

Ontologías y la Búsqueda de Información

Lucy Medina Velandia

Universidad Sergio Arboleda, Bogotá, Colombia, lucy.medina@usergioarboleda.edu.co

Andrés Abel Arenas Prada

Universidad Sergio Arboleda, Bogotá, Colombia, andres.arenasa@usergioarboleda.edu.co

RESUMEN

En el presente documento se escribe sobre las ontologías, las normas para desarrollarla, las características de los programas y los lenguajes asociados a éstas. En la primera parte se describe el problema, el cual trata básicamente del Por qué se crea la ontología. El estudio se realiza en Colombia, en la universidad Sergio Arboleda y la línea de investigación trabajada es el desarrollo de Software, de la cual hace parte esta investigación. Finalmente se da a conocer en qué fase se encuentra el proyecto, qué objetivos se han cumplido y cuáles son las siguientes fases a seguir para culminar este proyecto.

Palabras claves: Ontología, tesauros.

ABSTRACT

This paper is written about ontologies, rules develop the characteristics of the programs and associated languages to them. In the first part describes the problem, which is Why basically creates the ontology. The study was conducted in Colombia, at the Sergio Arboleda University and the research work is the development of software, which is part of this research. Finally we explain at what stage is the project, which objectives have been met and what are the next phases continue to complete this project.

Keywords: Ontology, thesaurus.

1. INTRODUCCIÓN

En la actualidad se pueden encontrar millones de documentos en la web que están a disposición de los usuarios en cualquier momento; teniendo la posibilidad de modificar y publicar un sin número de información dentro de la red; precisamente por el continuo crecimiento de contenidos en la red se hace cada vez más difícil encontrar, organizar, acceder y mantener la información exigida por el usuario, y teniendo que optar por entrar a varios sitios haciendo que la búsqueda se vuelva larga y no se culmine con éxito. Por lo anterior se crean las llamadas ontologías para que las solicitudes de información en la web se tomen factibles, prácticas, rápidas y de esta manera el usuario pueda encontrar las indagaciones que necesita de una forma fácil y acertada.

La universidad Sergio Arboleda en conjunto con la línea de investigación Desarrollo de Software se ha propuesto crear una ontología, con el fin de ayudar a los usuarios que necesiten investigar y aprender sobre el tema relacionado con algoritmia.

En el artículo se expondrá la descripción del proyecto, parta del estado del arte que se recopiló y las fases que lo componen para desarrollar la ontología; se hace una reseña de los programas y lenguajes que se utilizan para desarrollar ontologías y conclusiones.

2. DESCRIPCIÓN DEL PROYECTO

El proyecto tiene como función crear una ontología en la cual se reúnan los conocimientos y conceptos que giran en torno a la algoritmia en el área de sistemas, lo cual generará que el conocimiento se consolide y a con la unión de un sistema multiagentes se cree un tutor con el cual se ayude al aprendizaje y consolidación de la formación del alumnos en el campo de los sistemas.

La investigación se ha dividido en seis fases:

- Fase uno: En esta fase el investigador inicia la indagación y recolección de información relacionada con la ontología y todos los conceptos y definiciones que la rodean. También se revisaron y averiguaron algunos programas que se emplean para desarrollar ontologías, entre ellos: KIF, OWL, XML, DAML, RDF, SHOE, OIL.
De otra parte, se hizo un estudio sobre algunas herramientas que se utilizan para desarrollar las ontologías, entre las que se compararon: Protegé, RDF-Edit, Ontosaurus, kaon ,Onto-Edit, Joe, y Apollo para elegir la mejor herramienta la cual cumpla las expectativas de la investigación.
- Fase dos: Con un panorama más claro acerca del problema, se procedió a construir la pregunta problema (origen de la investigación) en la cual se definieron los objetivos principales y específicos y se realizó la justificación del tema de investigación.
Se definieron las Fases y se construyó el cronograma que guió el desarrollo del proyecto.
- Fase tres: Se realizó un conversatorio o reunión de grupo, donde se exponen varias ideas con los estudiantes de primero y segundo semestre de Ingeniería de Sistemas, Ingeniería Industrial e Ingeniería Electrónica de la Escuela de Ingenierías de la Universidad Sergio Arboleda, con el fin de identificar cuáles son las asignaturas y temas de mayor dificultad en sus carreras. Una vez obtenidos los resultados de este estudio, se concluyó que los alumnos presentaban mayor dificultad en aquellas asignaturas referidas especialmente a “sistemas”, así como también cierta complejidad en los temas relacionados con “algoritmia”.
- Fase cuatro: Dado que el investigador ha realizado todas las averiguaciones pertinentes sobre el tema, y tiene una perspectiva más idónea del problema y su posible solución y habiendo hecho previamente una selección del programa que puede ofrecer todas las herramientas que el investigador necesita, se hace factible empezar a crear la ontología.
- Fase cinco: Se finaliza la ontología y se tiene lista para ser unida con el “agente inteligente” que permitirá crear un tutor, que asista a todas las personas interesadas en los temas referidos a la algoritmia, los encuentre y ofrezca información clara, procedente de una fuente real, y sobretodo fundamentada.
Se realizaron las pruebas pertinentes, con el objeto de integrar la ontología, los agentes y búsquedas.
- Fase seis: En esta fase en la que se inició la difusión del producto terminado y capacitación de los estudiantes de los primeros semestres de la Escuela de Ingenierías de la Universidad Sergio Arboleda.

