

# Base de Datos para la Gestión de Ontologías

**Belén B. Bonilla M., Davis Arosemena, estudiantes**

**Víctor López C., MSc.**

Universidad Tecnológica de Panamá

belen.bonilla@utp.ac.pa, davis.rosemena@utp.ac.pa, victor.lopez@utp.ac.pa

**Resumen-** Una ontología es una representación formal del conocimiento donde los conceptos, las relaciones y las restricciones conceptuales son definidos mediante formalismos en un determinado dominio. Las bases de datos utilizadas para almacenar y hacer consultas sobre ontologías se denominan Base de Datos basadas en Ontologías. En este artículo se presenta un estudio sobre las Bases de Datos basadas en Ontologías y se describen las características y propiedades de algunas de estas bases de datos.

**Palabras Claves-** Ontología, Base de Datos, Sesame, OntoDB, RDFSuite.

## 1. Introducción

Actualmente, las ontologías son utilizadas en una gran cantidad de dominios. Por ejemplo, en la Web Semántica [1], se utilizan para describir términos, recuperar información e interconectar servicios web, y en el área de base de datos, para facilitar el intercambio e integración de los datos [2].

Debido al uso incremental de las ontologías, se han propuesto una gran cantidad de modelos y lenguajes para el manejo de tales ontologías, dentro de los cuales se pueden mencionar: Resource Description Framework (RDF), Resource Description Framework Schema (RDFS) [3], Ontology Web Language (OWL) [4], entre otros. Sin embargo, resulta necesario también, poder almacenar estas ontologías y sus instancias de tal manera que se puedan acceder y manipular fácilmente.

Varios trabajos han sido desarrollados para representar tanto las ontologías como la data asociada, en bases de datos. Una base de datos que ofrece esta funcionalidad es denominada Base de Datos basada en Ontologías (OBDB – Ontology-based Database) o base de datos ontológica.

Existen diferentes bases de datos para el almacenamiento y gestión de ontologías entre las que se pueden mencionar Sesame [5], OntoDB [6], entre otras. Cada una posee características, propiedades y ventajas, y la elección de una u otra depende de las necesidades del usuario y del lenguaje de definición de las ontologías.

El objetivo de este artículo es presentar un estudio sobre las bases de datos utilizadas para el almacenamiento y consulta de ontologías, su importancia y características.

En la siguiente sección se presenta el concepto de ontologías, los beneficios de su utilización, sus aplicaciones y los lenguajes utilizados para crearlas y definir las. En la sección 3 se presenta el concepto de base de datos basadas en ontologías o base de datos ontológicas, la necesidad de su utilización y diferentes bases de datos para la gestión de ontologías.

## 2. Ontologías

Las ontologías proceden del campo de la Inteligencia Artificial; son vocabularios comunes para las personas y aplicaciones que trabajan en un dominio específico [7].

Uschold y Gruninger [8], utilizan el término ontología para referirse a un entendimiento compartido de algún dominio de interés, el cual puede ser utilizado como un framework unificado para resolver problemas tales como interoperabilidad, confusión conceptual y terminológica, etc. Por otro lado, Gruber [9], definen ontología como "una especificación explícita y formal de una conceptualización compartida". El término "conceptualización" se refiere a un modelo abstracto de algún fenómeno del mundo construido mediante la identificación de los conceptos relevantes a ese fenómeno; el término "explícito" significa que los conceptos utilizados en la ontología, y las restricciones para su uso, están claramente definidos; el término "formal" se refiere al hecho de que debe ser comprensible para las máquinas, es decir, estar expresada mediante una sintaxis que permita a una computadora operar sobre ella; y finalmente, el término "compartida" refleja la noción de que contendrá conocimiento consensuado en algún grado [10].

### 2.1. Beneficios de las ontologías

Actualmente, las ontologías facilitan aspectos como comunicación, interoperabilidad y razonamiento automático.

**Comunicación.** Las ontologías reducen las confusiones de tipo conceptual y terminológico; de esta forma, habilitan el entendimiento compartido y la comunicación entre personas, organizaciones y aplicaciones con diferentes necesidades y puntos de vista sobre un contexto particular. Esto se logra ya que a través de las ontologías se pueden crear modelos normativos los cuales permiten agregar semántica a un sistema determinado y crear un modelo extensible que puede ser refinado posteriormente y que permite transformaciones semánticas entre diferentes contextos [8].

