Berners-Lee, 1998a).

Para conformar un esquema RDF se tienen que utilizar los conceptos de clases y subclases. Una clase se refiere a un conjunto de recursos con características similares. Una subclase es un subconjunto de recursos de la clase que cumplen una condición determinada. Gráficamente, se representa de la siguiente forma:

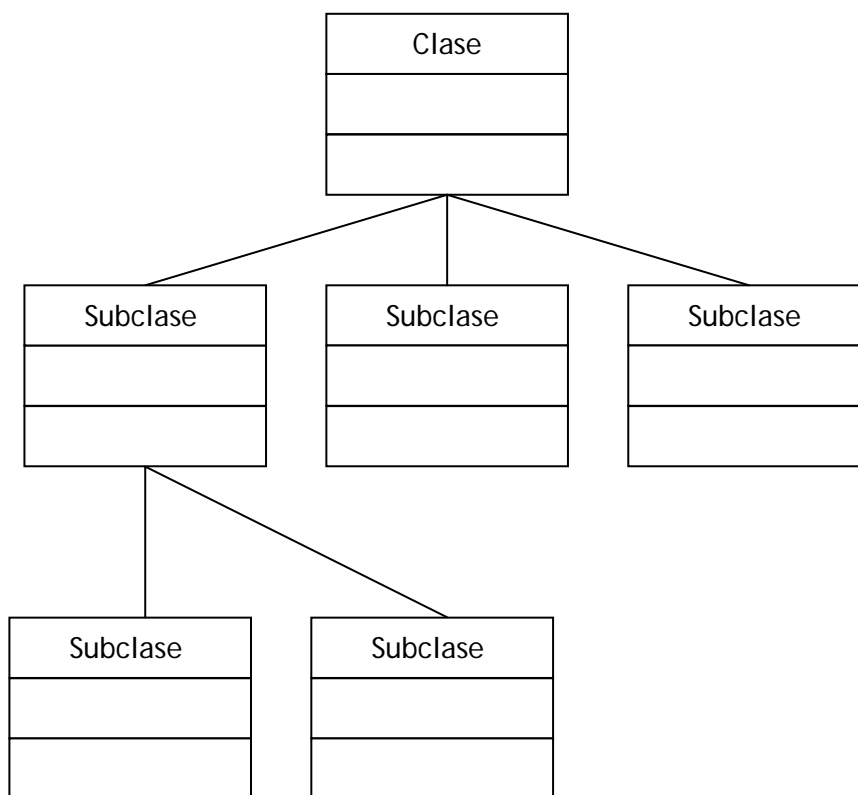


Figura 2.7 Representación gráfica de clases y subclases en RDF Schema

Por ejemplo, podemos tomar como clase al conjunto de los animales, y como subclase al conjunto de vertebrados (que cumplen la condición de tener huesos). A continuación crearemos el esquema RDF de esta situación.

Un esquema RDF tiene la siguiente estructura:

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
    ...
</rdf:RDF>
```

El esquema RDF utiliza los XML Namespaces llamados `rdf` y `rdfs`, por lo que se declaran desde el principio. El lugar donde están los puntos suspensivos es donde se definirá el vocabulario admitido por el esquema RDF. Primero se define la clase, y se hace de la siguiente manera:

```
<rdf:Description rdf:ID="Animal">
    <rdf:type rdf:resource=
        "http://www.w3.org/2000/01/rdf-schema#Class"/>
</rdf:Description>
```

La forma de declarar una clase es siempre la misma, sólo hay que ponerle el nombre deseado en el identificador de la clase (*ID*). Una vez declarada la clase se puede proceder a declarar subclases, y siempre se utiliza el siguiente segmento de código (sólo se tiene que modificar el *ID* para nombrar a la subclase del modo deseado, y el *resource*, que indica la clase de la que se deriva la subclase):

```
<rdf:Description rdf:ID="Vertebrado">
<rdf:type rdf:resource=
"http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#Animal"/>
</rdf:Description>
```

A pesar de que la notación de RDF luce complicada en un principio, es realmente fácil de manejarla, ya que todas las declaraciones de clases y subclases siguen exactamente el mismo formato. Se puede declarar una subclase de otra subclase. Por ejemplo, si queremos declarar la subclase felinos de la subclase vertebrados lo haríamos de esta manera (observar que se trata del mismo segmento de código anterior, lo único modificado es lo que está en negritas):