3. DESARROLLO DEL PROYECTO RECUPERACIÓN DE INFORMACIÓN EN LA WEB A TRAVÉS DE AGENTES INTELIGENTES

La universidad Sergio Arboleda en la cual se encuentra conformado el grupo de investigación Desarrollo Tecnológico y Ciencias del Espacio, al cual está vinculada la línea Desarrollo de Software, a la que pertenece el semillero de investigación, ha dividido el gran proyecto que consiste en crear un tutor Inteligente, compañero de estudio del tema algoritmos en varios frentes de trabajo.

De una parte, algunos integrantes del semillero están indagando sobre el componente pedagógico, en cuanto hace referencia a enseñanza – aprendizaje, de otra parte, se está creando la ontología sobre el tema de algoritmos, un tercer grupo trabaja sobre el correcto funcionamiento de los agentes inteligentes y por último otros integrantes del semillero se encargan de revisar cómo deberá fusionarse la ontología y los agentes para brindar el mayor y mejor uso.

4. MARCO TEÓRICO / ESTADO DEL ARTE

4.1 LOS TESAUROS

La palabra tesauros tiene su origen del latín thesaurus que significa tesoro o colección de palabras.

En el presente proyecto se utiliza el término tesauro para referirnos a una lista o un conjunto de términos que representan un concepto que a futuro se llamará clase principal, lo cual permite organizar y encerrar en un marco, un término o un concepto para que más adelante permita el entendimiento del que desee buscar concepto rápida y fácilmente.

En líneas generales, un tesauro comprende lo siguiente:

- Un listado de términos preferidos, que se ordenan en forma alfabética, temática y jerárquicamente.
- Un listado de sinónimos de esos términos preferidos, llamados descriptores, con la leyenda "útese (término preferido)" o una indicación similar.
- Una jerarquía o relaciones entre los términos. Esto se expresa con la identificación de "términos más generales" y "términos más restringidos".
- Las definiciones de los términos, para facilitar la selección de los mismos por parte del usuario.

4.2 LAS ONTOLOGÍAS

El término ontología se utiliza inicialmente en filosofía en el campo de la metafísica, que de un lado estudia la esencia del ser, y por otro, estudia las características básicas de la realidad como un todo.

En la actualidad en el campo de la filosofía que tiene como propósito la explicación de las existencias de una manera sistemática, trata sobre los tipos y estructuras de objetos, propiedades, eventos, procesos y relaciones relativas a cada porción de la realidad.

4.2.1 ¿QUÉ ES UNA ONTOLOGÍA?

En la filosofía se define como parte de la metafísica, que estudia el ser en general y sus propiedades trascendentales.

En este caso una ontología es una descripción rigurosa y formal de uno o varios conceptos dados, los cuales forman una estructura de conocimiento clara y fácil de entender para el usuario y esto hace que él pueda ampliar y reutilizar el concepto.

En términos generales una ontología está conformada por:

- Clase principal.
- Subclases
- Jerarquía de las clases.



Figure 1: Ejemplo de Ontología

En esta figura se ve la representación de una ontología muy superficial en la cual se nota la clase principal, que es vehículo de la que se dividen en tres subclases terrestres aéreas y marítimas y dentro de ellas se encuentran elementos de cada una de estas subclases.

