

ARTÍCULO

ESTADO ACTUAL DE LA WEB 3.0 O WEB SEMÁNTICA

Javier Salazar Argonza

Estado actual de la Web 3.0 o Web Semántica

Resumen

Se realiza un acercamiento a la tecnología de la Web 3.0 o Web semántica. Se habla de sus orígenes, así como de sus ventajas y desventajas. Se explican las principales características de los buscadores semánticos y se muestran los diferentes resultados que arrojan, en comparación con los tradicionales. Se analizan otras aplicaciones basadas en la Web semántica y se explican los beneficios que presentan. Finalmente, se presentan las tendencias que presenta su desarrollo.

Palabras clave:

Web 3.0, Web semántica, buscadores semánticos, aplicaciones semánticas, tecnologías de la información y la comunicación.

Current state of Web 3.0 or Semantic Web

Abstract

We performed a technology approach to Web 3.0 or Semantic Web, discussing their origins and their advantages and disadvantages. It explains the main features of semantic search engines, and shows the different results produced compared with traditional. It discusses other applications based on Semantic Web and explains the benefits they present. Finally, we present the development trends presented.

Keywords:

Web 3.0, semantic Web, semantic search engines, semantic applications, information and communication technologies.

Introducción

La Web semántica es una de las tecnologías de información y comunicación que posiblemente agradará más a los internautas, por las bondades que promete, entre ellas:

- Organizar la gran cantidad de Información suelta, redundante y de calidad dudosa, existente en la Web actual;
- Reducir los costos y los tiempos que invertimos en localizar información útil en la Web, porque actualmente debemos realizar el análisis semántico de la información;
- Establecer las reglas para integrar información con diferentes formatos, y
- Resolver los problemas de interoperabilidad entre diversos dispositivos y plataformas con los que accedemos a la Web.

Así, en este artículo se tratará de dar una visión del actual grado de desarrollo de esta tecnología.

El término "Web 3.0" apareció por primera vez en 2006, en un artículo del diseñador de páginas Web estadounidense Jeffrey Zeldman, crítico de la Web 2.0 y fundador de la empresa Happy-Cog para el desarrollo de páginas Web. Este término, utilizado por el mercado para promocionar las mejoras con respecto a la Web 2.0, básicamente describe la evolución del uso y la interacción con la Web a través de la incorporación de las siguientes tendencias tecnológicas:

- La transformación de la Web en una base de datos distribuida: a través del lenguaje de marcas extensible XML ("eXtensible Markup Language"), la estructura para la descripción de recursos, en base a metadatos, RDF ("Resource Description Framework"), y diversos microformatos que permiten agregar significado semántico a los contenidos.
- La introducción de la tecnología de Web semántica: emplea búsquedas en lenguaje natural y la minería de datos. Clasifica la información de manera más eficiente, a fin de devolver resultados más precisos a las solicitudes de búsqueda de los usuarios.
- Hacer los contenidos Web accesibles desde múltiples dispositivos: comprende el diseño de las interfaces para que puedan ser accedidas desde múltiples dispositivos, tales como teléfonos inteligentes, televisores digitales, iPad's, PDA's, etc.
- El uso de las tecnologías de inteligencia artificial: por medio de programas especializados (Agentes Inteligentes), para comprender mejor lo que la gente solicita.
- La Web geoespacial: que combine la información geográfica disponible de los usuarios, con la información que predomina en la Web, generando contextos que permiten realizar búsquedas u ofrecer servicios en base a la localización.
- Uso de la tecnología para 3D: que transforme la Web actual en espacios tridimensionales inmersivos, donde los usuarios puedan sumergirse e interactuar.

Los orígenes de la "Web semántica", por su parte, se remontan al año 2001, cuando el científico británico Tim Berners Lee (Inventor de la World Wide Web) y el investigador en inteligencia artificial James Hendler (del Departamento de Ciencias de la Computación de la Universidad de Maryland), presentaron un artículo en la revista Nature abordando por primera vez la posibilidad de generar este tipo de Web. En el artículo explican que la Web debería estar diseñada no sólo como espacio de información accesible para los seres

humanos, sino también como espacio donde pudieran acceder las computadoras. El objetivo del acceso por parte de las computadoras sería ayudar a los seres humanos a buscar y conseguir la información que necesitan, con la mayor rapidez y garantía de éxito posibles.

