

Búsqueda sobre catálogos basada en ontologías

Alianis Pérez Sosa, Yuniel Eliades Proenza Arias

Universidad de las Ciencias Informáticas. Carretera a San Antonio Km 2 $\frac{1}{2}$,

Reperto Torrens, La Lisa, Ciudad de la Habana. Cuba

apsosa@estudiantes.uci.cu, yproenza@uci.cu

Resumen

En la actualidad implementar sistemas capaces de realizar búsquedas en lugares donde existe gran cantidad de información es un reto para los desarrolladores de software. Con la introducción y utilización de las ontologías en los procesos que conforman la fase de organización y representación de la información, así como en el propio proceso de búsqueda, estas mejoran y el usuario obtiene lo que necesita. Las búsquedas ontológicas basadas en esquemas conceptuales generan resultados coherentes con el criterio del usuario sobre uno o varios dominios de información. Todo ello depende de que la información se describa mediante un mecanismo de representación del conocimiento donde los datos se relacionen entre sí, pasando a ser metadatos de información. El objetivo de este trabajo es explicar la importancia de realizar búsquedas sobre catálogos a través de ontologías. Además se exponen algunas de las metodologías, los lenguajes y herramientas mayormente utilizadas en la elaboración de este tipo de representación de la información.

Palabras Claves: *software, ontología, búsquedas ontológicas, representación del conocimiento, metadatos*

Introducción

En los lugares donde se almacena gran cantidad de información heterogénea es necesario contar con un método para recuperar la misma, por lo que se desarrollaron los buscadores de información. Existen fuentes de información como la Web que poseen buscadores que le devuelven al usuario la información solicitada. Estos motores de búsquedas realizan búsquedas sintácticas, devolviendo solo la información en la cual aparecen elementos constituyentes a la consulta realizada, tal y como ha sido especificada, provocando que en muchas ocasiones las respuestas arrojadas no sean las esperadas. Para solucionar estos problemas se pretende transformar la Web actual en una Web semántica (Castells, 2003), basándose en esta idea surge el término ontología que hasta principio de los años 90 solo era conocido en el campo de la inteligencia artificial.

Una ontología define un vocabulario común para quien necesita compartir información dentro de un determinado dominio, esto incluye definiciones de los conceptos básicos del dominio, así como sus relaciones. El objetivo principal de las ontologías es representar la información del dominio en cuestión a través de conceptos organizados jerárquicamente a partir de su significado, logrando un cierto grado de razonamiento automático que le permitirá a las máquinas entender por sí solas que es lo que el usuario necesita. De esta forma se podría intercambiar información de forma efectiva y eficiente.

Ontologías

Ontología en informática hace referencia a la formulación de un exhaustivo y riguroso esquema conceptual dentro de uno o varios dominios dados, para facilitar la comunicación y el intercambio de información entre diferentes sistemas y entidades. Según (Castells, 2003) *"Una ontología es una jerarquía de conceptos con atributos y relaciones de un dominio en específico, que define una terminología consensuada para definir redes semánticas de unidades de información interrelacionadas. Una ontología proporciona un vocabulario de clases y relaciones para describir un dominio, poniendo el acento en la compartición del conocimiento y el consenso en la representación de éste"*.

Basándose en esta definición, una ontología se puede definir como la creación de un modelo conceptual donde se establecen los conceptos relacionados con el dominio donde se va a desarrollar y sus relaciones. El objetivo sería crear una red semántica que le permita a la máquina entender lo que significa el término establecido en el criterio de búsqueda y pueda navegar por la jerarquía de conceptos definida, logrando ofrecer mejores resultados en las respuestas a las búsquedas realizadas. Existen varios criterios para clasificar una ontología, los mismos son expuestos en (Nuñez, 2007) .

Uso de las ontologías

En la actualidad existe un auge en el desarrollo de ontologías tanto en ámbitos científicos como empresariales. Según Ramos y Núñez (Nuñez, 2007) una ontología puede funcionar como un marco para la unificación de los diferentes puntos de vista y servir como base para:

- La comunicación entre personas con diferentes necesidades y puntos de vista.
- La interoperabilidad entre sistemas alcanzada por una traducción entre diferentes modelos, métodos, paradigmas, lenguajes y herramientas de software.

- Apoyar a la ingeniería de sistemas al momento de reusar un software, hacer el chequeo de consistencia (confiabilidad del software), y al momento de la especificación, sobre todo a la hora de levantar los requerimientos.

