

Web semántica. Ontologías, una introducción Ing. Alberto Barrón Cedeño^α

1. Introducción

Internet ha crecido de una manera impresionante y poco controlada, no es necesario contar con información muy relevante o siquiera veraz para tener un sitio web. Sin embargo, cada sitio tiene información potencialmente valiosa.

Sin embargo, en este momento la principal forma en que se accede a cualquier sitio web es por medio de un buscador¹ que no propiamente halla páginas, sino que las ordena con base no precisamente en la relevancia de la información, sino en la frecuencia con que una página es referenciada.

Otro problema grave es que las búsquedas son realizadas por medio de cadenas estáticas, realizando comparaciones exactas entre la búsqueda y los recursos. Así, si la cadena buscada es “computadora”, entre los resultados difícilmente habrá una página española, pues en aquel país, sólo se habla de “ordenadores”. Muchísimas páginas que pueden ser valiosas, quedan prácticamente fuera de alcance. Este hecho no sólo se ve reflejado en cuanto a diferencias geográficas o variedades dialectales, sino a estratos sociales y a perfiles educativos.

Es por esta razón que se busca dar un “significado” al web. La idea es no estar limitado a búsquedas rígidas de información, sino de asociar a las palabras con su concepto, sinónimos e incluso, reflejar la cercanía semántica de una palabra con respecto a otra². De esta manera, si de nuevo se realiza una búsqueda con base en “computadora”, se puede esperar no sólo contar con páginas que hablen de ordenadores, sino con otras que hablen de sistemas operativos, lenguajes de programación o electrónica.

Para lograr que el web cuente con estas propiedades, es necesario tomar en cuenta algunos aspectos que podrían incluso ser considerados ajenos a la computación (pero que definitivamente no lo son), como la lingüística, y para el caso particular de las ontologías, la terminología.

El objetivo del presente escrito no es realizar un extenso tratado de terminología, sin embargo, podemos señalar que para Dubuc la terminología es “el estudio sistemático de los términos o palabras y sintagmas³ especiales que sirven para denominar clases de objetos y conceptos”⁴. Los términos tienen una relación indisoluble con un concepto asociado. Sin embargo, dicha asociación sólo es válida dentro de un contexto en particular (un área de conocimiento). Definitivamente no es lo mismo hablar de una bomba en un contexto de autos que en uno de guerra⁵.

Al ser elementos del lexicón de áreas específicas de conocimiento y de sus expertos, los términos requieren ser ordenados con base no sólo en sus características semánticas, sino con respecto al dominio al que pertenecen. Las taxonomías y ontologías sirven, además de para contar con definiciones más precisas y claras, para ordenar los términos; *ordenar el conocimiento*. A

^α Posgrado en Ciencia e Ingeniería de la Computación, UNAM.

¹ Como Google o Altavista.

² Por ejemplo, computadora es cercana semánticamente a programa, televisión lo es a radio.

³ Una o más palabras asociadas a un mismo concepto.

⁴ Dubuc R. Manual práctico de terminología p. 10

⁵ Para profundizar en el tema puede consultar obras de Cabré, Wüster o Dubuc (incluidas en la bibliografía de este trabajo)

continuación, veremos de qué se trata una taxonomía.

2. Taxonomías

Una taxonomía es una organización jerárquica basada en un conjunto de atributos de los elementos clasificados. Dicha organización tiene la estructura de un árbol dirigido, por lo que no sólo no tiene ciclos, sino que un elemento no puede pertenecer al mismo tiempo a más de una clasificación (rama) dentro de la organización.

Hablando de manera más concreta, en una taxonomía se crean categorías jerárquicas para organizar las cosas en mapas simples del mundo. Como se señaló anteriormente, se trata de organizar el conocimiento, lo cual no es privativo de las taxonomías, sino de otras estructuras como los tesauros⁶ o como veremos más adelante, las ontologías.

Sin embargo, el hecho de que se traten de árboles dirigidos es un grave problema, pues en el mundo real, difícilmente un ente puede ser clasificado de manera tajante en un sólo lugar. Por ejemplo, un murciélago mexicano tiene la capacidad de volar, al igual que un águila calva (aunque tal vez ésta lo haga con mayor gracia), sin embargo, el murciélago es más cercano biológicamente a un perro que al águila (pues a diferencia de ésta, es mamífero y vivíparo). ¿En dónde debería ser colocado el murciélago en una taxonomía de animales, entre los que son capaces de volar o dentro de los mamíferos?