Ante la imposibilidad de compartir la información entre computadoras con la Web actual, estos autores propusieron adaptar los documentos de Internet con una nueva tecnología, que permitiera procesar los datos también por computadora, es decir, encontrar la información en forma rápida y eficaz, basándose en el significado y no en los términos.

La Web semántica define un nuevo modelo de Web (figura 1), con una capacidad muy superior a la que conocemos actualmente, regido por los principios de las búsquedas semánticas y la inteligencia artificial.

Bajo este nuevo modelo, las páginas Web adquieren un significado propio, el cual tiene como objetivo primordial facilitar las búsquedas de información, ajustándose más a satisfacer los intereses de los usuarios que las realizan.

La información que contienen actualmente las páginas de Internet, carece de una estructura preestablecida, lo que impide que las máquinas sean incapaces de procesar la información, como si se tratase de un ser humano, es decir, comprendiendo su significado.

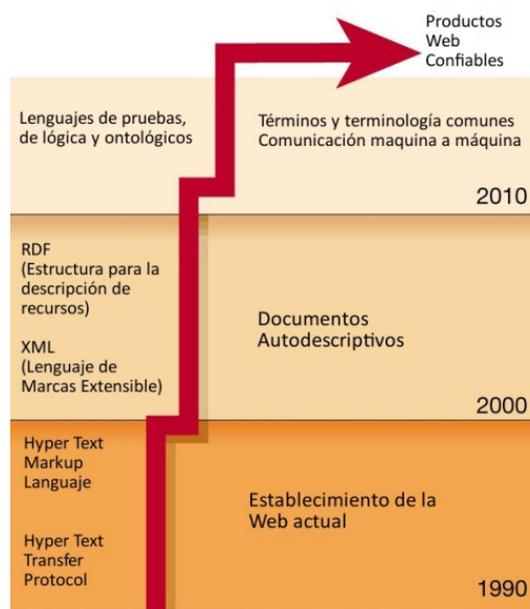


Figura 1: Capas del nuevo modelo tecnológico propuesto por Tim Berners Lee y James Hendler para la Web semántica. Fuente Imágen: <http://www.nature.com/nature/journal/v410/n6832/full/4101023b0.html>

La Web semántica surgió con el objetivo de mejorar el lado malo de Internet: "La sobreinformación". Tener acceso a tanta información, mucha de ella sin la calidad ni el rigor necesario, hace que el usuario, a menudo, se vea saturado de información y colmado de dudas ante lo que está leyendo.

La Web semántica está especialmente diseñada para dotar de significado a su contenido, de tal manera que cuando el buscador trata de localizarlo, no se fija en las palabras que contiene, sino precisamente en el significado, en lo que el usuario de verdad está buscando.

El mecanismo con que funciona la Web semántica se desarrolla a través de ontologías (esquemas conceptuales definidos para el intercambio de información), para añadir significado semántico a la Web, y taxonomías (reglas), para definir objetos y las relaciones que se pueden establecer entre ellos.

La Web semántica, pese a su escaso desarrollo actual, tiene un potencial ilimitado que permitirá, en poco tiempo, acceder a la información de Internet como nunca.

Ventajas de la Web semántica

- Incorpora contenido semántico a las páginas que se suben a Internet. Esto permite una mejor organización de la información, asegurando búsquedas más precisas por significado y no por contenido textual.
- Permite a las computadoras la gestión de conocimiento, hasta el momento reservada a las personas (hace uso de inteligencia artificial).

Desventajas de la Web semántica

- Es costoso y laborioso adaptar los documentos de Internet, para poder ser procesados de forma semántica (a esto hay que sumar los problemas del idioma).
- Es necesario unificar los estándares semánticos y proveer relaciones de equivalencia entre conceptos. Por ejemplo, en el caso del código postal, se debe establecer que CP es igual a ZC "zip code" en el caso del inglés.

El esfuerzo de dotar de significado a las páginas Web vale la pena, ya que producirá una World Wide Web mucho más asequible y entendible, con búsquedas de información más precisas.

Es lo mismo la Web 3.0 que la Web Semántica

Como ya se explicó anteriormente, el término Web 3.0 ha sido utilizado para describir el camino evolutivo de la red e incluye a la tecnología de la Web Semántica. Asimismo, las investigaciones académicas actuales están dirigidas a desarrollar programas que puedan razonar, basados en descripciones lógicas y agentes inteligentes. La Web semántica, por su parte, facilita las búsquedas de información al dotar de significado a los contenidos, lo cual hace que estos conceptos se relacionen ampliamente.