La principal utilidad de una ontología es ayudar a la compartición del conocimiento entre diferentes actores de un dominio determinado, como pueden ser personas, organizaciones y sistemas de software. Por tanto, las ventajas en el desarrollo de las ontologías son innumerables ya que presentan un alto nivel de descripción del vocabulario, fomentan la comunicación entre las personas y las máquinas, y realizan un desarrollo semántico profundo en los niveles de la jerarquía, logrando que las máquinas lleguen a un razonamiento que le permita obtener respuestas inteligentes.

Seleccionando la ontología adecuada

Varias ontologías pueden estar conectadas compartiendo información de diferentes dominios, como también una ontología existente pueda ser reutilizada por agentes de software que requieran dicha información, logrando un entendimiento compartido. Una ontología de un dominio determinado puede ser reutilizada por cualquier persona de acuerdo a sus necesidades, ahorrándose así tiempo y esfuerzo al poder adquirir parte del trabajo ya realizado. El proceso para seleccionar una ontología para ser reutilizada es bastante complicado, el usuario antes de seleccionarla debe analizar el contenido que la misma proporciona, los aspectos metodológicos y los tecnológicos.

En (Tello, 2002) se exponen algunos trabajos publicados por varios autores relacionados con las características del contenido de las ontologías, de los lenguajes usados en su construcción. El lenguaje en el que está desarrollado la ontología es un punto importante a la hora de tomar la decisión, puesto que puede ser que el sistema este implementado en un lenguaje que no sea compatible con la ontología seleccionada. Otro aspecto importante es la metodología que se siguió en su desarrollo, ya que una ontología es más fiable si en su construcción se siguió una metodología y la incorporación de nuevos términos será más fácil.

Cada ontología por lo general es creada con diferentes propósitos por lo que su dominio es diferente, es por esto que los esquemas de características analizados por los autores son difícilmente comparables. Existen otros métodos como *OntoMetric* que ayudan al usuario a seleccionar la ontología para reutilizar en su sistema, pero es un método que requiere demasiado

tiempo y cálculos por lo que se hace muy difícil para el usuario tomar una decisión.

Existen otros problemas que hacen complejo el proceso de selección de ontologías, por ejemplo, muchas de ellas se encuentran en la Web y puede suceder que cuando el usuario quiera acceder a la información la aplicación no logre conectarse con la ontología. Por las razones antes expuestas es que se recomienda desarrollar una ontología propia y no reutilizar las existentes, aunque se podrían reutilizar algunas taxonomías de conceptos ya definidas.

Construcción de ontologías.

Los profesionales de las tecnologías de la información enfrentan grandes desafíos al tener que administrar grandes cantidades de información heterogénea. Es por eso que necesitan ser dotados de las herramientas y tecnologías necesarias para poder mantener un mayor control de la información y así se agilizaría el proceso de gestión.

Entre las herramientas que se utilizan en la construcción de ontologías, la más utilizada es Protégé (Nuñez, 2007). Esta herramienta es un software libre de código abierto implementado en Java. Protégé brinda la posibilidad de crear clases y jerarquía de clases, declarar propiedades para las clases, crear instancias e introducir valores. Posee un entorno abierto y fácil de extender, tiene su propio lenguaje interno para definir ontologías, pero permite también trabajar con RDF y OWL. La herramienta puede correr en una aplicación local o a través de un cliente en una comunicación con un servidor remoto; permite a los usuarios compartir, crear y editar ontologías. Permite añadir plug-ins para adicionar nuevas funcionalidades o mejorar las existentes, visualiza los conceptos y relaciones de diferentes formas, entre ellas: árbol jerárquico de las clases y subclases de la ontología.

Existen lenguajes para codificar la semántica de los recursos haciéndole descripciones a través de ontologías. Entre los más usados están RDF (*Resource Description Framework*) y OWL (*Web Ontology Language*) ambos basados en la capacidad de representación de XML. RDF es un lenguaje creado por el W3C (*World Wide Web Consortium*) para representar datos en la Web, se basa en la idea de que los objetos a describir poseen propiedades que a su vez tienen valores. Permite definir relaciones semánticas entre distintas URIs asociándoles un conjunto de propiedades y valores. Está lenguaje está basado en triplas de la forma sujeto-predicado-objeto.

Recursos (*sujeto*): Son todos los objetos que pueden ser representados con expresiones RDF (páginas web, documentos, etc.). Es todo lo que puede ser referenciado por un identificador único de recursos (URI).

Propiedades (*predicado*): Atributo o relación que permite describir un recurso.