**Interoperabilidad.** La interoperabilidad es la condición mediante la cual sistemas heterogéneos pueden intercambiar procesos o datos. Para el tratamiento de la interoperabilidad, las ontologías se utilizan para soportar traducciones entre diferentes lenguajes y representaciones. En términos más prácticos, una ontología puede funcionar como traductora permitiendo la interoperabilidad entre dos aplicaciones de software [8].

**Razonamiento Automático.** Las ontologías resultan muy útiles para facilitar el razonamiento automático, es decir, sin necesidad de la intervención humana. Partiendo de unas reglas de inferencia, un motor de razonamiento puede usar los datos de las ontologías para inferir conclusiones de los mismos.

Una de las aplicaciones más importante del razonamiento automático es la validación de datos. También se usa para establecer relaciones entre ontologías, para descubrir relaciones ocultas o inesperadas entre los datos y para integrar esquemas de bases de datos [11].

Por otro lado, el uso de ontologías, aunado a sus bondades en cuanto a comunicación, interoperabilidad y razonamiento automático, permite que los usuarios organicen y estructuren la información de

manera que los agentes de software puedan interpretar el significado y, por tanto, puedan buscar e integrar datos mucho mejor que como se hace actualmente. Gracias al conocimiento almacenado en las ontologías, las aplicaciones podrán extraer automáticamente datos de las páginas web, procesarlos y sacar conclusiones de ellos, así como tomar decisiones y negociar con otros agentes o personas [12, 13].

## 2.2. Aplicaciones de las ontologías

Actualmente, las ontologías son ampliamente utilizadas en actividades como: comercio electrónico [14], gestión de conocimiento [15], procesamiento de lenguaje natural [16], recuperación de información [17] y Web Semántica [18], siendo estas dos últimas las de mayor impacto y relevancia.

Por otro lado, las ontologías también son utilizadas en las bases de datos, facilitando el intercambio e integración de los datos almacenados [19].

## 2.3. Lenguajes

Para crear y manipular ontologías, resulta necesario contar con lenguajes y tecnologías que provean las características adecuadas para su tratamiento y definición, según las necesidades de aplicación [20].

A continuación se detallan los lenguajes más utilizados para crear y definir ontologías.

**RDF/RDFS**, han sido desarrolladas por el World Wide Web Consortium (W3C) y en conjunto constituyen una herramienta de representación del conocimiento de uso general que proporciona un método neutral para describir un recurso o definir una ontología [3].

Una ontología puede ser definida a través de un RDFS y las instancias de esta ontología, a través de un documento RDF.

**DAML+OIL**, fue desarrollado por un comité conjunto de Europa y Estados Unidos en el contexto del proyecto DARPA, para permitir interoperabilidad semántica en XML. Es una extensión de RDF así como RDF es un lenguaje basado en XML. Su objetivo es apoyar la transformación de la Web de ser un foro de presentación de la información a ser un recurso de interoperabilidad, comprensión y razonamiento [21].

**DAML+OIL** coloca énfasis en la semántica clara e introduce varios constructores de lenguajes que incrementan el poder de expresividad en las ontologías.

**OWL** es un lenguaje de marcado para publicar y compartir datos utilizando ontologías. Tiene como objetivo proporcionar un lenguaje de marcado construido sobre RDF (Resource Description Framework) y XML [4].

OWL es un lenguaje para la definición e instanciación de ontologías y puede ser usado para representar el significado de los términos en vocabularios y la relación entre esos términos. Se podría describir como un conjunto de reglas que ofrecen una forma de definir conceptos y hechos del mundo real.

## 3. Bases de Datos basadas en Ontologías

Con la Web Semántica y todas las funcionalidades de la Web 3.0 relacionado con la capacidad de inferir conocimiento a través de los metadatos de los datos; se presentan nuevos retos para la gestión de dicha información en la Web.

Como solución se propone la utilización de ontologías que facilita la estructuración de la información con su respectiva semántica; de esta manera, las ontologías no solo son utilizadas para funciones en la Web sino para cualquier otra función en la que se requiera significado de lo que se está procesando en la computadora.

Muchas ontologías representadas en un lenguaje como OWL conjuntamente con sus instancias, representa una dificultad a la hora de gestionar sus operaciones ya que al encontrarse en un texto plano requiere la utilización de herramientas como Protégé [22] que levante toda la información en memoria principal para la realización de dichas operaciones; sin embargo, aunque estas herramientas son fáciles de diseñar están limitadas por la disponibilidad de memoria principal. Basado en esto se propone la utilización de bases de datos que facilite la gestión de las ontologías ya que por las características de un DBMS en la que la distribución, flexibilidad, heterogeneidad representan una gran ventaja, se garantiza la disponibilidad e integridad de la información [23].