Avances y tendencias de la Web 3.0 o Web semántica

A partir de 2007 comenzaron a surgir varias aplicaciones merecedoras de esta categoría, esencialmente buscadores. Uno de los primeros servicios de Web semántica presentados fue Twine, desarrollado por Radar Networks.

Los buscadores semánticos

Twine se trató de un servicio Web de índole social, que permitía subir piezas de información, tales como textos, archivos, etcétera, y realizaba análisis para generar automáticamente una serie de etiquetas que definían el contenido. Estas etiquetas podían estar presentes en el texto analizado o inferirse derivando información asociada a ellas. Con una estructura semejante a la de una Wiki, Twine trató de establecer una “red de conocimiento” capaz de poner en contacto a personas, de acuerdo con el conocimiento extraído de sus contribuciones.

Aunque Twine pudo ser un buscador encargado de indexar semánticamente toda la Web, arrancó con un objetivo más sencillo: hacerse cargo de aquellas partes que la gente considera valiosas. Con la adquisición de Twine.com por la Empresa EVRI, el 11 de marzo del 2011, esta interesante aplicación fue retirada de la red. En su lugar surgió un “motor de descubrimiento de contenido móvil” al que se denominó EVRI. <http://www.evri.com>.

EVRI es un buscador que se especializa en la devolución de resultados de las últimas noticias, temas seguidos regularmente, nuevos temas y recomendaciones de amigos. La plataforma central de la tecnología de EVRI se basa en el procesamiento del lenguaje natural y búsquedas semánticas, para proporcionar canales de contenido agregado en millones de temas.

Wolfram Alpha, desarrollado por la compañía Wolfram Research, es otra importante aplicación Web 3.0 que surgió el 15 mayo de 2009. Este buscador, al que se le llamó “Motor de Conocimiento Computacional”, es un servicio en línea que responde directamente las preguntas del usuario, en lugar de proporcionar una lista de los documentos o páginas Web que podrían contener la respuesta, como lo hace Google.

Wolfram Alpha se basa en la tecnología “Question Answering” o “QA”, para la recuperación de la información. QA es un método que procesa el Lenguaje Natural de una forma más compleja que otros sistemas de recuperación de documentos, razón por la que se le observa como un paso por delante de la tecnología de los buscadores tradicionales para Internet. Asimismo se basa en uno de los anteriores programas creados por Wolfram Research: “Mathematica”, que incluye el procesamiento de álgebra, cálculo numérico y simbólico, visualizaciones y capacidades estadísticas. <http://www.wolframalpha.com>.

Diferencias entre los buscadores semánticos y los tradicionales

A fin de ilustrar las diferencias entre las tecnologías de los buscadores tradicionales y los semánticos, se presenta el siguiente ejemplo: suponga que usted desea localizar en Internet los datos del presidente de México. A fin de trabajar de la misma manera con los buscadores, la búsqueda se realizará en inglés. La cadena a buscar será: “President of Mexico”.

Google, uno de los buscadores más conocidos y utilizados, es capaz de arrojar millones de resultados por cada palabra de búsqueda. Esta sobreabundancia de resultados responde al hecho de que las webs del momento son sintácticas, esto es, se sirven de las palabras, los términos, para ser localizadas por los buscadores.