4.2.2 ONTOLOGÍAS EN COLOMBIA

En esta investigación se inicia una fase de búsqueda de información y de consolidación de los datos, en el transcurso de esta tarea se encontró que gran parte de los desarrollos de las ontologías, programas y distintos planes son realizados en universidades como los Andes, Universidad Pontificia Javeriana, Universidad Nacional de Colombia y la Escuela de Ingenieros Julio Garavito, pero son pequeñas ontologías que no se han consolidado totalmente,

También se percibió a través de la indagación que en la mayor parte de los desarrollos de software que se refieren a ontologías, se hace en países diferentes, como por ejemplo en universidades europeas y americanas, por esta razón la Universidad Sergio Arboleda y la Línea de Investigación Desarrollo de Software, dirige su atención y esfuerzos a desarrollar la ontología sobre el tema de algoritmia, con el objeto de colaborar en el estudio y mejorar el entendimiento del tema, principalmente en los primeros semestres de las carreras: Ingeniería Industrial, Electrónica e Ingeniería de Sistemas y Telecomunicaciones, en las que se evidenció, empíricamente (luego de sostener diálogo con los estudiantes) cómo tienen falencias en las asignaturas que abarcan la disciplina de la algoritmia; dado que los estudiantes afirman que les acarrea problemas de análisis, comprensión y desarrollo, lo que genera cierta desvió a las asignaturas; como consecuencia, se presenta mucha deserción. De ahí que el semillero haya tomado la decisión de trabajar en pro de los estudiantes, con el objeto de facilitar a los alumnos de primeros semestres de la Escuela de Ingenierías un material que podrá ayudarles a entender, analizar y crear algoritmos de una manera más sencilla.

4.2.3 ¿POR QUÉ DESARROLLAR ONTOLOGÍAS?

En la actualidad se presenta un aumento en los usuarios que acceden a la web con el fin de hacer una búsqueda sobre un tema específico, por lo tanto existe la posibilidad de que éstos agreguen, publiquen y modifiquen la información. El problema generado por esta situación es que cada vez hay más información y ésta no está debidamente organizada ni bien dirigida para hacer una búsqueda; por esta razón al crear ontologías, se pretende generar una estructura clara y estable, lo que permite que se pueda acceder más fácil al conocimiento y crear una comunidad mundial que comparta información del tema en cuestión. dado que las ontologías son una estructura que se puede reutilizar y hacer cada vez más grande teniendo en cuenta la facilidad que tienen para absorber el conocimiento a nivel mundial.

Los beneficios de utilizar ontologías se pueden resumir de la siguiente forma:

- Proporcionan una forma de representar y compartir el conocimiento utilizando un vocabulario común
- Permiten usar un formato de intercambio de conocimiento
- Proporcionan un protocolo específico de comunicación
- Permiten una reutilización del conocimiento.

Como resultado de las primeras fases de la investigación, se encontró que en Colombia, el tema de ontologías no ha sido muy desarrollado ni implementado, por ese motivo la Universidad Sergio Arboleda ha decidido crear esta ontología teniendo en cuenta la problemática descrita a través del artículo.

5. LENGUAJES PARA DESARROLLO DE ONTOLOGÍAS A NIVEL WEB

5.1 LENGUAJE OWL

OWL Web Ontology Language o Lenguaje de Ontologías para la Web es un lenguaje de etiquetado semántico para publicar y compartir ontologías en la Web. Se trata de una recomendación del W3C (World Wide Web Consortium), y puede usarse para representar ontologías de forma explícita, es decir, permite definir el significado de términos en vocabularios y las relaciones entre aquellos términos (W3C, 2009).

5.2 LENGUAJE SHOE

Simple HTML Ontology Extensions fue el primer lenguaje de etiquetado para diseñar ontologías en la Web. Este lenguaje nació antes de que se ideara la Web Semántica. Las ontologías y las etiquetas se incrustaban en archivos HTML. Este lenguaje permite definir clases y reglas de inferencia, pero no negaciones o disyunciones. A su albur se desarrollaron muchos editores, buscadores, APIS, etc.; pero este proyecto fue abandonado a medida que se desarrollaron OIL y DAM; aunque también existe una serialización de este lenguaje en XML

5.3 LENGUAJE OIL

Ontology Inference Layer es un, derivado en parte de SHOE, fue impulsado también por el proyecto de la Unión Europea On-To-Knowledge. Utiliza ya la sintaxis del lenguaje XML y está definido como una extensión de RDFS. Se basa tanto en la lógica descriptiva (declaración de axiomas) y en los sistemas basados en frames (taxonomías de clases y atributos). OIL posee varias capas de sub-lenguajes, entre ellas destaca la capa base que es RDFS, a la que cada una de las capas subsiguientes añade alguna funcionalidad y mayor complejidad. La principal carencia de este lenguaje es la falta de expresividad para declarar axiomas.