Surge entonces, el concepto de Base de Datos basada en Ontologías (Ontology-based Database OBDB). Un OBDB es el almacenamiento de un modelo ontológico y sus instancias en una base de datos para la gestión de la información que se encuentra semánticamente estructurada [23].

La estructuración de los datos comprende dos modelos: orientado a objeto y basado en entidad-relación. Las ontologías utilizan el modelo orientado a objeto, ya que se estructura en clases, subclases y propiedades. De esta manera, un OBDB es una base de datos orientada a objetos virtual, sobre la base de datos relacional. Esto lleva a la necesidad de realizar un mapeo en la que una base de datos entidad-relación almacene información estructurada de forma orientada a objetos. Existen muchos modelos o esquema de mapeo para el mismo, tales como: la representación vertical y la representación binaria [24], en la que se combina representaciones verticales y horizontales; por otro lado, existen arquitecturas que emplean estos esquemas a través de un OBDB.

La consulta de la información almacenada en los OBDB puede ser realizada a través de lenguajes como SPARQL, RQL, OWL-QL, OntoQL y otros.

A continuación se describen algunos OBDBs que pueden ser utilizados.

### 3.1. RDF Suite

RDF Suite es una suite de herramientas para la validación, almacenamiento y consulta de ontologías y sus instancias que son descritas en RDF/RDFs [25]. De esta manera, los datos descritos en RDFs y las instancias de los datos en RDF serán representados en una base de datos relacional. Los datos en RDFs son almacenados en 4 tablas: clases, propiedades, subclases y subpropiedades. Para almacenar la información de los namespaces se utilizará otra tabla y las instancias de datos en RDF son guardadas en tablas separadas. Los *domain* y *range* que especifican una propiedad en las ontologías no son soportados por RDF Suite por lo que presenta poca flexibilidad y menos poder de expresión que otros lenguajes de ontologías como OWL. Por otro lado, RDF Suite utiliza el lenguaje de consulta RQL.

## 3.2. Sesame

Es una arquitectura genérica para el almacenamiento y consulta eficiente de una gran cantidad de metadatos estructurados en RDF/RDFs. El diseño e implementación de Sesame es independiente del modelo de estructura de datos, por lo que puede soportar base de datos relacional, *triple stores* y base de datos orientado a objeto [26]. Esto se debe a que presenta una API llamada SAIL (Storage and Inference Layer) el cual se encuentra en la capa superior del DBMS. De esta manera, el traslado de un método específico de RDF a cualquier método para el modelo de estructura de datos es responsabilidad del SAIL.

Esta arquitectura es diseñada para ser flexible y extensible en la utilización de más DBMS si fuese requerido. El lenguaje de consulta soportado es RQL. Sesame puede soportar el almacenamiento de *domain* y *range* dentro de las tablas en contraste con RDF Suite; sin embargo, ambos poseen esquema de base de datos similares cuando utiliza el modelo de estructura de datos objeto relacional. Los datos de RDF son guardados en tablas separadas.

Sesame, al utilizar RDF/RDFs, carece de poder de expresión de modelos ontológicos en comparación a otros lenguajes como OWL [26].

## 3.3 DLDB

Se define como una base de datos extensible relacional que soporta consultas de la semántica de la información utilizada en la Web semántica [27]. El diseño de DLDB se basa en la utilización de ontologías escritas en DAML+OIL que pueden ser almacenadas de manera híbrida, donde se combina la parte vertical (una simple tabla con un predicado, sujeto y objeto para cada columna) y la horizontal (una simple tabla con una columna para cada tipo y propiedad). Para esta arquitectura cada clase de la ontología es mapeada a una tabla independiente. Por consiguiente, cuando se guarda una gran cantidad de datos ontológicos, ésta podría representar un problema debido a la cantidad de tablas necesarias para la representación.

DLDB posee su propia implementación para las consultas a través de un conjunto de clases en Java; por lo que no necesita la utilización de ningún lenguaje específico. Estas clases mantienen el nombre de predicado y el número de términos, una vez que se ha ejecutado, se obtiene la traducción o el equivalente a las tablas y campos específicos que serán luego traducidos a consultas en lenguaje SQL; de esta manera, se ejecuta directamente en el DBMS la consulta generada en SQL del nombre del predicado y términos solicitados.

