

Tipo de artículo: Artículo original
Temática: Desarrollo de aplicaciones informáticas
Recibido: 14/3/2012 | Aceptado: 20/4/2012

Aplicación informática para gestionar repositorios, establecer semejanzas y caracterizar ontologías

Aplicación informática para gestionar repositorios, establecer semejanzas y caracterizar ontologías

Edgar Rojas Ricardo^{1*}, Dayné Gutierrez González², Flavio Enrique Roche Rodríguez², Michael Eduardo Marrero Clark², Leonel Vila Pérez², Elio Luis Toledo²

¹ Centro de Tecnologías de Gestión de Datos, (DATEC). Departamento de PostgreSQL. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370. Correo-e: erojas@uci.cu

² Centro de Tecnologías de Gestión de Datos, (DATEC). Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370. Correo-e: {[erojas](mailto:erojas@uci.cu), [dgutierrezg](mailto:dgutierrezg@uci.cu), [feroce](mailto:feroce@uci.cu), [memarrero](mailto:memarrero@uci.cu), [lvila](mailto:lvila@uci.cu), [eltoledo](mailto:eltoledo@uci.cu)}@uci.cu

* Autor para correspondencia

Resumen: el objetivo de la investigación es desarrollar una aplicación informática que gestione repositorios de ontologías, caracterice y establezca semejanzas entre ontologías representativas del conocimiento en la web. Para ello se realizó un análisis y selección de las tecnologías y herramientas adecuadas para la gestión de ontologías y el desarrollo de la aplicación. Se identificaron y documentaron las funcionalidades de la aplicación y se realizó el diseño implementación y pruebas de la misma. Esta aplicación permitirá gestionar repositorios de ontologías, obtener características y semejanzas entre las ontologías, para facilitar los procesos de almacenamiento, búsqueda y recuperación de ontologías, así como la reutilización del conocimiento almacenado en ellas.

Palabras clave: Dominio, ontologías, repositorio, semejanza.

Abstract: *The research objective is to develop a software application to manage ontologies repositories, characterize and establish similarities between ontologies representing the knowledge on the web. That's why was performed an analysis a selection of appropriate technologies and tools for ontology management and application development. The application functionality were identified and documented and were made the design, implementation and testing of it. This application will manage the ontologies repositories, obtaining characteristics and similarities between representative ontologies of the knowledge on the web, making easier the processes of storage, search and retrieval of ontologies and reuse of knowledge stored in them.*

Keywords: *Domain, likeness, ontologies, repository.*

1. Introducción

Debido al crecimiento de la información y recursos disponibles en la web, esta se ha convertido en la forma principal de intercambio de datos, proporcionando un medio flexible para desarrollar el comercio, los negocios y otras

actividades. Paralelamente a este crecimiento se ha hecho necesaria la incorporación de nuevas tecnologías que incrementen el potencial de la web y la adecúen a las necesidades crecientes de sus usuarios. La búsqueda y recuperación de información es un elemento fundamental para cubrir dichas necesidades; a pesar de que son numerosos los mecanismos y aplicaciones que se han desarrollado con este fin, la web presenta una característica que hace que este proceso en la actualidad sea ineficiente: la información contenida en ella está estructurada para el entendimiento de las personas y no de ordenadores o sistemas de computo.

La web actual es un espacio preparado para el intercambio de información diseñado para el consumo humano. Con los estándares del momento, no se puede diferenciar entre información personal, académica, comercial, etc. Es decir, cuando un buscador web realiza una consulta con algunas palabras claves, normalmente la información que se genera como respuesta no es útil en su totalidad porque no corresponde a las necesidades de los usuarios. Por otro lado, los agentes de búsqueda actuales no se diseñan para “comprender” la información que reside en la web, porque es prácticamente imposible conocer la representación de los datos ubicados en las páginas.

A finales de los noventa surge la visión de lo que se ha denominado la web semántica diseñada por Tim Berners-Lee, la cual propone superar las limitaciones de la “web actual”.

Con una semántica implícita los datos contenidos en las páginas web pueden ser utilizados y “comprendidos” por los ordenadores. Esta tecnología busca desarrollar una web más cohesionada, donde sea aún más fácil localizar, compartir e integrar información y servicios (Ortega, 2009).

