

Web semántica

Artículo de la Wikipedia, la enciclopedia libre

La Web semántica (del inglés semantic web) se basa en la idea de añadir metadatos semánticos a la World Wide Web. Esas informaciones adicionales —que describen el contenido, el significado y la relación de los datos— se deben proporcionar de manera formal, para que así sea posible evaluarlas automáticamente por máquinas de procesamiento. El objetivo es mejorar Internet ampliando la interoperabilidad entre los sistemas informáticos y reducir la necesaria mediación de operadores humanos.

El precursor de la idea, Tim Berners-Lee, intentó desde el principio incluir informaciones semánticas en su creación, la World Wide Web, pero por diferentes causas no fue posible. Por ese motivo introdujo el concepto de semántica con la intención de recuperar dicha omisión.

Relación con Internet

En la actualidad, la World Wide Web está basada principalmente en documentos escritos en HTML, un lenguaje de marcas que sirve principalmente para crear hipertexto en Internet. El lenguaje HTML es válido para adecuar el aspecto visual de un documento e incluir objetos multimedia en el texto (imágenes, esquemas de diálogo, etc.). Pero ofrece pocas posibilidades para categorizar los elementos que configuran el texto más allá de las típicas funciones estructurales, como sucede con otros lenguajes de maquetación (tipo LaTeX).

HTML permite mediante una herramienta de visualización (como un navegador o un agente de usuario) mostrar por ejemplo un catálogo de objetos en venta. El código HTML de este catálogo puede explicitar aspectos como "el título del documento" es Ferretería Acme; pero no hay forma de precisar dentro del código HTML si el producto M270660 es una "batería Acme", con un "precio de venta al público" de 200 €, o si es otro tipo de producto de consumo (es decir, es una batería eléctrica y no un instrumento musical, o un puchero). Lo único que HTML permite es alinear el precio en la misma fila que el nombre del producto. No hay forma de indicar "esto es un catálogo", "batería Acme" es una batería eléctrica, o "200 €" es el precio. Tampoco hay forma de relacionar ambos datos para describir un elemento específico en oposición a otros similares en el mismo catálogo.

La Web Semántica se ocuparía de resolver estas deficiencias. Para ello dispone de tecnologías de descripción de los contenidos, como RDF y OWL, además de XML, el lenguaje de marcas diseñado para describir los datos. Estas tecnologías se combinan para aportar descripciones explícitas de los recursos de la Web (ya sean estos catálogos, formularios, mapas u otro tipo de objeto documental). De esta forma el contenido queda desvelado, como los datos de una base de datos accesibles por Web, o las etiquetas inmersas en el documento (normalmente en XHTML, o directamente en XML, y las instrucciones de visualización definidas en una hoja de estilos aparte). Esas etiquetas permiten que los gestores de contenidos interpreten los documentos y realicen procesos inteligentes de captura y tratamiento de información.

Componentes de la Web Semántica

Los principales componentes de la Web Semántica son los metalenguajes y los estándares de representación XML, XML Schema, RDF, RDF Schema y OWL. La OWL Web Ontology Language Overview describe la función y relación de cada uno de estos componentes de la Web Semántica:

- XML aporta la sintaxis superficial para los documentos estructurados, pero sin dotarles de ninguna restricción sobre el significado.

- XML Schema es un lenguaje para definir la estructura de los documentos XML.
- RDF es un modelo de datos para los recursos y las relaciones que se puedan establecer entre ellos. Aporta una semántica básica para este modelo de datos que puede representarse mediante XML.
- RDF Schema es un vocabulario para describir las propiedades y las clases de los recursos RDF, con una semántica para establecer jerarquías de generalización entre dichas propiedades y clases
- OWL añade más vocabulario para describir propiedades y clases: tales como relaciones entre clases (p.ej. disyunción), cardinalidad (por ejemplo "únicamente uno"), igualdad, tipologías de propiedades más complejas, caracterización de propiedades (por ejemplo simetría) o clases enumeradas.