```
<rdf:Description rdf:ID="Felino">
  <rdf:type rdf:resource=
    "http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#Vertebrado" />
</rdf:Description>
```


2.4 La capa ontológica

El término “ontología” proviene de Filosofía. La Ontología es una rama de la Metafísica interesada en la identificación de los tipos de cosas que existen y cómo describirlos (Antoniou *et. al.*, 2004: 10). Provee una clasificación exhaustiva de los objetos y sus relaciones.

La Inteligencia Artificial fue la primera rama de Ciencias de la Computación que adoptó el término de ontología. Una ontología describe formalmente un dominio de discurso. Consiste en una lista finita de términos y las relaciones entre esos términos. Los términos denotan conceptos importantes (o clases de objetos) del dominio. (*Ibidem*: 10)

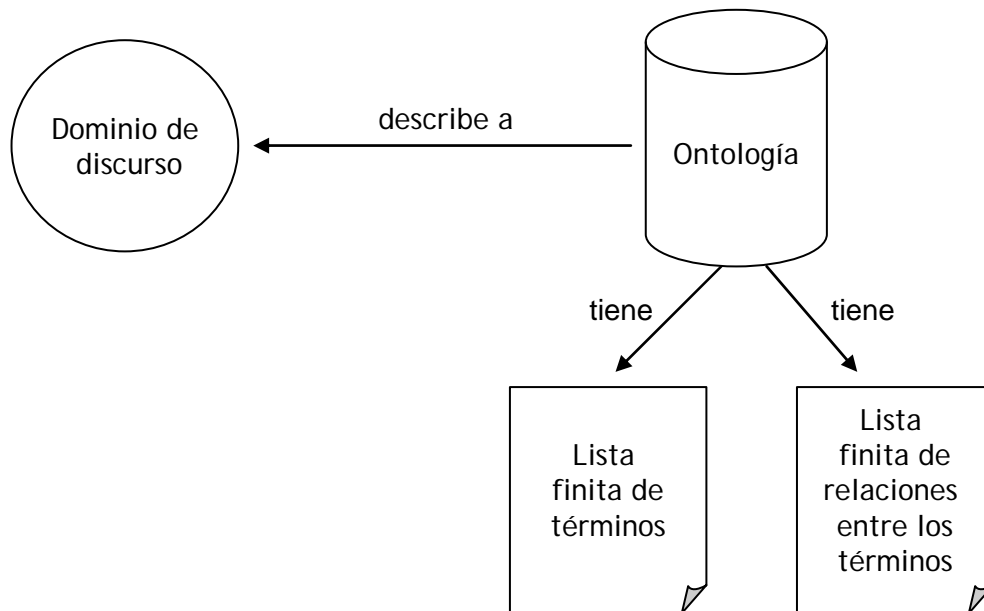


Figura 2.8 Definición de ontología

El objetivo de una ontología es capturar y hacer explícito el vocabulario usado en aplicaciones semánticas, garantizando de esta manera una comunicación libre de ambigüedades. (Breitman *et. al.*, 2007: 8)

Un dominio es un área de conocimiento específica que es centro de atención de una comunidad de interés. Una ontología define los términos usados para describir y representar un dominio.

2.4.1 Elementos de una ontología

Para entender mejor el concepto de ontología imaginemos una situación muy sencilla: deseamos crear una ontología acerca de comida en restaurantes.

Lo primero que debemos identificar es el dominio, el cual es "comida en restaurantes". De un dominio se debe generar una lista finita de términos: "comida", "entrada", "guisado", "postre", "botana", "bebida", "ensalada", "agua", etc. La lista resultante de términos es realmente extensa, por lo que enlistar de esta manera no tiene ningún sentido. Es conveniente hacer una agrupación de términos, de modo que se facilite la tarea de enlistar.

Dentro de la lista de términos, existen grupos de ellos que tienen características comunes. Éstos se agrupan en clases. Por ejemplo, los términos "café", "té", "agua", "refresco", se pueden agrupar dentro del concepto "bebida".

Una clase es un término que representa un grupo de objetos con propiedades similares. En el caso anterior, la clase es "bebida". Un individuo es un término que forma parte de una clase. Los términos "café", "té", "agua" y "refresco" son individuos de la clase "bebida".³

Hay que recordar que una ontología provee una clasificación exhaustiva de los objetos, así que se tiene que detallar tanto como sea posible.

Los términos (ya sean clases o individuos) pueden estar relacionados. Lee W. Lacy (2005: 36-39) identifica los cinco tipos de relación más importantes entre términos, los cuales se muestran en la siguiente tabla:

³ En la creación de ontologías se utiliza el término "individuo" para referirse a una instancia de una clase. Un individuo en Ontología es el equivalente a un objeto en Programación Orientada a Objetos.

Tipo de relación	Descripción	Ejemplo
Sinonimia	Especifica que dos términos son similares.	"refresco" es lo mismo que "soda"
Antonimia	Especifica que dos términos son contrarios.	"bebida con alcohol" es lo contrario a "bebida sin alcohol"
Hiponimia	Indica jerarquía (un término subordinado a otro)	"pastel de chocolate" es un "postre"
Meronomia	Indica agregación (un término está compuesto de varias partes)	"capuchino" está formado por "café", "leche", "chocolate" y "canela"
Holonimia	Indica composición (un término es parte de otro)	"café" es parte de "capuchino" "chocolate" es parte de "capuchino"

Tabla 2.2 Principales tipos de relación entre términos

Como se puede notar, las relaciones en una ontología no son solamente jerárquicas, sino que existen términos equivalentes, opuestos, etc. Esto permite dar mayor riqueza a la relación entre términos. Es necesario definir todas las relaciones existentes. Un término puede estar relacionado con cuantos términos sea necesario, formando una red conceptual de gran complejidad.

La definición de una ontología no es una tarea sencilla. Se necesita la participación de personas que tengan amplios conocimientos sobre el dominio para poder crear una ontología del mismo. Ante la complejidad de esta tarea se ha propuesto el concepto de repositorios de ontologías, que son sitios Web en donde se almacenan las ontologías ya creadas para que puedan ser utilizadas por otros grupos de trabajo sin tener que reescribirlas nuevamente. ⁴

⁴ Uno de los sitios que contiene repositorios de ontologías es <http://www.schemaweb.com>, donde se pueden encontrar ejemplos de ontologías sobre todo tipo de temas: tiempo, zoología, educación, aeropuertos, viajes, etc. Sin embargo, cualquier persona puede colaborar en la elaboración de ontologías en este sitio, por lo que no es una fuente 100% confiable.

2.4.2 OWL

OWL (Web Ontology Language⁵, o Lenguaje de Ontología de la Web) permite la codificación de ontologías que puedan ser entendidas por las máquinas. Permite representar en forma explícita los términos del vocabulario, así como la relación entre los términos.

Utilizando RDF Schema se pueden jerarquizar los tipos de recursos, por medio de clases y subclases. Pero al hablar de ontologías vimos que existen otros tipos de relación entre recursos, por ejemplo, de equivalencia, de oposición, etc. OWL es más poderoso que RDF Schema porque permite describir otros tipos de relaciones.⁶

La creación de una clase en OWL es muy sencilla. Por ejemplo, si queremos crear una clase de nombre "continente", lo hacemos de la siguiente forma:

```
<owl:Class rdf:ID="continente">  
</owl:Class>
```

Ya que en una ontología se deben listar todos los términos permitidos, OWL permite hacer un listado de todos los individuos que pueden derivarse de una clase. Por ejemplo, la clase anterior puede restringir su vocabulario de la siguiente manera:

⁵ Como dato curioso se puede mencionar que las siglas de Web Ontology Language deberían ser WOL, pero cuando fue creado se decidió ponerle OWL (o búho en español) en honor al personaje del mismo nombre que aparece en *Winnie de Pooh* (1926), libro infantil que fue escrito por el gran ontólogo Alan Alexander Milne (1882-1956). En cierta forma, Búho es una representación de una ontología, ya que es el personaje del libro que sabe todos los conceptos y las relaciones entre ellos, corrigiendo y enseñando constantemente a sus ignorantes amigos. (Hendler, 2008)

⁶ De hecho, en varios libros se considera a RDF Schema como si fuera una capa ontológica primitiva. Sin embargo, RDF no permite crear una verdadera ontología, ya que el tipo de relaciones que se pueden crear es muy restringido (únicamente relaciones de tipo jerárquico).