Esta arquitectura posee mayor capacidad de expresión que las implementaciones anteriores debido al lenguaje ontológico utilizado; sin embargo, es menos flexible ya que no acepta ningún lenguaje de consulta, creando una dependencia a la utilización del conjunto de clases.

## 3.4 OntoDB

OntoDB representa un modelo de OBDB relacional en la cual las ontologías y sus instancias son almacenadas a través de una estructura propia [24].

Este modelo propone dos estructuras para el almacenamiento: el enfoque de una sola tabla y el de doble esquema.

El primero utiliza tablas verticales para guardar la descripción de las clases, propiedades y sus instancias; en el segundo, el identificador de instancias representadas, propiedades y valores de las instancias son almacenadas en una tabla que posee tres columnas para cada valor mencionado.

El lenguaje ontológico utilizado es OWL por lo que posee mayor capacidad de expresión en la representación de la semántica y su lenguaje de consulta es el OntoQL que ha sido diseñado especialmente para este modelo.

Basado en OntoDB se propone OntoDB2 [28] el cual posee algunas mejoras como: permitir trabajar con varios modelos ontológicos en una misma estructura de base de datos.

## 3.5 OntoMS

OntoMS es un modelo para la representación de ontologías e instancias OWL en base de datos relacional [29]. Como lenguaje de consulta utiliza el OWL-QL, el cual está basado en patrones de consultas en forma de propiedades, sujeto y objeto.

OntoMS genera una clase basada en esquema relacional donde una relación es creada para cada clase. Cada clase relacionada contiene asociado tanto propiedades como atributos.

## Referencias

- [1] E. K. Jacob, "Ontologies and the Semantic Web," Bulletin of the American Society for Information Science and Technology, pp. 19-22, 2003.
- [2] T. Sellis, D. Skoutas, and K. Staikos, "Database interoperability through Web Services and ontologies," 8th IEEE International Conference on Bioinformatics and BioEngineering, 2008.
- [3] W3C. (2004) Resource Description Framework (RDF). [Online]. Disponible en: <http://www.w3.org/RDF/>
- [4] Z. zhihong and Z. mingtian, "Web Ontology Language OWL and its description logic foundation," Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies PDCAT'2003., pp. 157-160, 2003.
- [5] OpenRDF. (2002) User Guide for Sesame 2.3. [Online]. <http://www.openrdf.org/doc/sesame2/2.3.2/users/userguide.html>
- [6] S. Jean, H. Dehainsala, D. N. Xuan, G. Pierra, L. Bellatreche and Y. Aït-Ameur, "OntoDB: It Is Time to Embed Your Domain Ontology in Your Database," Advances in Databases: Concepts, Systems and Applications Lecture Notes in Computer Science, vol. 4443/2010, pp. 1119-1122, 2010
- [7] W3C. (2005) Guía Breve de Web Semántica. [Online]. Disponible en: <http://www.w3c.es/divulgacion/guiasbreves/websemantica>
- [8] M. Uschold, and M. Gruninger, "Ontologies: Principles, Methods and Applications", Knowledge Engineering Review, vol.11, 1996.
- [9] T.R. Gruber, "A translation approach to portable ontologies", Knowledge acquisition, vol.5, 1993, pp.199-220.
- [10] A. Gómez-Pérez, and D. Manzano-Macho, "An overview of methods and tools for ontology learning from text", The knowledge engineering review, vol.19, pp. 187-212, 2005.
- [11] M. Viinikkala. (2004) Ontology in Information Systems. [Online]. Disponible en: <http://www.cs.tut.fi/~kk/webstuff/Ontology.pdf>
- [12] R. Pedraza-Jiménez, L. Codina, and C. Rovira, "Web semántica y ontologías en el procesamiento de la información documental", El profesional de la información, vol.16, pp. 569-578, 2007.
- [13] E. P. Redondo, Y. H. Montero, E. Herrera-Viedma, and J.C. Herrera, "Ontologías, metadatos y agentes: recuperación semántica de la información", 2002.
- [14] A. Malucelli, "Combining ontologies and agents to help in solving the heterogeneity problem in e-commerce negotiations," Proceedings of International Workshop on Data Engineering Issues in E-Commerce, pp. 26-35, 2005.
- [15] W. Yanmei, "Ontology-based Web knowledge management," Proceedings of the 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing and the Fourth Pacific Rim Conference on Multimedia vol. 3, pp. 1859-1863, 2003.