Expresión (*objeto*): Un recurso o bien un literal específico junto con el nombre de una propiedad, más el valor de ésta formarán una expresión RDF.

OWL como su nombre lo indica es un lenguaje para representar ontologías en la Web, posee un nivel de representación semántica de la información mayor que RDF. Este lenguaje añade más vocabulario para describir propiedades y clases, relaciones entre clases, cardinalidad, igualdad, más tipos de propiedades, características de propiedades y clases enumeradas. Existen dos versiones OWL-1 y OWL-2.

En el proceso de construcción de ontologías es aconsejable emplear una adecuada metodología, siguiendo los procesos de desarrollo, las actividades y técnicas indicadas. Las metodologías proporcionan un conjunto de directrices que indican cómo hay que llevar a cabo las actividades identificadas en el proceso de desarrollo, qué técnicas son las más apropiadas en cada actividad y qué produce cada una de ellas.

Una de las metodologías más utilizadas es Methontology (Nuñez, 2007) recomendada por FIPA (*Fundación de Agentes Físicos Inteligentes*) la cual tiene sus raíces en las actividades identificadas por la IEEE para el proceso de desarrollo de software. Esta metodología está basada en prototipos evolutivos permitiendo añadir, cambiar y eliminar términos en cada nueva versión de la ontología.

Las ontologías en el proceso de búsqueda

Las ontologías permiten estructurar la información para agilizar el proceso de búsqueda. Las Búsquedas ontológicas son búsquedas basadas en ontología, estas se caracterizan por el uso de semántica explícita, se realizan inferencias basadas en lógica, a diferencia de la búsqueda tradicional donde se trata de encontrar los resultados sin tratar de comprender el término de búsqueda especificado. La búsqueda semántica le permite al usuario y aplicación entender qué significa el término establecido en la consulta en base al conocimiento semántico explícito dado en la ontología. Esto permite relacionar el término de búsqueda con un concepto del dominio cuyo significado esté definido y con ello conocer las propiedades y las relaciones que dicho concepto posee. Estas búsquedas le darán la posibilidad al usuario de obtener varias respuestas en base a su solicitud.

Búsquedas basadas en ontologías sobre catálogos temáticos

Cuando el término ontología empezó a ser usado se pensó en la idea de que se desarrollara para realizar búsquedas en la Web, lo que implicaría representar semánticamente el contenido de la Web. Una ontología también puede realizar búsquedas sobre un catálogo temático, es decir, sobre a un conjunto de datos ordenados y clasificados, en este caso por temáticas. Una ontología que funcione sobre un catálogo permitirá describir completamente el esquema de la base de datos que se desee representar mediante la instanciación de la misma, por lo que se convierte en un intermediario entre la aplicación y la base de datos.

Los usuarios especifican su criterio de búsqueda en función de conceptos y atributos que se establecen inteligentemente; una vez establecida la consulta se navega por la estructura jerárquica que define la ontología para recuperar los términos que tengan relación con la petición realizada. De esta forma se encontraría no solo la respuesta de la palabra clave, sino también otros términos de igual significado, puesto que existen disímiles términos para referirse a un mismo elemento y en muchas ocasiones el usuario no encuentra lo que busca.

Conclusiones

La búsqueda basada en ontología es un paso importante en la generación de resultados más precisos y con niveles mayores de exactitud en la información. Mientras más ordenada y mejor procesada se encuentre la información será más fácil desarrollar la selección y posterior adquisición. Las principales ventajas de esta nueva evolución en la tecnología es el desarrollo de aplicaciones con esquemas de datos comunes y búsqueda de información con inferencias. Desarrollar una ontología para su uso en el proceso de búsqueda sobre catálogos temáticos a través de palabras claves, le brinda a la aplicación la posibilidad de devolver la información requerida por el usuario.

Referencias Bibliográficas

Castells, Pablo. 2003. *"La Web Semántica"*. Universidad Autónoma de Madrid : s.n., 2003.

Núñez, Esmeralda Ramos y Haydemar. 2007. *"Lecturas en Ciencias de la Computación"*. Caracas- Venezuela : Facultad de Ciencias-Escuela de Computación : s.n., 2007.

Tello, Adolfo Lozano. 2002. "*Métrica de Idoneidad de ontologías*".
Cáceres : s.n., 2002.

SOCIEDAD DE LA INFORMACION

www.sociedadelainformacion.com

Edita:



Director: José Ángel Ruiz Felipe

Jefe de publicaciones: Antero Soria Luján

D.L.: AB 293-2001

ISSN: 1578-326x