```

<owl:Class rdf:ID="continente">
  <owl:oneOf rdf:parseType="Collection">
    <owl:Thing rdf:about="#America"/>
    <owl:Thing rdf:about="#Europa"/>
    <owl:Thing rdf:about="#Asia"/>
    <owl:Thing rdf:about="#Africa"/>
    <owl:Thing rdf:about="#Oceania"/>
  </owl:one of>
</owl:Class>

```

De esta manera, existen únicamente cinco individuos que pueden pertenecer a la clase "continente", y son "América", "Europa", "Asia", "África" y "Oceanía".

La creación de una subclase es también muy sencilla. Por ejemplo, si queremos crear la subclase "círculo" lo declaramos de la siguiente manera (suponiendo que una clase llamada "figura" ya está creada):

```

<owl:Class rdf:ID="circulo">
  <rdfs:subClassOf rdf:resource="#figura">
</owl:Class>

```

Dos clases pueden ser equivalentes aunque tengan distinto nombre (sinonimia). La clase "empleado" es lo mismo que la clase "trabajador". Esto lo codificamos en OWL de esta manera:

```

<owl:Class rdf:about="#empleado">
  <equivalentClass rdf:resource="#trabajador"/>
</owl:Class>

```

Con OWL podemos indicar que una clase es opuesta a otra (antonimia). Por ejemplo, si queremos indicar que la clase "empleado" es opuesta a la clase "desempleado", lo hacemos de esta forma:

```
<owl:Class rdf:ID="empleado">
  <owl:complementOf>
    <owl:Class rdf:about="#desempleado"/>
  </owl:complementOf>
</owl:Class>
```

Aunque sólo se mostraron algunos casos básicos, con los ejemplos anteriores podemos notar que OWL es un lenguaje muy poderoso para describir los conceptos y sus relaciones. La lista de posibles relaciones y estructuras que podemos crear con OWL es muy extensa y se requeriría de muchas páginas para ejemplificarlas todas.

2.5 Las capas aún no desarrolladas

Las capas más altas de la Web semántica están todavía lejos de ser completamente desarrolladas. En esta sección se describe brevemente en qué consiste cada una de ellas.

2.5.1 La capa lógica

Teniendo como base las ontologías ya creadas, esta capa contendrá reglas lógicas que permitirán a las computadoras hacer inferencias que deriven en nuevo conocimiento.

Por medio de ontologías se puede describir un dominio, que tiene un conjunto finito de términos y un conjunto finito de relaciones entre los términos. Pero haciendo uso de la lógica, se puede obtener muchas más información a partir de la ontología.

Por ejemplo, imaginemos que una ontología que describe el dominio "comida de restaurantes" provee la siguiente información:

1. "Soda" es lo mismo que "refresco".
2. "Refresco" es una "bebida".

En la capa lógica se necesita de un lenguaje que traduzca la información anterior a proposiciones como las siguientes:

p: Es una soda

q : Es un refresco

r: Es una bebida

p -> q Si es una soda entonces es un refresco

q -> r Si es un refresco entonces es una bebida

Una de las reglas de inferencia para proposiciones es el silogismo hipotético, el cual al aplicarlo a las proposiciones anteriores genera el siguiente resultado:

$p \rightarrow q$	Si es una soda entonces es un refresco
<u>$q \rightarrow r$</u>	<u>Si es un refresco entonces es una bebida</u>
$\therefore p \rightarrow r$	Si es una soda entonces es una bebida

Aunque este ejemplo es sencillo, la lógica provee muchas reglas de inferencia para proposiciones que pueden ser utilizadas. De esta manera, las computadoras pueden obtener nueva información a partir de la información que ya tienen disponible.

Aunque esta capa no se ha implementado, se ha propuesto a *Prolog*⁷ como el lenguaje de programación que será utilizado para dotar a la Web semántica de lógica.

2.5.2 La capa de evidencia

La capa de evidencia o prueba, implementará lenguajes y mecanismos para distinguir entre recursos Web con diferentes grados de honradez (fiabilidad en la información que tienen). (Geroimenko *et. al.*, 2003: 14)

Esto es necesario ya que en la Web existe libertad de participación y no toda la información que se encuentra en ella es correcta. Esto se puede ilustrar con el siguiente caso.

Supongamos que un usuario desea investigar especies de árboles en Centroamérica. En la Web existen muchos sitios sobre Botánica, en donde se puede encontrar información acerca de especies de árboles con su respectiva descripción. Pero también existen sitios de Criptobotánica, que es una pseudociencia que estudia plantas que forman parte de leyendas locales o relatos literarios, pero de los que no se posee evidencia alguna sobre su

⁷ Prolog es uno de los lenguajes de programación lógica más utilizados en Inteligencia Artificial, y ya ha sido probado y utilizado en muchas otras aplicaciones (como sistemas expertos, bases de datos inteligentes, lingüística computacional, etc.). Ahora, se intentará vincularlo con la implementación de esta capa de la Web semántica.

existencia. Dentro de la Criptobotánica, el “árbol come hombres” es una especie que se cree existe en las profundas selvas de Centroamérica. El usuario podría recibir junto a la información correcta sobre árboles reales, datos sobre este árbol carnívoro del que no existe prueba alguna sobre su existencia.

Entonces, ¿qué se puede hacer para distinguir entre fuentes de información con poco grado de fiabilidad y fuentes de información sustentadas científicamente?. La tarea de la capa de evidencia será verificar que la información tenga pruebas para considerarla como válida.

2.5.3 La capa de confianza

La capa de lógica junto con la de evidencia permitirán construir la “Web de la confianza”, donde agentes de software autónomos serán capaces de tomar decisiones importantes de diferentes tareas, por ejemplo, comprar comida, agendar citas, reservar boletos para un viaje, entre otras muchas cosas, sin intervención humana (*Ibidem*: 14). Se le llama capa de confianza porque los seres humanos podrán confiar en las computadoras al delegarles tareas de mayor importancia que resolverán mediante el uso de la información disponible en la Web semántica.

En el tercer capítulo de este trabajo se explicarán detalladamente qué son los agentes inteligentes, los cuales son el punto clave de esta capa. Sin embargo, esta capa todavía está muy lejos de ser desarrollada.

2.6 Concepto de firma digital

La Web actual se caracteriza por la libertad de participación. Restringir esa libertad a costa de la “Web de la confianza” implicaría modificar la filosofía de universalidad seguida por la Web hasta el día de hoy.

Por ello se ha pensado en la idea de la firma digital. Esto consiste en seguir permitiendo a cualquier persona participar en la conformación de la Web, pero únicamente las fuentes de confianza tendrán esta firma digital para indicar a las computadoras y agentes inteligentes que un sitio Web cuenta con información que está correctamente organizada y evidencias que respaldan su contenido.

De esta manera, los sitios Web se clasificarán de acuerdo a su grado de confiabilidad en: muy confiables, confiables, poco confiables, no confiables.

Aplicaciones de la Web semántica

El camino hacia la Web semántica está trazado, y las tecnologías que la harán posible se están generando y perfeccionando poco a poco. Aunque la fecha en que podremos verla funcionando es imprecisa, es seguro que los usuarios de la Web seremos testigos de su cambio gradual.

