Web Semántica: La evolución de una Web con significado

La Web Semántica es una de las tendencias tecnológicas dominantes en nuestros días y una de las que más impacto tiene para los usuarios de la Web. El concepto de la Web Semántica fue introducido en el artículo The Semantic Web [1], publicado en el 2001 por Berners-Lee, quien la definió como una extensión de la Web actual que habilita una Web de datos global. Su objetivo es permitir que los datos almacenados en la Web puedan ser procesados por las máquinas de manera inteligente, facilitando a las personas la búsqueda, integración y análisis de la información disponible en la Web. A partir de su definición en 2001 se han desarrollado un conjunto de tecnologías que conforman la base de su implementación, y estas han sido utilizadas en aplicaciones de ciudades inteligentes, telemedicina, investigación y colaboración científica, entre otras.

La primera tecnología de esta tendencia fue RDF, un modelo de datos genérico que permite estructurar y ligar datos para describir recursos en la Web. RDF surge en el año 1999, sin embargo es hasta el año 2004 cuando se populariza su uso, cuando las aplicaciones de la Web Semántica lo utilizan para generar información estructurada en un formato estándar, accesible y manejable.

RDF además sirvió como base en la especificación de tecnologías como RDFS, diseñado para agregar significado a los recursos RDF mediante el desarrollo de ontologías; posteriormente el RDFS fue extendido por el lenguaje OWL para soportar la representación de ontologías más complejas y la inclusión de reglas de inferencia. Algunos ejemplos del uso de estas primeras tecnologías son herramientas semánticas para la gestión de contenidos Web como ODESeW [2] y OntoWebber [3], las cuales ofrecieron soluciones a problemáticas específicas de la década pasada.

Linked Data: la nueva perspectiva para la publicación de datos

En el año 2006, surge el paradigma de datos enlazados (*Linked Data*) [4], el cual propone la inclusión de enlaces entre los datos publicados en la Web Semántica por diferentes proveedores facilitando el descubrimiento de contexto de los datos. Las tecnologías principales que desde ese entonces dan soporte a *Linked Data,* además de RDF, son las HTTP URIs (un medio genérico para identificar entidades o conceptos en la web) y SPARQL (un lenguaje de consulta para la recuperación de información descrita en RDF).

Con el objetivo de estandarizar el proceso para publicar y conectar datos estructurados en la Web, Berners-Lee propuso un conjunto de principios que establecen mejores prácticas para *Linked Data*. Estos principios evolucionaron para crear un esquema de 5 niveles (estrellas) para la publicación de datos enlazados abiertos (*Linked Open Data*), el cual promueve la publicación de datos enlazados provenientes de diversas fuentes, instituciones u organizaciones de manera abierta (libre de regalías). El esquema establece que a mayor número de estrellas, más fácil es el consumo de los datos por diversas personas o aplicaciones. A raíz del surgimiento de *Linked Open Data* se generó una nube de conjuntos de datos expuestos bajo este esquema, siendo el dataset central DBPedia [5]. El conjunto de datos central de esta nube, contiene datos estructurados extraídos de Wikipedia, con cerca de 45 millones de enlaces a conjuntos de datos externos. Una aplicación real de la utilización de DBPedia es el IBM Watson [6], el cual es una tecnología cognitiva que procesa la información de manera similar a un ser humano mediante la comprensión del lenguaje natural, la generación de hipótesis basadas en la evidencia y el aprendizaje.

Como complemento a la iniciativa de *Linked Open Data*, la industria ha ofrecido sus propios enfoques en el desarrollo de la Web de Datos. En este contexto, las principales aplicaciones se relacionan principalmente con motores de búsqueda que operan sobre un espacio de datos global, tal es el caso de Knowledge Graph de Google [7] y el Open Graph Protocol de Facebook [8].

¿Qué nos espera en el futuro de la Web Semántica?

La creación de un ecosistema de aplicaciones y servicios web inteligentes, ubicuos y conscientes de sí mismos y del entorno, que no sólo serán consumidores de la información disponible en la Web, sino que además serán publicadores de nueva información proveniente tanto del mundo digital como del mundo real, donde incluso los dispositivos puedan comunicarse directamente entre sí. Este ecosistema converge de manera natural con los esfuerzos hechos en diversos frentes (científicos, tecnológicos, sociales, económicos, políticos) para avanzar hacia el Internet del Futuro, el cual consiste en una infraestructura global de dispositivos inteligentes y objetos físicos interconectados en Internet para soportar una plataforma de aplicaciones y servicios Web inteligentes que ofrezcan diferentes beneficios a la sociedad.

El futuro no está muy lejos, hoy en día existen propuestas como el proyecto Ready4SmartCities [9] que propone el uso de tecnologías de *Linked Data* para mejorar la eficiencia de los sistemas de energía en ciudades inteligentes.

Referencias:

[1] http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html
[2] http://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/old-technologies/74-odesew
[3] http://infolab.stanford.edu/OntoAgents/OntoWebber/
[4] http://www.w3.org/DesignIssues/LinkedData.htm
[5] http://www.dbpedia.org/
[6] http://www.ibm.com/smarterplanet/us/en/ibmwatson/
[7] http://www.google.com/insidesearch/features/search/knowledge.html
[8] https://www.facebook.com/about/graphsearch
[9] http://www.ready4smartcities.eu/

RESEÑA CURRICULAR:

Blanca Vázquez, Eliel Morales, Karen Nájera y Samuel Vieyra son investigadores de la Gerencia de Desarrollo de Nuevos Produc-tos y Servicios en INFOTEC. Actualmente realizan investigación en Web Semántica, Modelado Organizacional, Datos Abiertos e Internet del Futuro en colaboración con universidades de México y de la Unión Europea.
{blanca.vazquez, eliel.morales, karen.najera, samuel.vieyra}@infotec.com.mx