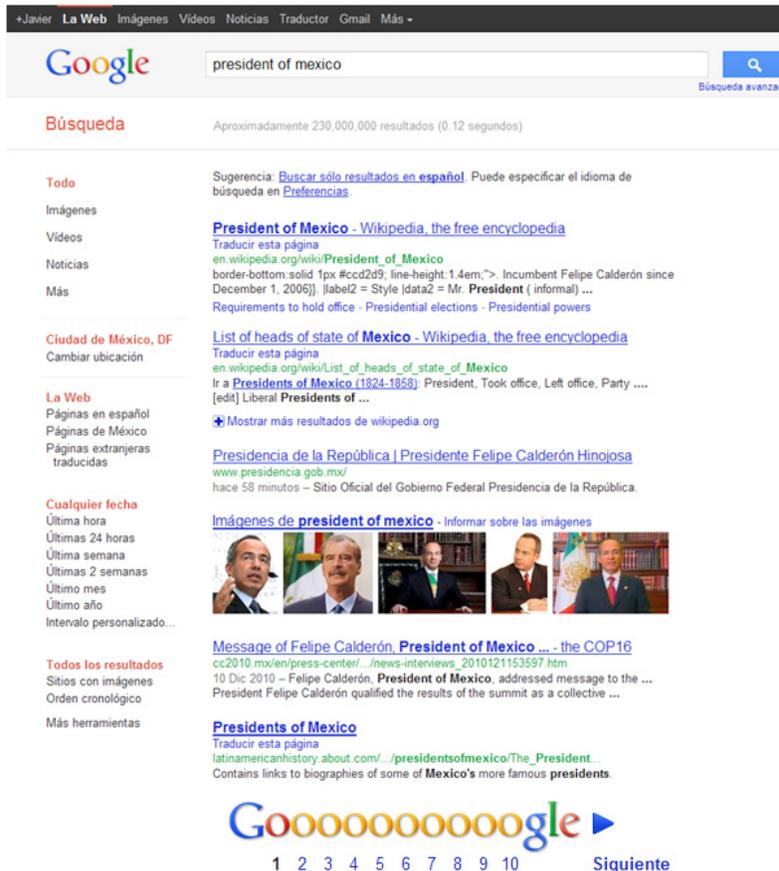


Figura 2. Resultados de la búsqueda “President of Mexico”, realizada con Google.

En la figura 2 note que Google entregó una lista con todas las páginas que encontró con la información solicitada (230,000,000 de resultados en 0.12 segundos), sin duda muchas en un tiempo muy corto. No obstante, si analizamos los resultados obtenidos, nos damos cuenta que tendremos que revisar cada una de las ligas para extraer los datos que requerimos.

La misma búsqueda, pero ahora realizada con Wolfram Alpha, nos arroja resultados que son sin duda más apegados a las necesidades y deseos del usuario final (figura 3).

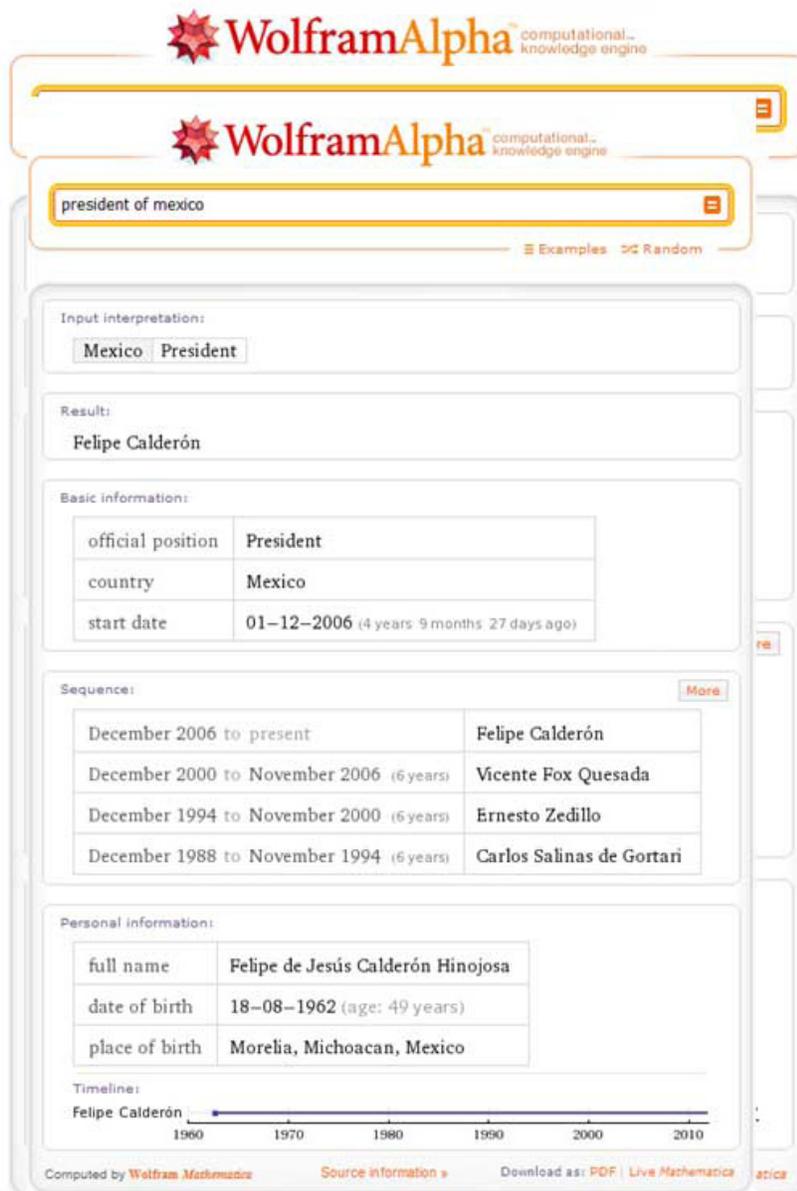


Figura 3. Resultados de la búsqueda “President of Mexico”, realizada con Wolfram Alpha.