La Web semántica está en la mira de muchas personas. Hay científicos, ingenieros, políticos, militares, empresarios, e incluso usuarios comunes, que están atentos a cada movimiento en el desarrollo de la nueva Web y pensando qué posibles aplicaciones se le puede dar a este suceso tecnológico que está por llegar en unos cuantos años.

Las aplicaciones son muchas. Los científicos piensan como pueden aprovechar la Web para crear nuevas formas de almacenar el conocimiento humano. Biólogos, historiadores, geógrafos, etc., saben que el acervo cultural de la humanidad podría ser almacenado en la Web semántica para facilitar su acceso y organización.

Los políticos saben que la apropiación de una nueva tecnología tiene impacto en el desarrollo de un país. No es fortuito que el gobierno de Estados Unidos esté tan interesado en el desarrollo de la Web semántica. Además, toda nueva tecnología se puede aplicar de uno u otro modo en el aspecto bélico. Y es que en este mundo quien tiene la información tiene el poder.

Hasta las personas que no entienden el funcionamiento de una nueva tecnología saben que pueden sacar ventajas de ella de alguna manera. Los grandes empresarios piensan como usarla en sus negocios para redituar grandes ganancias. Los usuarios comunes buscan como ahorrar tiempo y esfuerzo en sus labores cotidianas.

La Web semántica es para todos. Desde el contador de una empresa internacional buscando como almacenar y administrar toda la información financiera, hasta un niño encargando a un agente inteligente buscar por él una tarea escolar.

Este último capítulo está enfocado e describir brevemente algunas de las aplicaciones de la futura Web.

3.1 Agentes inteligentes personales

Los servicios proveídos por la Web del futuro serán mejorados si la semántica se añade a los recursos Web presentes. Las computadoras serán capaces de hacer citas con el doctor sincronizadas con nuestra agenda, encontrar nuevos proveedores para los productos que consumimos, hacer arreglos para viajes, entre otras tareas. (Breitman *et. al.*, 2007: 9-10)

Para lograr lo anterior es necesaria la creación de agentes inteligentes. Un agente es una pieza de software que trabaja en forma autónoma y proactiva. "Un agente personal en la Web Semántica recibirá ciertas tareas y preferencias de una persona, buscará información en los recursos de la Web, se comunicará con otros agentes, comparará información acerca de los requerimientos y preferencias del usuario, seleccionará ciertas opciones, y dará respuestas al usuario". (Antoniou *et. al.*, 2004: 14)

Imaginemos la siguiente situación. En un futuro no muy lejano, un adolescente de 15 años está pensando qué hacer durante el fin de semana. Después de un rato de no dar con una respuesta en su mente, decide preguntarle a su agente inteligente. Así como hoy en día todos podemos tener una cuenta de correo electrónico, en el futuro de la Web semántica se pretende que cada quien tenga su propio agente inteligente.

Este agente inteligente personal sabe de antemano datos sobre su dueño (género, edad, lugar donde vive, etc.). Estos datos fueron pedidos a su dueño directamente en algún momento (quizá la primera vez que utilizó a su agente), o los fue deduciendo en base al

tipo de cosas que hace en la Web (puede deducir el nivel económico, tipo de música que le gusta, entre muchas otras cosas, a partir de los sitios Web que frecuenta).

Este usuario enciende su computadora, introduce su contraseña para que su agente empiece a trabajar, y le pide realizar una búsqueda sobre “cosas para hacer el fin de semana”. Entonces, el agente inteligente empieza a realizar una investigación en la Web semántica. El agente inteligente hace uso de las ontologías en la Web para deducir que “cosas para hacer” es lo mismo que “actividades”, y que “fin de semana” está compuesto de “sábado” y “domingo”. En base a los conocimientos que tiene sobre el usuario, el agente empieza a buscar actividades para realizar en sábado o domingo que le gusten a alguien con el perfil de su dueño.

El agente busca en toda la Web, por lo que encontrará una gran cantidad de resultados. Después realizará un proceso de filtrado para encontrar las mejores opciones en base a las características del usuario. Buscará actividades que queden cerca de donde vive, que vayan acorde con su edad y gustos, que pueda pagar, etc. También debe verificar que los sitios Web de donde obtiene información sean confiables para no poner en riesgo a su usuario.

Una vez terminada la búsqueda, que llevará unos cuantos minutos, le presentará al usuario una lista de posibilidades. Todas las opciones que recibe el usuario son correctas, pero de ellas elegirá la que más le guste. El agente toma nota de cual de todas las opciones eligió como una retroalimentación para mejorar los conocimientos que tiene acerca de las preferencias de su usuario.

El ejemplo anterior demuestra que tan útil puede ser la Web semántica. Si la información está bien organizada, los agentes inteligentes pueden hacer este tipo de búsquedas de un modo relativamente sencillo.¹

¹ A pesar de que esto parece una historia sacada de un libro de ciencia ficción, ya hay avances al respecto. Mientras la Web semántica se está desarrollando, la Inteligencia Artificial está diseñando en paralelo estos agentes inteligentes. Es posible encontrar en la Web una gran cantidad de información al respecto.

Un agente inteligente también se puede encargar de tareas más complejas, por ejemplo, comprar un libro. El usuario le podría decir: “compra un libro de poesía”. El agente buscaría entonces un libro de poesía que le agrade a su usuario (en base a lo que conoce de él), y después buscaría el mejor sitio para comprarlo (que sea seguro y barato).

De acuerdo con James A. Hendler (1999), un agente debe tener las siguientes cualidades:

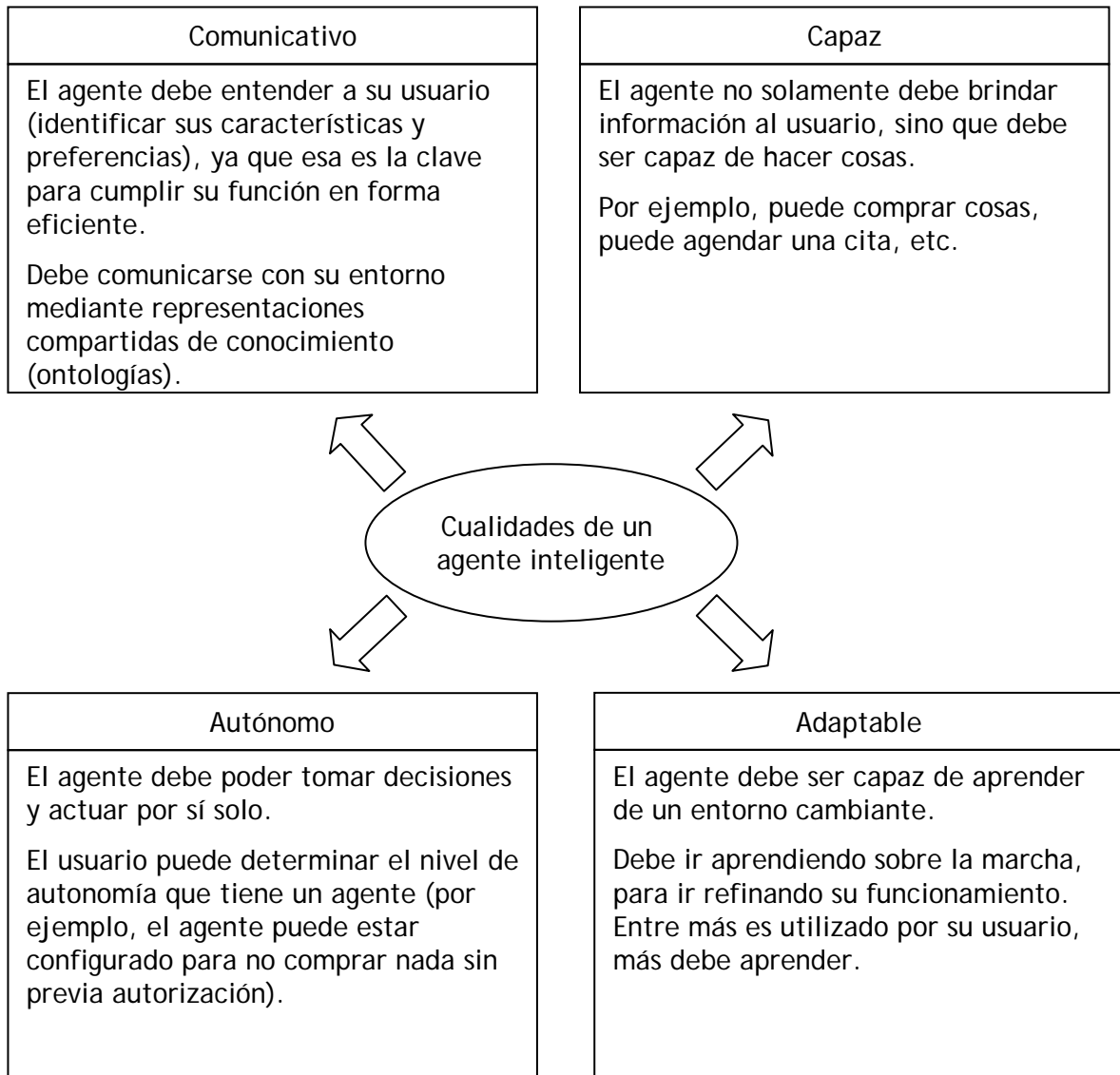


Figura 3.1 Cualidades de un agente inteligente

3.2 La Web semántica social

En los primeros años de la Web únicamente las personas con conocimientos en HTML podían publicar información. Pero en el año 2003 se suscitó una gran ola de socialización cuando aparecieron los blogs² y los wikis³, que permitieron a más usuarios participar con contenido en la Web sin necesidad de saber utilizar HTML. Los blogs y wikis permitieron a cualquier individuo o grupo de personas solicitar un espacio personal en la Web y llenarlo con sus ideas de forma sencilla. (Mika, 2007: 24)