- [16] I. Zafraayes, L. Gu, F. Giunchiglia, W. Pan, Q. Ji, M. Chi and X. Huang, "From Web Developer to Collaborative Natural Language Processing Challenges," *The Semantic Web Lecture Notes in Computer Science*, vol. 4625, pp. 693-696, 2007.
- [17] M. Eghami, "Associational agents and ontology for semantic information search," *Third International Conference on Research Challenges in Information Science*, pp. 421-425, 2008.
- [18] W3C. (2008) W3C Semantic Web Activity. [Online]. Disponible en: <http://www.w3.org/2007/sem/>
- [19] T. Galla, D. Stizman, and K. Stalova, "Database interoperability through Web Services and ontologies," *5th IEEE International Conference on Distributed and Bio-Inspired Computing*, 2009.
- [20] A. Gomez-Perez and D. Corcho, "Ontology languages for the Semantic Web," *Intelligent Systems IEEE*, vol. 17, pp. 84-90, 2002.
- [21] D. L. McGuinness, "DAML+OIL: an ontology language for the Semantic Web," *Intelligent Systems IEEE*, vol. 17, pp. 72-80, 2002.
- [22] Stanford Center for Biomedical Informatics Research. (2010) The Protégé Ontology Editor and Knowledge Acquisition System. [Online]. Disponible en: <http://protege.stanford.edu/>.
- [23] H. Dehnavi, "Benchmarking data Schemes of Ontology based Database," in *Proceeding of database system for advanced applications*, 2007.
- [24] H. Dehnavi, "DaidDB: An Ontology-Based Database for Data Intensive Applications," in *Proceeding of database system for advanced applications*, 2007.
- [25] S. Alcock, "The RDFSuite: Managing Voluminous RDF Description Bases," in *2nd International Workshop on the Semantic Web*, pp. 1-13, 2001.
- [26] J. Simola, "Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schemas", in *Proceedings of the first Intl Semantic Web Conference (ISWC 2002)*, vol. 2942, pp. 84-96, 2002.
- [27] Z. Pan and J. Hertz, "QLDB: Extending Relational Databases to Support Semantic Web Queries," 2003.
- [28] C. Fensler, "OntoDB2: Support of Multiple Ontology Models within Ontology Based Database," *Proceedings of the 2005 EDBT Ph.D. Workshop*, 2005.
- [29] M. J. Peris, "An Efficient and Scalable Management of Ontology," *Proceedings of the 12th International conference on Database systems for advanced applications*, 2007.

**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ**  
Sistema de Ingreso Universitario

**Fechas de Inscripción:**

- **1ra. Convocatoria**  
**9 de abril al 13 de julio de 2012**  
Aplicación de la prueba PAA: 21 de julio de 2012
- **2da. Convocatoria**  
**23 de julio al 14 de septiembre**  
Aplicación de la prueba PAA: 22 de septiembre de 2012

**o comunicarse a través de:**

● Bocas del Toro	706-8104/810-2306	<a href="http://www.fch.bocastorobocas.com">www.fch.bocastorobocas.com</a>
● Azuero	902-9821/810-2404	<a href="http://www.fca.azuero.com">www.fca.azuero.com</a>
● Chiriquí	774-1000/774-1000/774-1000	<a href="http://www.fcc.chiriqui.com">www.fcc.chiriqui.com</a>
● Coclé	897-9123/897-7323	<a href="http://www.fcc.cocle.com">www.fcc.cocle.com</a>
● Colón	472-4125/472-4125/472-4126	<a href="http://www.fcc.colon.com">www.fcc.colon.com</a>
● Veraguas	300-4047	<a href="http://www.fcv.veraguas.com">www.fcv.veraguas.com</a>
● Panamá Oeste	304-6807/304-6808	<a href="http://www.fcp.panamaoeste.com">www.fcp.panamaoeste.com</a>
● Panamá Sur	300-9271/300-3209	<a href="http://www.fcs.panamasur.com">www.fcs.panamasur.com</a>
● Escuela de Telecía	300-8137/300-9234	<a href="http://www.fct.com">www.fct.com</a>

Para mayor información ingrese a  
**[www.utp.ac.pa](http://www.utp.ac.pa)**