Es en este punto donde entran las ontologías.

3. Ontologías

Una ontología sirve para definir los términos usados para describir y representar un área de conocimiento. Incluye definiciones de conceptos de especialidad además de las relaciones entre ellas. Además, son accesibles *por la computadora*⁷.

Para lograr su objetivo, el web semántico requiere de ontologías que cuenten con un alto grado de estructuración para especificar descripciones de clases (las cosas), relaciones entre clases y las propiedades que éstas tienen.

Las ontologías son útiles para diversas aplicaciones inteligentes como es el caso de recuperación de información, tratamiento de lenguaje natural o comercio electrónico. Si el objetivo se logra, dichas aplicaciones inteligentes trabajarán en un nivel conceptual más cercano al del ser humano, lo cual permitiría confiar a un agente la adquisición de información.

Si bien, otros enfoques, como el de XML, han logrado cierta aproximación a este manejo de la información, no ha resultado suficiente, por lo que han surgido otras herramientas como RDF, con el que es posible crear ontologías limitadas pero útiles.

Con el afán de ser un tanto más específicos en el uso que en la actualidad se da ya a las ontologías, podemos mencionar portales web como el *dmoz open directory project*⁸ que cuenta con una ontología con la terminología de los temas que contiene y permite realizar búsquedas más precisas que aquellas realizadas con buscadores basados en frecuencia y cadenas específicas.

Otro caso de uso de las ontologías es la administración de colecciones multimedia en las que a cada archivo de imagen, vídeo o audio, le puede ser asociada información semántica que permita

⁶ Colección de vocablos, sin su definición, clasificados en una jerarquía, se le puede considerar una estructura arbórea

⁷ No se entienda solamente como la capacidad de que un programa logre acceder a la información, sino de que realmente sepa de qué se trata lo que está obteniendo para lograr una mejor explotación

⁸ Véase dmoz.org

acceder a él de manera más rápida y precisa.

La administración de sitios corporativos es otro caso de éxito, pues permite que cualquier elemento de la empresa (o ajeno a ella), logre obtener la información que requiere sin importar las características de la búsqueda.

A diferencia de una taxonomía, una ontología es un grafo sin alguna jerarquía en particular, y definitivamente no es un árbol, por lo que el problema mencionado anteriormente deja de existir.

Por último, es importante señalar que no se trata de generar conocimiento, el conocimiento ya está en el web, se trata de clasificarlo para poder hallarlo de manera más eficiente.

Como se señaló antes, RDF es una herramienta que puede ser utilizada para crear ontologías, sin embargo, no es suficiente. En la siguiente entrega, se observará una ontología verdadera además de uno de los lenguajes creados específicamente para realizarlas: OWL. Además, se verá un análisis de los diversos editores que existen para ello y se intentará dar una justificación para recurrir a una herramienta que no garantiza ser enteramente computable.

Referencias.

Cabré, M. T. La Terminología. Teoría, métodos, aplicaciones. Barcelona, ed. Antártida, 1993

Dubuc, Robert. Manual práctico de terminología. 3º edición. Pontificia Universidad Católica de Chile.

McComb, Dave. Semantics in business systems. Morgan Kaufmann ed. EUA, 2004

Wüster, Eugen. Introducción a la Teoría General de la Terminología y a la lexicografía terminológica. Barcelona, IULA 1998.

dmoz open directory project. Netscape, 2005

<http://dmoz.org>

consulta: septiembre de 2005

Expressing Qualified Dublin Core in RDF / XML. Dublin Core Metadata Initiative, 2005

<http://dublincore.org/documents/dcq-rdf-xml/>

consulta: septiembre de 2005

OWL Web Ontology Language Guide. W3C Recommendation, 10 February 2004

<http://www.w3.org/TR/owl-guide/h>

consulta: septiembre de 2005

OWL Web Ontology Language. XML Presentation Syntax. W3C Note, 11 June 2003

<http://www.w3.org/TR/owl-xmlsyntax/>

consulta: septiembre de 2005

Web Ontology Language (OWL). W3C, 2005

<http://www.w3.org/2004/OWL/>

consulta: septiembre de 2005