La mayoría del contenido Web tiene sus orígenes en la gente común que colabora con sus opiniones, con sus puntos de vista sobre los temas, o con la creación de páginas Web sencillas y rudimentarias. En realidad es poco el contenido de la Web que es elaborado por instituciones científicas o gubernamentales en comparación con el que es realizado por los usuarios comunes. La Web es resultado de un trabajo colectivo entre todos los usuarios. Debido a esto, se ha pensado en la idea de dejar a los mismos usuarios la tarea de dotar de semántica a la actual Web.

La Web semántica social es una iniciativa que pretende crear un sistema de conocimiento colectivo en base a la participación de toda la humanidad. Esto se realiza dejando en manos de todos los usuarios de la Web el establecimiento de la clasificación y relación entre los conceptos.

Esto no quiere decir que cada usuario se ponga a escribir sus propios documentos rdf. La idea es crear un mecanismo que permita recolectar el modo de pensar del colectivo humano. La idea sería de gran utilidad, ya que la cantidad de conceptos y relaciones existentes es enorme.

² Un blog es una especie de "diario electrónico". Es una página Web que es creada en forma automatizada. El usuario no necesita crearla desde HTML, sino seguir una serie de pasos sencillos en una interfaz gráfica (la cual se ve en el navegador y es por lo general un formulario HTML). De esta manera, hasta un niño puede tener su propio blog.

³ Un wiki es una colección de páginas Web enlazadas mediante hipervínculos, y que pueden ser modificadas por cualquier persona. El objetivo es que la construcción de uno de estos sitios sea un trabajo cooperativo. Wikipedia, una enciclopedia en línea, es el mejor ejemplo de un sitio creado mediante la colaboración de muchas personas.

Recientemente se ha comenzado a utilizar un sistema de clasificación llamado Folksonomía⁴. Las folksonomías se distinguen del resto de sistemas de organización documental porque son de clasificación colaborativa, no es un experto documentalista quien realiza la clasificación, sino personas comunes que al encontrar un recurso en la Web pueden agregarle descriptores o etiquetas. Este método de organización es subjetivo, espontáneo, caótico y desordenado, pero no por ello ineficiente. Por el contrario, se convierte en una forma simple de describir y relacionar contenidos en línea que los mismos usuarios categorizan, según su perspectiva y la utilidad que el contenido les representa. (López Guzmán, marzo 2008)

Un ejemplo de folksonomía es el sitio *Flickr*⁵, el cual permite a los usuarios compartir fotos. Cuando un usuario agrega una nueva fotografía le agrega una serie de etiquetas que describen su contenido. Por ejemplo, si un usuario sube una foto de sus vacaciones, puede agregar las etiquetas "playa", "verano", "soleado", "mar", entre otras, para describir el significado de la foto.

Existen varios proyectos orientados a impulsar la Web semántica social. Uno de ellos es *Stumpedia*⁶, el cual es una herramienta que permite a cualquier usuario crear una página sobre un determinado término de búsqueda (un concepto o tema importante), y después crear links dentro de su página hacia recursos que estén vinculados con el término en cuestión (imágenes, blogs, artículos, videos, etc.). El objetivo es realizar un filtro de información, ayudando a otras personas que investigan sobre ese concepto a ahorrarse la tarea de tener que clasificar entre resultados útiles e inútiles.

Otro ejemplo en la Web es *Semandeks*⁷, el cual permite a cualquier usuario participar en la creación de conceptos y relaciones importantes. Los usuarios pueden revisar todo lo que ya está creado o añadir nuevas cosas. Existen términos creados, como "gato", "Paulo Coelho", "Día de la independencia", "Lenguaje de programación". También existe una lista de relaciones que puede unir a esos conceptos, tales como "es un", "es hijo de",

⁴ Folk es una palabra en inglés que significa popular. De esta manera, folksonomía significa "clasificación hecha por el pueblo".

⁵ Para consultar *Flickr* el link es: <http://www.flickr.com>

⁶ Para consultar *Stumpedia* el link es: <http://www.stumpedia.com>

⁷ Para consultar *Semandeks* el link es: <http://semandeks.com>

“tiene un”, “es parte de”, etc. Un usuario puede crear nuevos conceptos o relaciones si considera que hacen falta, o simplemente unir las cosas ya existentes. Por ejemplo, un usuario interesado en aportar información sobre el autor Oscar Wilde, podría crear los conceptos “Oscar Wilde”, “El retrato de Dorian Gray”, “El príncipe feliz”, etc. También puede crear nuevas relaciones, por ejemplo, “fue escrito por”, “escribió”, etc. Después puede conectar las cosas que ha creado con las cosas que ya existían. Cualquier conexión debe seguir el formato Concepto-Relación-Concepto. Por ejemplo:

Concepto	Relación	Concepto
Oscar Wilde	es un	escritor
Oscar Wilde	escribió	El retrato de Dorian Gray
Oscar Wilde	escribió	El príncipe feliz
El retrato de Dorian Gray	es una	novela
El príncipe feliz	es un	cuento
cuento	es un	género literario
novela	es un	género literario

Tabla 3.1 Ejemplos de conexiones en *Semandeks*

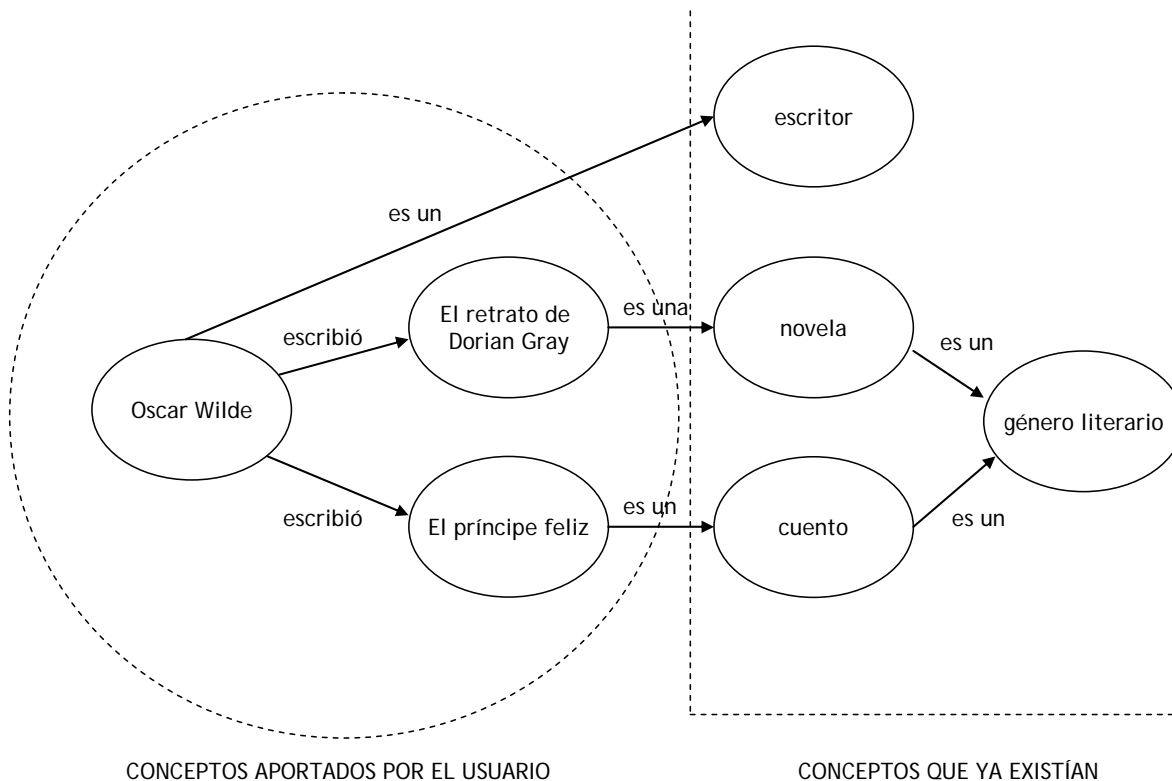


Figura 3.2 Representación gráfica de una ontología en *Semandeks*

De esta manera, con la colaboración de todos los usuarios se puede crear una gran ontología que represente el conocimiento humano.

Otro ejemplo es *DBpedia*⁸, esfuerzo comunitario para extraer información estructurada de *Wikipedia*, con el fin de poder hacer búsquedas más sofisticadas. A pesar de que *Wikipedia* es la base de información más completa que existe en la actualidad, todavía trabaja con motores de búsqueda por palabras clave. El objetivo es lograr estructurar a *Wikipedia* para que un buscador semántico pueda realizar búsquedas como “Dame todas las ciudades de Francia que tengan más de 10 000 habitantes” o “Dime los actores de teatro que más premios han recibido”.

