



Web Semántica – un socio para el aprendizaje

Vitor Gonçalves

Profesor Adjunto de la "Escola Superior de Educação do Instituto Politécnico de Bragança"

La Web se asume como una mediateca mundial. Los motores de búsqueda convencionales todavía dejan mucho que desear, sobre todo en lo que respecta a la comprensión del significado de los contenidos, ya que los resultados devueltos corresponden a cientos o incluso miles de páginas web que contienen palabras iguales a las palabras clave presentadas en la búsqueda. Por una parte, esto es debido a que el lenguaje HTML describe cómo se deben presentar los datos en lenguaje natural, pero no ofrece ninguna descripción semántica de los datos o de las hiperconexiones entre ellos. Por otro lado, los motores de búsqueda tienen dificultades con la relación entre dos conceptos y no devuelven conocimiento implícito. Además, se añade el hecho de que hay información útil almacenada en bases de datos (*deepweb* o *web profunda*), a la cual un motor de búsqueda difícilmente puede acceder.

El *WWW Consortium* (W3C) ha contribuido de manera significativa a la construcción de una infraestructura tecnológica para soportar una «web de datos» (un sitio web que relacione datos tales como los que se encuentran en las tradicionales bases de datos), sustituyendo gradualmente la convencional "web de documentos" (una web formada por enlaces o conexiones entre recursos web). La web de datos puede ser entendida como una nube de datos conectados (*linked data*), un enorme conjunto de documentos (actualmente constituida por miles de millones de hechos y millones de enlaces), cuyos datos se describen en un lenguaje comprensible también para las máquinas. Esta web de datos conectados es la base del objetivo de la "Web Semántica". La Web Semántica no es más que una extensión de la web actual, en el sentido de que la información en forma de texto en lenguaje natural sea complementada por su semántica basada en la representación formal del conocimiento.

Las tecnologías de la Web Semántica permiten la representación explícita del conocimiento y su posterior procesamiento, creando nuevos conocimientos a partir del conocimiento implícitamente oculto. Para ello, primero hay que estructurar los datos e identificar los recursos de la web. El URI (Uniform Resource Identifier) es la base para el direccionamiento. El Unicode proporciona la legibilidad necesaria. La XML (*eXtensible Markup Language*) permite representar la estructura de los datos y

proporciona un formato para el intercambio de información. Pero, ¿cómo representar hechos simples? Para ello, se recurre a la tecnología RDF (*Resource Description Framework*). El RDFS proporciona el vocabulario que permite que la información sea vista como clases de instancias, lo que permite inferencias sobre la misma. Podemos decir que el RDFS permite representar un modelo del mundo y que las bases de datos RDF –*triple stores*– permiten almacenar los datos semánticos. Para cuestionar los datos semánticos se puede utilizar la *Query Language SPARQL*. Pero, ¿cómo obtener la semántica en la web? Por el momento, la forma más inmediata es a través de RDFa y microdatos. Los datos están incrustados en el HTML de forma transparente para el usuario, pero se interpretan por los motores de búsqueda de Yahoo (*SearchMonkey*) y Google (*RichSnippets*). Es conveniente representar el mundo a través de *Description Logics*, representando la lógica en la Web Semántica con el lenguaje ontológico OWL (*Ontology Web Language*), así como las reglas de inferencia a través del lenguaje RIF (*Rule Interchange Format*). Las diferentes ontologías y otros sistemas de representación del conocimiento deben articularse para generar sinergias. Por lo tanto, se puede recurrir al SKOS (*Simple Knowledge Organization System*), un modelo de datos común para el intercambio y conexión de los sistemas de organización del conocimiento a través de la Web. Tecnologías que garanticen la fiabilidad, la veracidad, los derechos autor y la confidencialidad de la información son también elementos importantes para la creación de una Web más inteligente.

Estas tecnologías de la Web Semántica (que permiten almacenar, organizar y enlazar datos en la Web, construir vocabularios y escribir las reglas para el manejo de datos) forman la arquitectura de la Web Semántica.

Cuestionar la web en lenguaje natural y obtener respuestas en lenguaje natural en vez de un gran conjunto de páginas web es una visión que corresponde a un futuro mucho más cercano de lo que podemos imaginar.

¿A quién no le gustaría tener su propio agente inmobiliario, de viajes o de aprendizaje? Como resultado de varios proyectos de investigación, ya existen muchos ejemplos en los campos de la farmacia y la medicina, integrando datos heterogéneos de investigación o, in-



cluso, de observación clínica; aplicaciones que conectan datos de bolsas de empleo, de aerolíneas, de hoteles y restaurantes o de agencias de viajes. Hace un año que la mismísima Google utiliza el *Knowledge Graph*. Este gráfico del conocimiento no sólo representa gráficamente las relaciones existentes entre los elementos de datos de fuentes públicas (Freebase, Wikipedia, CIA WorldFactbook, etc.), como también utiliza una base de conocimiento estructurado, que se obtiene a través de búsquedas que los usuarios realizan todos los días.

En la última década, a nivel europeo, se denota una participación cada vez mayor en proyectos que se centran en el desarrollo y explotación de tecnologías y aplicaciones para la Web Semántica. En la *CORDIS (Community Research and Development Information Service)* de la Comisión Europea se identifican más de un centenar de proyectos financiados en este campo (alrededor de tres docenas en contexto educativo), de los cuales una docena tiene como socios instituciones portuguesas (por lo menos dos se centran específicamente en la educación). La Fundación para la Ciencia y la Tecnología (FCT) ha financiado, al menos dos proyectos más que promueven la aplicación de la Web Semántica en el contexto educativo. Adicionando, incluso, algunas tesis de doctorado disponibles en los repositorios digitales de algunas instituciones portuguesas, podemos decir que Portugal está en la cola de la I + D en el contexto de la Web Semántica aplicada a la Educación.

En la actualidad, son pocos los que son conscientes de la presencia de las tecnologías de la Web Semántica, pero cuando empiecen a trabajar con un ordenador capaz de entender lo que necesitan y cuáles sus preferencias, que sea capaz de responder en lenguaje natural y capaz de aprender con ellos mismos, volviéndose un verdadero socio de los procesos de aprendizaje, entonces estaremos en la era de la Web Semántica. Mientras tanto, podemos hacer uso de algunas aplicaciones para intranets o incluso extranets (por ejemplo, entre sistemas de e-Learning). El camino se está edificando, pero «Roma y Pavía no se construyeron en un día».

Directorios Web 2.0

[Semantic Web](#)

Hitzler, P.; Krötzsch, M.; Rudolph, S. (2009). *Foundations of Semantic Web Technologies*, CRC Press.

Heath, T. & Bitzer, C. (2011). *Linked Data - Evolving the Web into a Global Space*, Morgan & Claypool