En la figura 3 note que aparece un resumen muy elaborado sobre el tópic □President of Mexico□. Esta

es la promesa de los buscadores semánticos o Web 3.0: ajustar las búsquedas de información a las necesidades de los usuarios.

La misma búsqueda, pero ahora realizada con el buscador “EVRI”, sin duda arroja resultados actualizados e interesantes, pero con el enfoque comercial para el cual fue concebido: “mostrar sólo las últimas novedades del presidente de México” (figura 4).

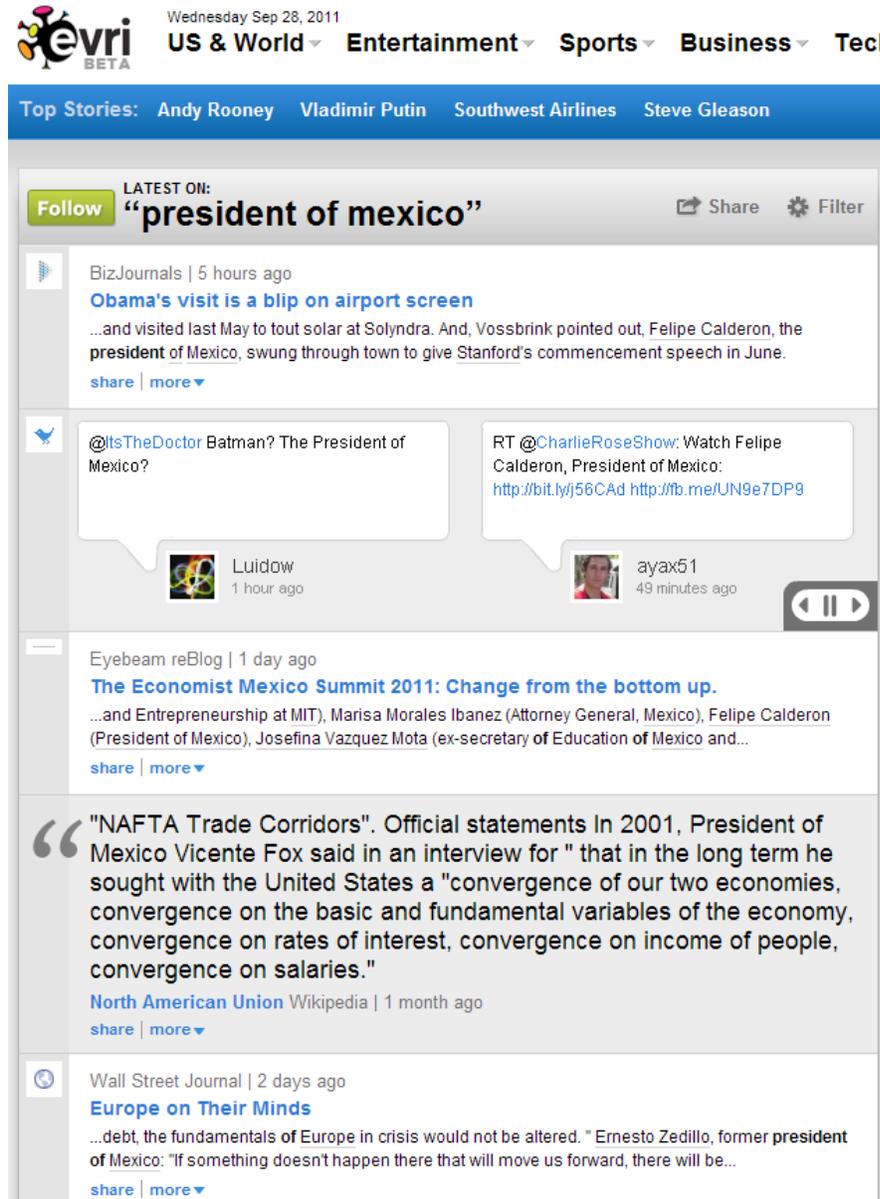


Figura 4. Resultados de la Búsqueda “President of Mexico”, realizada con el buscador EVRI.