La usabilidad y aprovechamiento de la Web y sus recursos interconectados puede aumentar con la web semántica gracias a:

- Los documentos etiquetados con información semántica (compárese ésta con la etiqueta <meta> de HTML, usada para facilitar el trabajo de los robots). Se pretende que esta información sea interpretada por el ordenador con una capacidad comparable a la del lector humano. El etiquetado puede incluir metadatos descriptivos de otros aspectos documentales o protocolarios.
- Vocabularios comunes de metadatos (Ontología (Informática)) y mapas entre vocabularios que permitan a quienes elaboran los documentos disponer de nociones claras sobre cómo deben etiquetarlos para que los agentes automáticos puedan usar la información contenida en los metadatos (p.ej. el metadato author tenga el significado de "autor de la página" y no el del "autor del objeto descrito en la página").
- Agentes automáticos que realicen tareas para los usuarios de estos metadatos de la Web Semántica
- Servicios Web (a menudo con agentes propios) que provean de información a los agentes (por ejemplo un servicio de garantías a quien un agente pudiera consultar sobre si un comercio electrónico tiene un historial de mal servicio o de generar correo basura).

Los proveedores primarios de esta tecnología son las URIs que identifican los recursos junto con XML y los namespaces. Si a esto se añade un poco de lógica, mediante una RDF, u otras tecnologías como los mapas temáticos y algo de razonamiento basado en técnicas de inteligencia artificial, Internet podría estar cerca de alcanzar las aspiraciones iniciales de su inventor, Tim Berners-Lee.

Críticas y reacciones escépticas

El límite de la web semántica está dado no por las máquinas y/o sistemas biológicos que se pudieran usar, sino porque la lógica con que se intenta construir carece del uso del tiempo, ya que la lógica formal es puramente metonímica y carece de la metáfora, y eso es lo que marcan los teoremas de Gödel, la tautología final de toda construcción y /o lenguaje metonímico (matemático), que lleva a contradicciones. Esta lógica consistente es opuesta a la lógica inconsistente que hace uso del tiempo, propia del inconsciente humano, pero el uso del tiempo está construido en base a la falta, no en torno a lo positivo sino en base a negaciones y ausencias, y eso es imposible de reflejar en una máquina porque la percepción de la falta necesita de la conciencia de sí mismo que se adquiere con la ausencia. El problema está en que pretendemos construir un sistema inteligente que sustituya nuestro pensamiento, al menos en las búsquedas de información, pero la particularidad de nuestro pensamiento humano es el uso del tiempo el que permite concluir, por eso no existe en la mente humana el problema de la parada o detención del cálculo, o lo que es lo mismo ausencia del momento de concluir. Así que todos los esfuerzos encaminados a la web semántica están destinados al fracaso a priori si lo que se pretende es prolongar nuestro pensamiento humano en las máquinas, ellas carecen de discurso metafórico, pues sólo son una construcción matemática, que siempre será tautológica y metonímica, ya

que además carece del uso del tiempo que es lo que lleva al corte, la conclusión o problema de la parada.

Como demostración vale la del contraejemplo, o sea que si suponemos que es posible construir la web semántica, como un lenguaje con capacidades similares al lenguaje humano, que tiene el uso del tiempo, entonces si ese es un teorema general, con un solo contraejemplo se viene abajo, y el contraejemplo está dado en el caso particular de la máquina de Turing y el problema de la parada. Luego no se cumple la condición necesaria y suficiente del teorema, nos queda la condición necesaria que es que si un lenguaje tiene el uso del tiempo, carece de logica formal, usa la lógica inconsistente y por lo tanto no tiene el problema de la parada esa es condición necesaria para la web semántica, pero no suficiente y por ello ninguna máquina, sea máquina de Turing, computador o dispositivo aleatorio como un cuerpo negro en física, puede alcanzar el uso de un lenguaje que no sea el matemático con la paradoja de la parada, consecuencia del teorema de Gödel.

Nota en Wikipedia: Se otorga permiso para copiar, distribuir y/o modificar este documento bajo los términos de la Licencia de Documentación Libre GNU, Versión 1.2 o cualquier otra versión posterior publicada por la Fundación para el Software Libre; sin Secciones Invariantes ni Textos de Cubierta Delantera ni Textos de Cubierta Trasera.