Las ontologías constituyen la base para soportar la representación del conocimiento que necesita la web semántica, en este ámbito pretenden introducir descripciones explícitas sobre el significado formal de los objetos y sus relaciones. Estas permiten almacenar el conocimiento, de modo que la información se encuentre ordenada y mejor procesada. De esta forma sistemas automatizados como los agentes de software podrán interpretar conceptos, comprender los datos almacenados en las páginas web, buscarlos e integrarlos de manera eficiente, procesarlos y sacar conclusiones de ellos; así las búsquedas y el manejo de datos en internet serán más específicos.

Actualmente existen múltiples ontologías desarrolladas y publicadas en internet que son construidas en diferentes partes del mundo, sin tener en cuenta que se puede reutilizar el conocimiento almacenado en otras existentes, debido a que se tratan conceptos que son iguales en terminologías diferentes, pues no existe un vocabulario común entre los especialistas de un mismo dominio en diferentes escenarios. Además es importante poder analizar características de una o varias ontologías que arrojen resultados válidos para una investigación en curso. Muchas de estas ontologías son agrupadas en repositorios estructurados de forma muy heterogénea y en algunos casos el acceso a ellos se dificulta por el uso excesivo de requisitos de autenticación. Además carecen de información sobre su contenido, que pudiera hacer más eficiente la búsqueda. Luego de analizar lo anteriormente descrito se definió como objetivo general desarrollar una aplicación informática que gestione repositorios de ontologías, caracterice y establezca semejanzas entre ontologías representativas del conocimiento en la web para facilitar los procesos de almacenamiento, búsqueda y recuperación de ontologías.

2. Materiales y métodos

En la rama de la Informática: “Una ontología define los términos a utilizar para describir y representar un área de conocimiento. Son utilizadas por las personas, las bases de datos, y las aplicaciones que necesitan compartir un

dominio de información. Incluyen definiciones de conceptos básicos del dominio, y las relaciones entre ellos, que son útiles para los ordenadores” (Hendler, 2005).

2.1. Componentes de una ontología

Las diferentes disciplinas que se encargan del estudio de las ontologías no identifican con exactitud los mismos elementos que la componen o presentan en ocasiones ambigüedad en los términos con que los designan. De manera general se ha llegado al consenso de que las ontologías están compuestas por cinco elementos que permitirán representar el conocimiento de algún dominio: conceptos, relaciones, funciones, instancias y axiomas.

Conceptos: son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc. Las clases en una ontología se suelen organizar en taxonomías a las que se les pueden aplicar los mecanismos de herencia.

Relaciones: representa la interacción y enlace entre los conceptos del dominio. Suelen formar la taxonomía del dominio.

Funciones: son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología.

Instancias: se utilizan para representar objetos determinados de un concepto.

Axiomas: son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Permiten junto al mecanismo de la herencia de conceptos, inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos (Gruber, 1993).

Cada componente de la ontología tiene una implicación en la gestión del conocimiento. Así, los conceptos, las instancias y las relaciones pueden representar el conocimiento tácito de los integrantes de la organización. Las funciones son muy usadas para describir los pasos para desarrollar un proceso. Y los axiomas, permiten hacer inferencias, lo que es de gran utilidad para la toma de decisiones (Ramírez, 2005).

Para la representación del conocimiento en la web mediante ontologías se han creado un grupo de lenguajes, fundamentalmente contruidos en sintaxis XML (del inglés, *Extensible Markup Language*) o RDF(S) (del inglés, *Resource Description Framework*). Dichos lenguajes, aún en continua evolución, poseen semánticas que permiten razonar sobre una determinada expresión por lo que ofrecen una integración e interoperabilidad de datos muy ricos entre comunidades descriptivas, siendo comprensibles por las máquinas, y cuya mayor virtud está en la estructuración de los contenidos en la web (Revista de Ciencias Sociales, 2007).

Entre la amplia gama de lenguajes que en la actualidad se han creado para la representación de ontologías, se seleccionó OWL (del inglés, *Web Ontology Language*) para ser gestionado por la aplicación. OWL es el lenguaje de especificación de ontologías más completo, principalmente porque está basado fundamentalmente en lógica descriptiva. Es desde febrero de 2004 una recomendación de W3C, por lo que la mayoría de desarrolladores e investigadores de la web semántica van a centrar sus esfuerzos en desarrollar herramientas y sistemas orientados a este lenguaje (Samper, 2005).