Como puede apreciarse, hay una diferencia significativa entre los resultados arrojados por cada uno de los 10 -xx

buscadores.

Vale la pena mencionar también que Google ha estado trabajando en el desarrollo de un producto de búsqueda semántica. De hecho el pasado 3 de junio de 2009 fue lanzado Google Squared. Este buscador permitía extraer datos estructurados de toda la Web y presentaba sus resultados en un formato parecido al de una hoja de cálculo. Google Squared actualmente no está disponible en la red, pues se cerró después de una serie de pruebas a fin de mejorar las búsquedas. También en Julio de 2010, Google adquirió a la empresa Metaweb. Esta empresa es una compañía líder en el tema de Web semántica. Son quienes están detrás de Freebase, una inmensa base de datos, de la cual Google ya venía usando servicios para su sitio de noticias.

Otro buscador Web de importancia, que está tratando de introducir el uso de la tecnología semántica, es Bing de Microsoft, que fue puesto en línea el 3 de junio de 2009. Bing está basado en la tecnología semántica desarrollada por la empresa Powerset y que Microsoft compró en 2008 para el enriquecimiento de sus búsquedas.

Por otra parte, en lo que se refiere al tema de los buscadores semánticos, el pasado 2 de junio de 2011 Google, Yahoo y Bing, anunciaron la creación de un proyecto al que llamaron schema.org para acercar la Web semántica a cualquier página basada en la tecnología Rich Snippets (“Fragmentos Ricos”) de Google, que surgió en el 2009.

Entre las características y mejoras que ofrece schema.org, destacan las siguientes:

- Posibilidad de etiquetar muchos tipos de información, como por ejemplo: restaurantes, productos, eventos, sonidos, recetas, etcétera.
- Unificar la estructura de los datos. Hasta ahora podíamos trabajar con RDFa, microdatos o microformatos. A partir de ahora todo se centrará en microdatos; eso sí, el resto seguirán teniendo soporte, al menos por el momento.

Otras aplicaciones basadas en la Web semántica

A pesar de que la Web 3.0 tiene poco tiempo de haber aparecido, las empresas de vanguardia, sobre todo en los medios de comunicación y en línea, están empezando a adoptar los elementos de esta tecnología. Por ejemplo, la tienda en línea Amazon está utilizando la inteligencia artificial para hacer recomendaciones a sus clientes, basadas no sólo en sus propias historias de navegación y compra, sino también en el comportamiento de otros clientes con historias similares. Best Buy por su parte ha añadido la tecnología semántica en su página Web, para permitir a los motores de búsqueda encontrar información detallada en

las páginas de sus productos.

Actualmente también están en marcha programas de investigación para combinar aspectos de Internet con la Web semántica, con el fin de mejorar las cadenas de suministro de las empresas y desarrollar oportunidades de ventas.

Si bien la mayoría de la tecnología semántica está todavía en desarrollo, a continuación se listan algunos ejemplos de aplicaciones que ya se han instrumentado con ella:

- El RSS (Really Simple Syndication, Redifusión Web o Sindicación Web). RSS es un vocabulario RDF (Resource Description Framework), basado en XML (eXtensible Markup Language), que permite catalogar la información (noticias y eventos) de tal manera que sea posible encontrar información precisa, adaptada a las preferencias de los usuarios.

Los archivos RSS contienen metadatos ("datos que describen a otros datos"), sobre fuentes de información especificadas por los usuarios, cuya función principal es avisar a éstos que los recursos que han seleccionado para formar parte de esa RSS, han cambiado sin necesidad de comprobar directamente la página, es decir, notifican de forma automática cualquier cambio que se realice en esos recursos de interés seleccionados.

El servicio de RSS permite difundir información actualizada a usuarios que se han suscrito a una o más fuentes de contenidos. Es un formato desarrollado para compartir contenidos Web en sitios que se actualizan con frecuencia. Los sitios Web que brindan el servicio de RSS, tienen un botón anaranjado (figura 5), el cual permite suscribirse a ellos. Una vez realizado este proceso, bastará con un simple navegador para consultar las novedades del sitio Web solicitado, sin tener que entrar a éste.



Figura 5. Símbolo utilizado para el servicio RSS.