Procesada por programas o aplicaciones, en oposición a situaciones donde el contenido solamente necesita ser presentado a los seres humanos. OWL surge como una revisión al lenguaje DAML-OIL y es mucho más potente que éste. Al igual que OIL, OWL se estructura en capas que difieren en la complejidad y puede ser adaptado a las necesidades de cada usuario, al nivel de expresividad que se precise y a los distintos tipos de aplicaciones existentes (W3C, 2009).

5.4 LENGUAJE KIT

Knowledge Interchange Format es un lenguaje para representar ontologías basadas en la lógica de primer orden. KIF está basado en la lógica de predicados con extensiones para definir términos, meta conocimiento, conjuntos, razonamientos no establecidos, etc.; y pretende ser un lenguaje capaz de representar la mayoría de los conceptos y distinciones actuales de los lenguajes más recientes de representación del conocimiento. Se trata de un lenguaje diseñado para intercambiar conocimiento entre sistemas de computación distintos, diferentes lenguas, etc.; y no para la interacción entre seres humanos (W3C, 2009).

5.5 LENGUAJE FOAF

Aunque no es exactamente un lenguaje de ontologías ya que se trata de un vocabulario con definiciones que usa el lenguaje RDFS/OWL, FOAF hace más fácil que el software procese los términos del vocabulario FOAF para describir documentos. FOAF permite crear una base de datos unificada de información al normalizar una forma de definir categorías, tipos de relaciones (W3C, 2009).

6. CONCLUSIONES

- Una ontología puede estar construida desde un elemento general hasta sus elementos más específicos o viceversa, dependiendo del tipo de usuario y la necesidad por la cual se desarrolle la ontología.
- Las ontologías sobre los temas de algoritmia que han sido cerradas por las universidades en Colombia, no han sido desarrolladas completamente, para generar la ayuda necesaria al usuario y facilitar las búsquedas y entendimiento de conceptos.
- Después de bajar, montar y probar los distintos programas para desarrollar las ontologías, el que ofrece todos los elementos que suple las necesidades del proyecto es el Protegé.
- Gran parte de los lenguajes creados y fusionados son enfocados en la utilización de ellos en un entorno web.
- Se describieron y aclararon los conceptos y términos de los tesauros mostrando, que se tienen en cuenta en los primeros pasos para la creación de una ontología.

REFERENCIAS

Kaon.(2009). “software Kaon ” <http://kaon.semanticweb.org/>,10/12/2009,10/12/2009.
OntoAgente. (2009). “software OntoAgente” <http://www.i-u.de/schools/eberhart/ontoagent>, 09/12/2009.
Ontoedit. (2009). “software Ontoedit”[Http://www.ontoknowledge.org/tools/ontoedit.shtml](http://www.ontoknowledge.org/tools/ontoedit.shtml), 10/12/2009.
Ontology matching. (2009), “software Ontology matching” <http://www.ontologymatching.org/>, 10/12/2009.
Ontotext .(2009). “software Ontotext” <http://www.ontotext.com/>, 10/12/2009.
OntoAgente. (2009). “software OntoAgente” Passenger Data for 1990-2000, <http://www.i-u.de/schools/eberhart/ontoagent>, 09/12/2009.
Odra.(2009). “software Odra” <http://www.mindswap.org/~golbeck/downloads/odra.shtml>,10/12/2009.
Protégé.(2009). “software Protégé”<http://protege.semanticweb.org>,10/12/2009.
Sewese.(2009). “software sewese” <http://www-sop.inria.fr/edelweiss/wiki/wakka.php?wiki=Software>,10/12/2009.
W3C (2009). “Información general de las ontologías” <http://www.w3.org/TR/owl-features/>
W3C (2009). “Lenguaje OIL” <http://www.w3.org/TR/daml+oil-reference>,02/04/2009
W3C (2009). “Lenguaje OWL” <http://www.w3.org/2007/09/OWL-Overview-es.html>,02/04/2009
W3C (2009). “Panorama general de las ontologías” <http://www.w3.org/TR/2009/WD-owl2-overview-20090327/>,02/04/2009
W3C (2009). “Preguntas frecuentes sobre OWL” <http://www.w3c.es/traducciones/es/sw/2005/owlfaq>,02/04/2009.
W3C (2009). “lenguaje KIF” <http://www.w3.org/2000/07/hs78/KIF.html>,02/04/2009

Autorización y Renuncia

Los autores autorizan a LACCEI para publicar el escrito en las memorias de la conferencia. LACCEI o los editores no son responsables ni por el contenido ni por las implicaciones de lo que esta expresado en el escrito.