2.2. Herramientas y tecnologías

Para el desarrollo de la aplicación se seleccionó la metodología XP en busca de simplificar el desarrollo del software. Para la gestión de ontologías se decide utilizar Jena 2.6 teniendo en cuenta que incluye una API para el manejo de ontologías y soporta OWL (del inglés, *Web Ontology Language*), lenguaje en el que han sido representadas las ontologías que se procesan en la aplicación. Para la realización de la aplicación se utilizó como lenguaje de programación Java y al tratarse de una aplicación web se utilizó tecnología JSP (Java Server Page). Para mostrar las gráficas se utiliza la librería Highchart. Es una librería escrita en Javascript y ajax capaz de crear gráficos estadísticos interactivos. Para el desarrollo de la interfaz de la aplicación se decide utilizar Dojo 1.3, este proporciona facilidades para el diseño y utiliza tecnología AJAX, la que posibilita realizar cambios en secciones de las páginas sin que sea necesario la recarga constante de las mismas. Como Entorno de Desarrollo Integrado se utilizó Netbeans 6.9 hecho principalmente para el lenguaje de programación java pero que además integra lenguajes como JSP, javascript y CSS, permitiendo desarrollar aplicaciones web. Se utilizó además como servidor de aplicaciones Apache Tomcat 6.0 y como Sistema Gestor de Base de Datos PostgreSQL 8.4.

3. Resultados y discusión

En concordancia con lo que propone la metodología XP, para el desarrollo de la aplicación se identificaron y describieron Historias de Usuario (HU) donde se reflejaron las características que debía cumplir la aplicación. Se redactaron un total de 29 HU, de ellas 13 con un fuerte impacto en las funcionalidades de la aplicación por lo que se les designó un nivel de prioridad alto. Se agruparon las HU teniendo en cuenta la implicación funcional de una con otra y la complejidad de las mismas, quedando establecidas 5 iteraciones para el desarrollo del proyecto. La aplicación se diseñó para el trabajo con ontologías escritas en lenguaje OWL (del inglés, *Web Ontology Language*) por ser este el más popular y completo entre los existentes. Su funcionamiento se basa en la gestión de un repositorio local, aunque permite incorporar contenidos de repositorios remotos con el objetivo de brindar a los usuarios la mayor cantidad de información posible. Las ontologías contenidas en los repositorios, están clasificadas por dominios del conocimiento, lo que posibilitará caracterizar en cierta medida los repositorios al poder establecer los dominios a los que pertenece su contenido. Además es posible, mediante gráficos, establecer comparaciones entre los diferentes repositorios, elemento que facilitará en gran medida al usuario decidir en cual de estos realizar la búsqueda de la ontología que necesite. Para incorporar los dominios del conocimiento que luego permitirán caracterizar el contenido de los repositorios se estableció una gestión de dominios; dicha gestión la realizará un usuario gestor que previamente autenticado podrá insertar, modificar o eliminar un dominio del conocimiento.

Otra característica de la aplicación desarrollada que le confiere gran utilidad para el usuario es la posibilidad de visualizar la ontología seleccionada en forma de árbol y de grafo, así como brindar parámetros estadísticos como la media, la moda y la varianza de dicha ontología; elementos que pudieran ser útiles al usuario para el trabajo o investigación que este realizando y para el cual requiera el uso de dicha ontología. Además la aplicación permite establecer el grado de semejanza entre dos o más ontologías, elemento que cobra mayor importancia si tenemos en cuenta que en ocasiones son usadas terminologías distintas para describir un mismo fenómeno de la realidad; esta característica permitirá establecer si las ontologías describen un mismo fenómeno y el grado de semejanza que existe entre una u otra.

En la aplicación se realiza la gestión de usuario, puesto que, si bien la mayoría de las funcionalidades están disponibles a cualquier usuario (sin necesidad de autenticación), existen algunas que son específicas de determinado rol. Se establecieron cuatro roles (usuario común o sin autenticar, usuario autenticado, gestor y administrador).

Cada usuario, según el rol que ocupen en la aplicación podrá acceder a las funcionalidades propias y a las de los roles de menor prioridad o rango. La prioridad de los roles está establecida de forma descendente en el siguiente orden: administrador, gestor, usuario autenticado y usuario común.