El hecho de permitir la participación de cualquier persona tiene un riesgo implícito: la inexactitud de la información. Siempre habrá errores en la conexión de conceptos ya que

⁸ Para consultar *DBpedia* el link es: <http://dbpedia.org>

son realizadas por personas que no necesariamente son expertas en la materia. Sin embargo, la cantidad de conocimiento humano es tan grande que quizá sea la única forma de lograr organizar todo el conocimiento humano.

3.3 Administración del conocimiento.

El funcionamiento de las empresas necesita del manejo de una gran cantidad de información: lista de clientes, lista de proveedores, catálogo de productos, registros de ventas, lista de empleados, lista de cuentas por cobrar, programas de entrega, inventario del almacén, etc.

Grigoris Antoniou y Frank Van Harmelen (2004: 3-4) explican que la administración del conocimiento se interesa en la adquisición, acceso y mantenimiento del conocimiento que tiene una organización. Ésta ha emergido como una actividad clave en los grandes negocios porque ellos ven su conocimiento interno como un capital intelectual del cual pueden obtener más grande productividad, crear nuevos valores, e incrementar su competitividad. La administración del conocimiento es particularmente importante para grandes organizaciones con departamentos dispersos geográficamente.

Sin embargo, desde la perspectiva de la administración del conocimiento, la tecnología actual sufre limitaciones en las siguientes áreas:

- ✓ Búsqueda de información.- Las compañías usualmente dependen de motores de búsqueda basados en palabras claves.
- ✓ Extracción de información.- Tiempo y esfuerzo humanos son requeridos para seleccionar de los documentos recuperados información relevante.
- ✓ Mantenimiento de información.- Comúnmente hay problemas, como inconsistencias en la terminología y fallas en remover la información obsoleta.
- ✓ Descubrimiento de información.- El nuevo conocimiento implícito existente en las bases de datos es extraído utilizando Minería de Datos⁹. Sin embargo, esta

⁹ La Minería de Datos es un conjunto de técnicas provenientes de Inteligencia Artificial para extraer conocimiento implícito en bases de datos (es decir, encontrar relaciones que no se ven a simple vista e identificar patrones de comportamiento).

tarea es todavía difícil para colecciones de documentos estructurados débilmente.

La lista de problemas anteriores es muy parecida a los que tiene la Web actual. Por lo tanto, las ideas desarrolladas en busca de la Web semántica podrían ser usadas en la administración del conocimiento de las empresas. La Web semántica representa un conjunto de tecnologías que pueden ser usadas en redes corporativas internas. (Daconta *et. al.*, 2003: 4)

3.4 Comercio Electrónico (E-commerce)

El comercio electrónico es la actividad de compra y venta de productos o servicios a través de la Web. Esta actividad ha tomado cada vez mayor importancia, pues cada día son más los usuarios que por este medio realizan compras de ropa, libros, música, juguetes, computadoras, etc.

La Web semántica tendrá un gran impacto en el comercio electrónico. Las grandes empresas se adecuarán a las nuevas tecnologías de modo que sus productos estén inmersos en la estructura semántica para que puedan ser encontrados por los agentes inteligentes. Dentro de los términos incluidos en la gran ontología de la Web encontraremos los nombres de productos, marcas, servicios, etc. Además, se espera que la Web semántica proporcione más confianza por medio de la firma digital (que únicamente tendrán sitios confiables), de modo que mayor número de usuarios realizarán compras por medio de sus agentes inteligentes sin preocupación alguna.

3.5 Herencia cultural

Catalogar nuestra herencia cultural ha sido una actividad principalmente de los museos y otras instituciones culturales en todo el mundo. Hoy en día, casi todos los grandes museos hacen sus colecciones disponibles en la Web, frecuentemente con gran calidad. (Breitman *et. al.*, 2007: 11)

Una de las aplicaciones más importantes de la Web semántica será almacenar el acervo cultural de la humanidad. Para ello se requerirá de la participación de biólogos, físicos, matemáticos, ingenieros, geógrafos, historiadores, artistas, filósofos, médicos, etc. De esta manera, se construirá una gran enciclopedia estructurada semánticamente que permita almacenar el saber colectivo.

3.6 Relaciones interpersonales

La forma en que se dan las relaciones interpersonales se ha modificado por las nuevas tecnologías disponibles en la Web. Los servicios de mensajería instantánea, el chat, las redes sociales, etc., se han convertido en medios para conocer gente nueva, encontrar pareja, mantener contacto con familiares y amigos, etc.

Si la Web actual ha tenido un gran impacto en la forma en que los seres humanos establecen sus relaciones sociales y afectivas, todavía se deben esperar cambios más importantes con la instauración de la Web semántica, pues ella traerá consigo nuevas herramientas de interacción social más sofisticadas, donde los agentes inteligentes harán búsquedas más precisas sobre gente con gustos similares a los nuestros.

Conclusiones

La Web es una de las invenciones con mayor trascendencia en la humanidad. En los albores del siglo XXI la Web se fusionó rápidamente con la cotidianidad de muchas personas, quienes hoy la ven como una herramienta fácil de utilizar para obtener información, divertirse en los ratos de ocio, establecer vínculos sociales, entre otras muchas cosas.

La Web se caracteriza por su descomunal tamaño, por su apertura a la participación de cualquier persona sin distinción alguna, por su estructura anárquica, y lo más importante, por su continuo cambio.

La Web está en un proceso de evolución, perfeccionándose poco a poco, por medio de la incorporación de nuevas tecnologías que permiten proveer mejores servicios a la comunidad de usuarios. La siguiente fase de este cambio continuo será la Web semántica.