- Zemanta. Es un programa que facilita la elaboración de blogs (bitácoras colaborativas). Utiliza búsquedas semánticas a lo largo de la captura y la redacción de los contenidos del blog, que permiten al usuario disponer e incorporar imágenes, ligas, etiquetas, enlaces y/o los artículos más relevantes relacionados con el tema de su blog. Zemanta se distribuye como un plug-in que se instala en el navegador de Internet. Una vez activo analiza su escritura para realizar la descarga de contenidos (<http://www.zemanta.com>).

- Inform. Es un sitio Web que permite ver noticias como piezas interconectadas de un rompecabezas más grande y lo hace de forma automatizada. Inform recoge las noticias de los medios de comunicación en línea y las compañías de información, y se encarga de interconectar y relacionar los temas empleando la tecnología de la Web Semántica. Esto le permite mostrar un panorama más completo alrededor de las noticias (<http://inform.com>).
- Siri. Es una aplicación de inteligencia artificial y Web semántica para el iPhone, que adquirió la empresa Apple en el 2010 por 200 millones de dólares, según Mashable.com.¹ La empresa que la desarrolló se llama igual que su aplicación Siri. Con este programa Apple logró instrumentar un nuevo asistente de voz en su iPhone, que permite responder a preguntas naturales y ejecutar tareas fácilmente, sin utilizar el teclado.



Figura 6. Logotipo de la nueva aplicación Siri de Apple.

Gracias a Siri, el iPhone es capaz de responder a preguntas de voz como “¿Qué tiempo hace hoy?”, “¿Cuántos días faltan hasta Navidad?” o “¿Qué hora es en París?”. Pero también podrá activar una alerta a la hora deseada: “Quiero que me despiertes a las seis de la mañana”; apuntar una cita en el calendario del teléfono; contestar en cascada a los mensajes del usuario; responder a sus acciones, etcétera. Asimismo, Siri será compatible con las aplicaciones de terceros (<http://www.apple.com/iphone/features/siri.html>).

Existen más aplicaciones Web 3.0 que están surgiendo, tales como Tripit (para viajeros frecuentes (<http://www.tripit.com>)) o Swotti (que recoge los comentarios de los usuarios y las opiniones de la Web, a fin de brindar información útil sobre marcas, productos, servicios o personas, acerca de sus atributos, adjetivos y/o competidores (<http://www.swotti.com>)). De hecho, a lo largo de 2010 empezó a desarrollarse con mayor fuerza la tecnología Web 3.0 y los expertos aseguran que continuará esta tendencia hasta finales de 2020, por lo que las principales novedades las podremos apreciar durante la próxima década.

Conclusiones

- Es una tecnología que promete lograr que las computadoras entiendan de manera exacta lo que se les pide buscar o hacer.
- Están en proceso de estandarización las reglas con que se deberá operar y unificar la estructura de los datos de la Web semántica.

1

<http://mexico.cnn.com/tecnologia/2011/10/04/el-nuevo-iphone-se-adentra-en-el-software-de-inteligencia-artificial>.

- Es costoso, laborioso y tardado adaptar los documentos de Internet para poder ser procesados de forma semántica.
- Ya comienzan a liberarse servicios TIC que permiten a los usuarios aprovechar esta tecnología.

Los principales avances de esta tecnología los veremos en los próximos años.

Referencias bibliográficas

VICTORIA ESAINS "Google compra Metaweb para darle un toque semántico a sus búsquedas", ALT1040 [en línea]. 17 de Julio de 2010. [Consultado: 25 de Septiembre de 2011].

Disponible en Internet:

<http://alt1040.com/2010/07/google-compra-metaweb-para-darle-un-toque-semantico-a-sus-busquedas>

BILL GOODWIN "Businesses should prepare for Web 3.0, says Booz&Co" ComputerWeekly.com [en línea]. 25 de Julio de 2011. [Consultado: 18 de Septiembre de 2011].

Disponible en Internet:

<http://www.computerweekly.com/Articles/2011/07/25/247388/Businesses-should-prepare-for-Web-3.0-says-Booz&Co.htm>

TIM BERNERS-LEE& James Hendler "Publishing on the semantic Web" Nature [en línea]. 26 de Abril de 2001. [Consultado: 10 de Septiembre de 2011].

Disponible en Internet:

<http://www.nature.com/nature/journal/v410/n6832/full/4101023b0.html>.

HEBE BRAVO "Cuando Tim Berners Lee presentaba el WorldWideWeb" Maestros del Web [en línea]. 26 de Febrero de 2008. [Consultado: 12 de Septiembre de 2011].