La aplicación en general es una propuesta novedosa que posibilita de una forma fácil y usable la compartimentación y reutilización de la información entre la comunidad de desarrolladores e investigadores que participan en proyectos vinculados a la denominada web semántica. No es un simple repositorio con sólo las clásicas operaciones de insertar, actualizar o eliminar la información, sino que se le incorporan una serie de funcionalidades que la convierten en una herramienta útil para el desarrollador de la web del futuro.

Para la obtención de los parámetros estadísticos que permitirán caracterizar las ontologías se realizó un estudio de la teoría de conceptos de la estadística descriptiva.

La estadística descriptiva tiene por objeto fundamental describir y analizar las características de un conjunto de datos, obteniéndose de esa manera conclusiones sobre las características de dicho conjunto y sobre las relaciones existentes con otras poblaciones, a fin de compararlas (Canavos, 1998). La estadística en conjunto con la informática permite manejar de una manera eficiente, confiable y relativamente fácil grandes volúmenes de información y obtener resultados que serán sometidos al análisis e interpretación de los profesionales.

En el caso de las ontologías, se analizan un conjunto de datos que son representados a través de gráficos de pastel y de barra. Estos gráficos muestran los tipos de propiedades (*ObjectProperty*, *DatatypeProperty*) presentes en una ontología y la cantidad de cada tipo. Además de la cantidad de clases con instancias y sin instancias.

Otra de las características que se analizan de una ontología se obtiene a partir de los cálculos estadísticos que se explican a continuación. Donde se toman como referencia parámetros ya definidos dentro de la estadística descriptiva y son interpretados en función de ofrecer un resultado referente a las ontologías.

Media o Media aritmética: es un promedio estándar que a menudo se denomina promedio.

En una ontología la Media expresará el promedio de instancias por clases y se calcula de la siguiente manera:

$$M = \frac{\sum_{i=1}^n X_i}{n}$$

M: representa la Media (Promedio de instancias por clases).

N: representa el tamaño de la población (Cantidad de clases).

Xi: representa cada uno de los valores de la población (Cantidad de instancias de cada clase).

Moda: Es el valor que más veces se repite dentro del conjunto de datos.

En una ontología la Moda indica el tipo de propiedad que más se observa en el conjunto de clases que la componen.

Varianza: Esta medida permite identificar la diferencia promedio que hay entre cada uno de los valores respecto a su punto central (Media).

En una ontología la Varianza indica la diferencia promedio que hay entre cantidad de atributos por clases.

$$Vr = \frac{\sum_{i=1}^n (Xi - M)^2}{2a}$$

M: representa la Media (Promedio de instancias por clases).

N: representa el tamaño de la población (Cantidad de clases).

Xi: representa cada uno de los valores de la población (Cantidad de instancias de cada clase).

Vr: representa la Varianza (Cuanto se dispersan la cantidad de instancias por clases alrededor de su promedio de instancias).

Para el desarrollo de la funcionalidad obtener semejanza entre ontologías fue necesario establecer algunas métricas.

”Las métricas son medidas que se le aplica a un producto, estas medidas proporcionan una indicación cuantitativa de extensión, cantidad, dimensiones, capacidad y tamaño de algunos atributos de un proceso producto” (Pressman).

La IEEE siglas en ingles de *The Institute of Electrical and Electronics Engineers* (Instituto de ingenieros eléctricos y electrónicos) en “*Standard Glossary of Software EngeringTerms*” (Glosario Estándar de Términos de Ingeniería de Software) define métrica como una medida cuantitativa del grado en que un sistema, componente o proceso posee un atributo dado (IEEE,1992).

Atendiendo a los conceptos antes mencionados se concluye que una métrica es una forma de evaluar y una escala, definidas para realizar mediciones de uno o varios atributos. Para la investigación se propone una métrica que permita dado una ontología definir cuantas existen semejantes a ella en un repositorio determinado, basándonos en el esquema Noy y Hafner, proponen un marco para comparar ontologías con el propósito de estudiar la diversidad de los diseños de ontologías e indicar las fortalezas y debilidades de cada uno. Las características las clasifican como se muestra en la Figura 1.