Aunque estamos ya acostumbrados al modo de funcionamiento de la Web, siempre hay cosas que mejorar. Hoy en día, la Web tiene una serie de problemáticas, entre ellas, el hecho de que los motores de búsqueda nos regresan una enorme lista de sugerencias sobre documentos que contienen las palabras que introducimos, pero que no nos dan una respuesta concreta a nuestras preguntas. La labor de seleccionar cuál es la información útil todavía es tarea de los humanos.

La Web semántica se postula como una solución a este problema, planteando la idea de estructurar la información de un modo que pueda ser entendida tanto por humanos como por computadoras. Una estructura de este tipo permitiría que los motores de búsqueda regresaran respuestas concretas a los usuarios, ahorrándoles la tarea de tener que revisar, analizar y clasificar la información.

La construcción de la Web semántica será paulatina, por medio de la incorporación sucesiva de nuevas tecnologías sobre las ya existentes. Algunas de estas tecnologías ya están disponibles:

- ✓ El estándar Unicode se encargará de garantizar la universalidad de la nueva Web, al proveer una lista de caracteres provenientes de todas partes del mundo, y cuyo uso será válido en la Web semántica.
- ✓ Por medio de los URI's se puede identificar a los recursos disponibles en la Web, de modo que exista un único modo de referirnos a un recurso particular, independientemente de los lenguajes humanos.
- ✓ XML provee una sintaxis para estructurar los documentos que conformarán a la Web semántica. Estos documentos se caracterizan por describirse a sí mismos, pues los datos y los metadatos se almacenan en el mismo documento.
- ✓ XML Schema restringe el vocabulario contenido en un documento XML con el objetivo de estandarizar la forma de estructurar grupos de documentos que tengan características semejantes.
- ✓ RDF permite describir a los recursos de la Web, así como las relaciones que existen entre ellos.
- ✓ RDF Schema se encarga de la jerarquización de recursos en la Web utilizando los conceptos de clases y subclases.
- ✓ OWL permite la creación de ontologías, relacionando los recursos en la Web. Además de las relaciones jerárquicas que se pueden construir en RDF Schema, OWL permite otro tipo de relaciones (semejanza, oposición, etc.).

Aún se necesita definir qué tecnologías serán necesarias para proveer a la Web de la capacidad de utilizar lógica para extraer nuevo conocimiento del ya existente, así como de la habilidad para discernir entre fuentes de información con distinto grado de confiabilidad.

La Web semántica abre un panorama de grandes posibilidades en cuanto a las aplicaciones que puede tener. Los agentes inteligentes personales, la Web semántica social, la Administración del conocimiento, el comercio electrónico, etc., son sólo algunas de las tantas aplicaciones que tendrá la nueva Web. No hay duda que la evolución de la Web seguirá dando nuevas sorpresas y tomando senderos inimaginables.

Fuentes

- ✓ Alesso, H. Peter y Smith, Craig Forsythe (2004). *Developing Semantic Web services*. Estados Unidos: AK Peters.
- ✓ Antoniou, Grigoris y Van Harmelen, Frank (2004). *A semantic Web primer*. Cambridge: MIT Press.
- ✓ Benson, Edward (2008). *The Art of Rails*. pp. 4-5. Cambridge: Wiley & Sons Ltd.
- ✓ Berners-Lee, Tim (1998a). "What the Semantic Web can represent", en <http://www.w3.org/DesignIssues/RDFnot.html>.
 - y Fielding, R.; Irvine, U.C. y Masinter, L. (1998b). "Uniform Resource Identifiers (URI): Generic Syntax", en <http://tools.ietf.org/html/rfc2396>.
 - y Fischetti, Mark y Dertouzos, Michael (2000a). *Tejiendo la red*. España: Siglo XXI.
 - (2000b). "Semantic Web XML2000", en <http://www.w3.org/2000/Talks/1206-xml2k-tbl/>.
 - y Hendler, James y Lassila, Ora (2001, mayo). "The Semantic Web", en *Scientific American*, núm. 284, pp. 34-43. Estados Unidos: Scientific American Inc.
- ✓ Breitman, Karin K.; Casanova, Marco Antonio y Truszkowski, Walt (2007). *Semantic Web*. Estados Unidos: National Aeronautics and Space Administration.
- ✓ Daconta, Michael C.; Obrst, Leo Joseph y Smith, Kevin T. (2003). *The Semantic Web. A Guide to the Future of XML, Web Services, and Knowledge Management*. Indianápolis: Wiley.
- ✓ Geroimenko, Vladimir y Chen, Chaomei (2003). *Visualizing the semantic Web*. Nueva York: Springer.
 - (2004). *Dictionary of XML technologies and the semantic Web*. Nueva York: Springer.

- ✓ Hendler, James (1999). "Is There an Intelligent Agent in Your Future?", en <http://www.nature.com/nature/webmatters/agents/agents.html>.
- (2003). "Frequently Asked Questions on W3C's Web Ontology Language (OWL)" en <http://www.w3.org/2003/08/owlfaq.html>.

- ✓ Kashyap, Vipul; Bussler, Christoph y Moran, Matthew (2008). *The Semantic Web. Semantic for Data and Services on the Web*. Nueva York: Springer.

- ✓ Lacy, Lee W. (2005). *OWL: Representing Information using the web ontology* . Victoria: BC Trafford.

- ✓ López Guzmán, Clara (2008, marzo). "Las folksonomías: organización a través del desorden", en *Enter@te*, num. 67, pp.5-6. México: UNAM.

- ✓ Mika, Peter (2007). *Social networks and the Semantic Web*. Nueva York: Springer.

- ✓ Powers, Shelley (2003). *Practical RDF*. Estados Unidos: O'Reilly & Associates.

- ✓ Van der Vlist, Eric (2002). *XML Schema*. Estados Unidos: O'Reilly & Associates.

Índice de figuras y tablas

Capítulo 1

Figura 1.1 La Web vista como un gran iceberg de la información.	10
---	----

Capítulo 2

Figura 2.1 "Layer cake" de la Web semántica propuesto por Berners-Lee.	20
Tabla 2.1 Fragmento de la tabla Unicode.	21
Figura 2.2 Ejemplo de un árbol XML	30
Figura 2.3 Proceso de validación de documentos XML.	36
Figura 2.4 Despliegue en el navegador utilizando XSL.	40
Figura 2.5 Ejemplo de grafo RDF.	43
Figura 2.6 Grafo RDF en el que se relacionan dos recursos Web.	44
Figura 2.7 Representación gráfica de clases y subclases en RDF Schema	45
Figura 2.8 Definición de ontología.	48
Tabla 2.2 Principales tipos de relaciones entre términos.	50

Capítulo 3

Figura 3.1 Cualidades de un agente inteligente.	61
Tabla 3.1 Ejemplos de conexiones en Semandeks.	64
Figura 3.2 Representación gráfica de una ontología en Semandeks.	65