Disponible en Internet:

<http://www.maestrosdelweb.com/editorial/cuando-tim-berners-lee-presentaba-el-worldwideweb/>.

MARCOS CALDERÓN BLANCO "¿Qué es la Web 3.0?" PÁGINA DE RECUPERACIÓN Y ACCESO A LA INFORMACIÓN, Universidad Carlos III de Madrid [en línea]. [Consultado: 10 de Septiembre de 2011].

Disponible en Internet:

<http://web30websemantica.comuf.com/web30.htm>.

JEFFREY ZELDMAN "Web 3.0" A List apart [en línea]. 17 de Enero de 2006. [Consultado: 10 de Septiembre de 2011].

Disponible en Internet:

<http://www.alistapart.com/articles/web3point0/>.

“Jeffrey Zeldman, Biographical article” Wikipedia the free encyclopedia. [en línea]. 6 de Octubre de 2011. [Consultado: 10 de Septiembre de 2011].

Disponible en Internet:

http://en.wikipedia.org/wiki/Jeffrey_Zeldman.

VIV DEHAES “Web 3.0 o 3D” Interacciones. [en línea]. 29 Noviembre de 2007. [Consultado: 11 de Septiembre de 2011].

Disponible en Internet:

<http://www.interacciones.com.ar/web-30-o-3d/>

“Web semantic” Wikipedia the free encyclopedia [en línea]. 26 de septiembre de 2011. [Consultado: 12 de Septiembre de 2011].

Disponible en Internet:

http://es.wikipedia.org/wiki/Web_Sem%C3%A1ntica

ALONSO ÁLVAREZ “¿Ya tenemos aplicaciones Web 3.0?” La Cofa, Observatorio Tecnológico, Telefónica Investigación y Desarrollo, S.A. [en línea]. 30 de Octubre de 2007. [Consultado: 22 de Septiembre de 2011].

Disponible en Internet:

<http://www.lacofa.es/index.php/innovacion/ya-tenemos-aplicaciones-web-30>.

STEFAN NAGTEGAAL y Steven Wittens “Evri acquires Twine.com” Fusion Corporate Partners [en línea]. 11 de Marzo de 2010. [Consultado: 20 de Septiembre de 2011],

Disponible en Internet:

<http://fusiondiginet.com/2010/03/11/evri-acquires-twine-com/>

“Guía Breve de Web Semántica” W3C [en línea]. 6 de mayo de 2010. [Consultado: 1 de Octubre de 2011].

Disponible en Internet:

<http://www.w3c.es/divulgacion/guiasbreves/websemantica>.

M. BAMIEH “A Glimpse at Web 3.0: 13 Semantic Web Applications Reviewed” Thoughtpick [en línea]. 24 de Agosto de 2009. [Consultado: 2 de Octubre de 2011].

Disponible en Internet:

<http://blog.thoughtpick.com/2009/08/a-glimpse-at-web-3-0-13-semantic-web-applications-reviewed.html>.

SERGIO RODRÍGUEZ “Siri, el asistente de voz del iPhone, eclipsa al teléfono” EL MUNDO.es [en línea].

05 de Octubre de 2011. [Consultado: 05 de Octubre de 2011].

Disponible en Internet:

<http://www.elmundo.es/elmundo/2011/10/04/navegante/1317755191.html>

MARK "RIZZIN" HOPKINS "Siri Raises \$8.5 Million for Personal Artificial Intelligence" Mashable Business [en línea]. 14 de Octubre de 2008. [Consultado: 5 de Octubre de 2011].

Disponible en Internet:

<http://mashable.com/2008/10/14/siri/>.

LOGO WEB 3.0 Y ARTÍCULO "Web 3.0 la web semántica?" Universia Noticias España [en línea]. 21 de Enero de 2011. [Consultado: 7 de Septiembre de 2011].

Disponible en Internet:

<http://noticias.universia.es/ciencia-nn-tt/reportaje/2011/01/21/782381/web-3-0-web-semantica.html>.

KAVI GOEL y Pravir Gupta "Introducing schema.org: Search engines come together for a richer web" Google Webmaster Central Blog [en línea]. 02 de Junio de 2011 [Consultado: 27 de Septiembre de 2011].

Disponible en Internet:

<http://googlewebmastercentral.blogspot.com/2011/06/introducing-schemaorg-search-engines.html>.