Las características seleccionadas para conformar la métrica definida para el desarrollo de la aplicación informática fueron:

- Su tamaño, identificado por el número de conceptos, propiedades, instancias y propiedades de las instancias.
- Los términos representados: qué conceptos incorpora la ontología.
- Si los conceptos tienen estructura interna (si tiene un conjunto de propiedades que le dan significado a cada categoría).
- Los tipos de atributos que están presentes en un concepto.

Según la métrica confeccionada dos ontologías se pueden definir como semejante atendiendo a un porcentaje determinado, de acuerdo al número de características que contengan similitud entre ellas. El porcentaje de semejanza

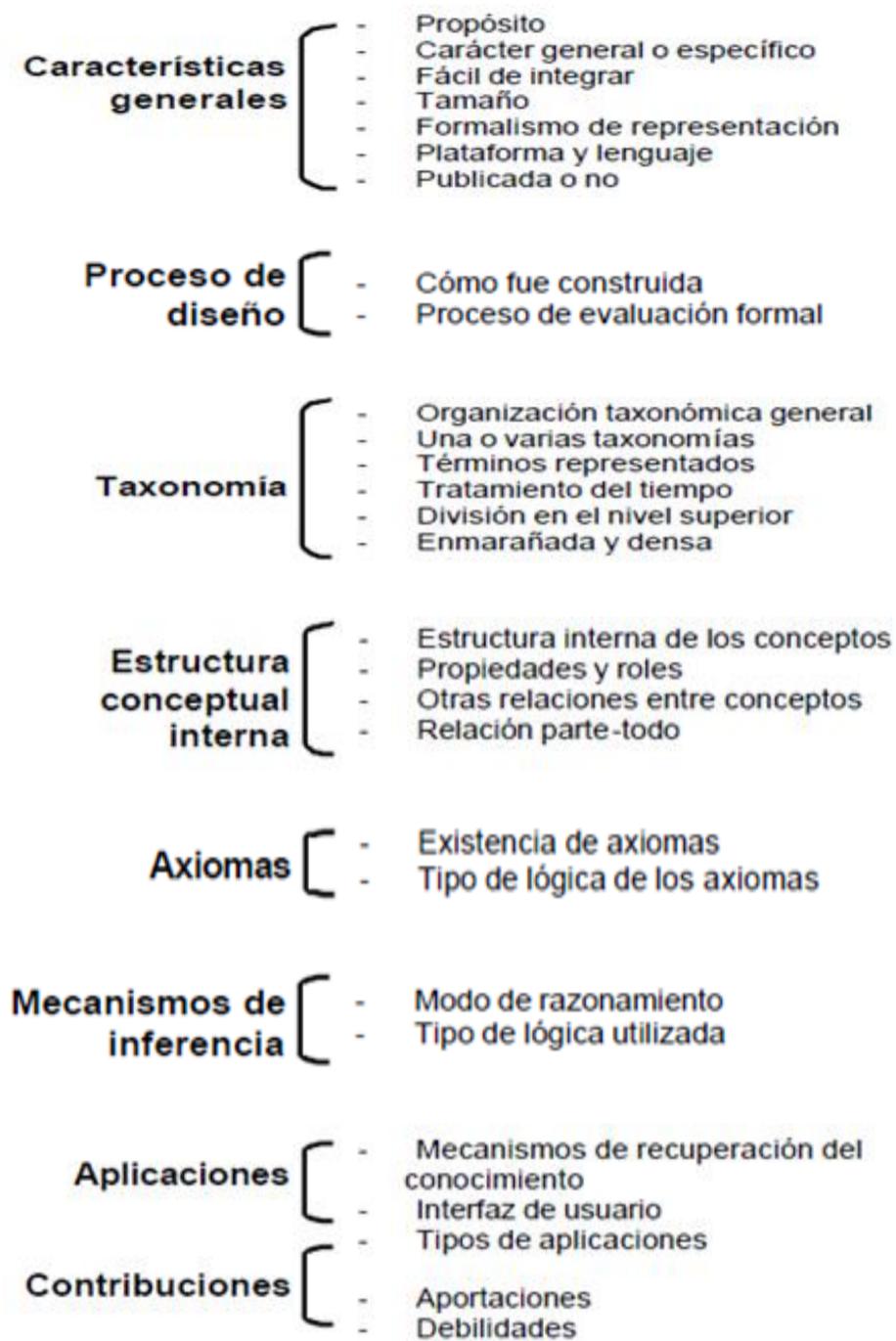


Figura 1. Taxonomía de características de Noy y Hafner.

será igual al número que represente la similitud de las características entre el total de características definidas en la métrica por cien, quedando la fórmula representada de esta forma:

$$Ps = \frac{Ns}{Tc} * 100$$

Donde:

Ps = porcentaje de semejanza.

