

La web semántica y el desafío del conocimiento

El autor adelanta los nuevos pasos en la evolución de internet relacionada con el software que cada día será más “inteligente”, cada día incorpora más comprensión a la información.

29/09/2011 | Las nuevas tecnologías -es absurdo continuar llamándolas nuevas- de información y comunicación son la infraestructura de la revolución científica y tecnológica actual, cada día más extensa y más acelerada. Pero están lejos de ser algo definitivo porque al comprar la computadora más reciente, ya es obsoleta, porque la I+D del propio fabricante lo sacará del mercado en año y medio.

En materia de software, la próxima etapa importante en la evolución de internet, será la web semántica, es decir, aquella en que las tecnologías permiten a los ordenadores procesar el significado de las páginas web, en lugar de sus nombres o sólo descargar información.

El Millennium Project en un ejercicio prospectivo sobre las posibilidades para la educación y el aprendizaje en 2030 publicado en el informe 2007 State of the Future, anunciaba la aparición de la Web 17.0, que significaría la culminación de la integración de los datos en la web -proceso iniciado por Google Earth, Wikipedia, los cursos en línea del MI- organizada conforme un marco lógico de conceptos -tanto precisos como inciertos- con una interfaz de lenguaje natural, actualizada en forma dinámica y con un subsistema inteligente que "comprende" las reglas lógicas que gobiernan las interacciones de las entidades, probabilidad anticipada por el inventor de la web, Tim Berners-Lee al comenzar el siglo. Esto hará que el software acceda al significado de los contenidos en línea, en una visión conocida como la web semántica.

Hace poco, Google, Microsoft y Yahoo, tres grandes de internet, lanzaron una iniciativa, llamada: Schema.org que define como un vocabulario de términos interconectados que pueden agregarse al código HTML de una página web para comunicar el significado de los conceptos en la página. Así si en un texto se habla de un Palacio de Justicia, Schema.org lo entiende como un tipo específico de edificio gubernamental.

Se pueden definir personas y eventos, también atributos como: distancia, masa, o duración. Estos datos permitirán a los motores de búsqueda en una determinada consulta comprender mejor la utilidad de una página, por ejemplo, dejando claro que al mencionar la palabra pentágono, se refiere al Departamento de Defensa de EE.UU. y no a formas regulares de cinco lados.

La medida finalmente fomentará el uso generalizado de la tecnología que hará que la información en línea sea tan comprensible a las computadoras como lo es para los seres humanos.

Se puede esperar una ola de otras aplicaciones y servicios inteligentes capaces de entender la información en línea casi tan bien como nosotros.

Si muchas páginas web aceptan las sugerencias de Schema.org, no sólo beneficiaría las búsquedas. "Estos datos pueden ser utilizados por cualquier software de correlación cruzada, para identificar cosas que están

relacionadas o para entender la relación entre la información de diferentes fuentes".

También Wikipedia, la enciclopedia en línea, está explorando maneras de abrazar la web semántica. Este recurso global construido a partir del tiempo libre de millones de voluntarios, Wikipedia puede ser la personificación de la web 2.0, pero la Fundación Wikimedia, organización sin fines de lucro que dirige Wikipedia, está pensando entre otros proyectos, en cómo convertirse una pieza clave de la web 3.0 o la web semántica.

Ello implica hacer que muchos de sus 15 millones de artículos sean comprensibles para las computadoras como lo son a los seres humanos. El software podría saber, que los números mostrados en una de las columnas de una tabla de presidentes de EEUU son fechas. Ello permitiría que aplicaciones basadas en Wikipedia generen automáticamente plazos históricos o respondan a preguntas de cultura general que normalmente requieren de una persona para localizar y leer una entrada correspondiente en el sitio.

Las universidades alemanas Libre de Berlín y Leipzig desarrollaron otra base de datos sin fines de lucro llamada DBpedia, un esfuerzo comunitario para extraer información estructurada de Wikipedia y vincular esta información con otros conjuntos de datos de la web haciendo que esta nueva información sea accesible a todos.

DBpedia permite consultas más sofisticadas que las usuales a la enciclopedia, describiendo actualmente más de 3,5 millones de cosas. De ellas 1,67 millones se clasifican en una ontología coherente, que incluye 364.000 personas, 462.000 plazas, 99.000 álbumes de música, 54.000 películas, 17.000 video-juegos, 148.000 organizaciones, 169.000 especies y 5.200 enfermedades.

Los datos DBpedia definen y etiquetan las características y los resúmenes de estos 3,5 millones de cosas en 97 idiomas diferentes, con 1.850.000 enlaces a las imágenes y 5.900.000 enlaces a páginas web externas; 6.500.000 enlaces externos a otros conjuntos de datos RDF, 633.000 categorías Wikipedia, y 2.900.000 categorías YAGO. Ésta podría ser utilizada en nuevas e interesantes formas e inspirar nuevos mecanismos para la navegación, la conexión y la mejora de la propia enciclopedia.

Una persona, en su trabajo diario, encuentra en su escritorio gran cantidad de información proveniente de múltiples fuentes: mails, búsquedas web, agendas, notas, hojas de cálculo, presentaciones y documentos. La clasificación de esta información es difícil y se hace mayormente de una manera ad hoc.

En unos meses más Radar Networks, lanzará Twine una herramienta de marcadores gratuita, llamada a cambiar la forma de buscar y organizar información ofreciendo una forma más inteligente de aumentar la productividad. Mediante conexión sin cables puede recoger automáticamente todas las páginas web visitadas, mails enviados y recibidos, y así sucesivamente.

Al disponer de esa información, comienza a analizarla y ordenarla automáticamente en categorías que incluyen personas involucradas, conceptos discutidos, lugares, organizaciones y empresas. Así cuando un usuario busca algo, puede acceder rápido a la información relacionada al respecto. También utiliza elementos de las redes sociales para que un usuario tenga acceso a la información recopilada por otros en su red. Nova Spivack, CEO y fundador de Radar Networks, afirma que todo esto crea una especie de "inteligencia colectiva".

Entretanto, los humanos aún parecen disponer de un pequeña ventaja en la conexión de sentido ya que recientes investigaciones han encontrado que los verdaderos expertos, tienen algo por lo menos tan valioso como la maestría en las reglas de clasificación, que es un instinto, una comprensión instantánea del tipo de problema que se enfrenta.

El cerebro humano es una suerte de máquina de reconocimiento de patrones y cuando se enfoca adecuadamente, rápidamente puede profundizar su capacidad de captar una cosa, situación o persona de un primer vistazo, lo que en alguna profesión se denomina ojo clínico. Con este instinto se puede "ver" el mejor movimiento a realizar.

Ahora bien no hay razón alguna para que alguien con buen ojo para la moda o juegos de palabras no pueda, a partir de éste, desarrollar una intuición para la clasificación de rocas o mamíferos o ecuaciones algebraicas, sumando al conocimiento perceptual un poco de interés o motivación.

¿Será posible que escuelas, maestros y estudiantes puedan aprovechar mucho las tecnologías de la información en la web semántica y de esta capacidad llamada aprendizaje perceptivo, para aumentar nuestra inteligencia colectiva?

Las opiniones vertidas en este espacio, no necesariamente coinciden con la línea editorial de Diario Los Andes. .

URL <http://www.losandes.com.ar/notas/2011/9/29/semantica-desafio-conocimiento-596749.asp>