Ns = número que represente la similitud de las características.

Tc = total de características definidas

De manera general la aplicación cuenta con un grupo de funcionalidades que están diseñadas para minimizar los tiempos de búsqueda, análisis y recuperación de ontologías, así como para permitir la reutilización del conocimiento, con ello se preservaría el conocimiento de los expertos en cualquier dominio.

Existen numerosas ontologías publicadas en internet pero los repositorios carecen de información acerca de su contenido y se hace más difícil encontrar la información deseada. La aplicación ofrece una estructura de repositorios organizados por dominios del conocimiento que facilitan la búsqueda de ontologías.

Los valores estadísticos permiten obtener información de los componentes de una ontología, ahorrando tiempo y recurso a los usuarios que deseen analizar una o varias ontologías para llevar a cabo un proyecto.

Obtener semejanza entre ontologías atendiendo a su estructura interna y los conceptos que en ella se representan, permite la reutilización del conocimiento, teniendo en cuenta que en ocasiones son tratados iguales conceptos con terminologías diferentes.

4. Conclusiones

Las ontologías y los lenguajes para su representación constituyen la base para el desarrollo de la web del futuro o web semántica. Serán los mecanismos que posibilitarán dotar a la información contenida en Internet de una estructura que los sistemas informáticos podrán interpretar. El desarrollo de aplicaciones para apoyar este proceso constituye un elemento fundamental para una rápida implantación de estas tecnologías. Con el propósito de contribuir a este proceso se implementó una aplicación informática para la gestión de repositorios de ontologías representativas del conocimiento en la web, que además, las caracteriza y establece semejanzas entre ellas a partir de métricas y parámetros estadísticos; con el objetivo de hacer más viable el proceso de gestión y reutilización de ontologías en el desarrollo de aplicaciones para la web semántica.

5. Referencias

CANAVOS, GEORGE C. *Probabilidad y Estadística*. Caracas. s.l. : McGraw-Hill, 1998.

GRUBER, THOMAS.TOWARD. *Principles for the Design of Ontologies Use for Knowledge Sharing*. Stanford Knowledge Systems Laboratory: s.n., 1993.

HENDLER, JIM. *Preguntas frecuentes sobre el Lenguaje de Ontologías Web (OWL) del W3C*. [en línea] 2005. [Consultado el: 15 de octubre de 2010]. Disponible en: [<http://www.w3c.es/Traducciones/es/SW/2005/owlfaq>].

IEEE, *IEEE Standard for Software Quality Management System*. 1992.

ORTEGA MORÁN, JUAN FRANCISCO. *Cálculo de modos y tiempos de desplazamiento en una ciudad usando fuentes públicas*. [en línea] 2009. [Consultado el: 25 de septiembre de 2010]. Disponible en: [<http://e-archivo.uc3m.es/bitstream/10016/5837/1/PFC%20Juan%20Francisco%20Ortega%20Moran.pdf>].

PRESSMAN, ROGER S. *Ingeniería de Software: Un enfoque práctico*. 5ta. Edición. s.l.: McGraw-Hill, 2002. ISBN 84-481-3214-9.

RAMÍREZ CÉSPEDES, ZULIA. *Las ontologías como herramienta en la Gestión del Conocimiento*. Departamento de Bibliotecología y Ciencia de la Información, Universidad de La Habana: s.n., 2005.

SAMPER ZAPATER, JOSÉ JAVIER. *Ontologías para servicios web semánticos de información de tráfico: descripción y herramientas de explotación*. Valencia: s.n., 2005. 84-370-6270-5.

TALAVERA PEREIRA, ROSALBA y MARCANO AULAR, YELITZA JOSEFINA. *Los lenguajes de representación semántica y su uso en la construcción de ontologías*. Revista de Ciencias Sociales. 1, 2007, Vol. XIII. 1315-9518.