



TLATEMOANI
Revista Académica de Investigación
Editada por Eumed.net
No. 17 – Diciembre 2014
España
ISSN: 19899300
revista.tlatemoani@uaslp.mx

Fecha de recepción: 12 de noviembre de 2014
Fecha de aceptación: 17 de diciembre de 2014

ONTOLOGÍAS. INTEGRACIÓN DE ESQUEMAS

Dayana de la Caridad Rivero Hernández¹.
Profesora Asistente Universidad de Las Tunas. Cuba.
E-mail: dayanarh@ult.edu.cu

Javier Vila Labrada².
Profesor Asistente Universidad de Las Tunas. Cuba.
E-mail: javiervl@ult.edu.cu

Yadira de las Mercedes Rivero Hernández³.
Profesora Asistente Universidad de Ciencias Médicas. Las Tunas. Cuba.
E-mail: yadirar@ltu.sld.cu

Resumen. Una ontología define los términos y las relaciones básicas para la comprensión de un área del conocimiento, así como las reglas para poder combinar los términos para definir las extensiones de este tipo de vocabulario controlado. En este trabajo se realiza un estudio de las principales características de las ontologías, así como sus posibles aplicaciones y los conceptos fundamentales del tema. Además se analiza algunos de los esquemas de integración de datos existentes, así como la integración entre ontologías y folksonomías.

Palabras clave: Ontología, Integración de Datos, Folksonomías, Integración de Esquemas, Integración de Ontología y Folksonomías.

Abstract. An ontology defines the basic terms and compression of a field of knowledge relationships and the rules to combine terms to define extensions of this

¹ Ingeniera en Telecomunicaciones y Electrónica. Máster en Informática Aplicada. Cursa Doctorado en Ciencias de la Computación.

² Ingeniero en Automática. Cursa Maestría en Informática Aplicada.

³ Licenciada en Educación, Especialidad Matemática. Máster en Didáctica de la Educación Superior.

type of controlled vocabulary. In this paper a study of the main characteristics of ontologies and their potential applications and fundamental concepts of the subject is made. Also some of the integration schemes existing data, as well as integration between ontologies and folksonomies is discussed.

Keywords: Ontology, Data Integration, Folksonomies, Schema Integration, Integration of Ontology and Folksonomies.

Introducción

El rápido crecimiento de Internet ha simplificado el acceso a los datos, lo que ha provocado un incremento en el número de fuentes de datos creadas. A pesar de este incremento, en la mayoría de los casos estos repositorios de datos son accedidos de forma manual. A este problema se une el hecho de que estos datos tienen representaciones heterogéneas y la extremada volatilidad de la Web, en la que los recursos aparecen con la misma velocidad con que se crean. La heterogeneidad mencionada puede ser de tres tipos: intencional (diferencias en los contenidos), semántica (diferencias en la interpretación) y esquemática (tipos de datos, etiquetado, estructuras, etc.).

Debido al incremento de información disponible y la complejidad de tratar con grandes cantidades de datos heterogéneos se ha producido un incremento en la investigación sobre la integración de datos. La comunidad de bases de datos es una de las más importantes y con experiencia contrastada en el estudio de soluciones a los problemas de heterogeneidad y dispersión de los datos. Sin embargo otras comunidades como recuperación de la información y representación del conocimiento han aportado soluciones a los problemas mencionados.

El uso de ontologías para la integración de datos tiene algunas ventajas sobre los sistemas basados en palabras clave. Las ontologías proveen un vocabulario compartido común (conceptos) para representar la información incluida en los documentos (contenidos). Además las ontologías permiten definir relaciones entre los conceptos (roles). Tanto conceptos como roles pueden ser usados para realizar consultas más complejas y recuperar de forma precisa la información en la que el usuario está interesado.

Actualmente la integración basada en ontologías es un área de investigación muy activa, que recibe diferentes nombres dependiendo del objetivo del sistema propuesto: mediación semántica, mediación conceptual, integración semántica de datos, etc. Algunas propuestas para el uso de ontologías en sistemas de integración son: Carnot [1], SIMS [2], OntoMerge [3], OIS Framework [4] y Observer [5].

Materiales y Métodos

1. Ontologías

Una ontología define los términos y las relaciones básicas para la comprensión de un área del conocimiento, así como las reglas para poder combinar los términos para definir las extensiones de este tipo de vocabulario controlado.

Se trata de convertir la información en conocimiento, mediante unas estructuras de conocimiento formalizadas (las ontologías) que referencien los datos, por medio de metadatos, bajo un esquema común normalizado sobre algún dominio del conocimiento. Los metadatos no sólo especificarán el esquema de datos que debe aparecer en cada instancia, sino que también podrán contener información adicional de cómo hacer deducciones sobre ellos, es decir, cómo establecer axiomas que podrán, a su vez, aplicarse en los diferentes dominios que trate el conocimiento almacenado.

De esta forma, los buscadores podrán obtener información al compartir los mismos esquemas de anotaciones web y los agentes de software no sólo encontrarán la información precisa, sino que podrán realizar inferencias de forma automática buscando información relacionada con la que se encuentra situada en las páginas web y con los requerimientos de las consultas realizadas por los usuarios. Además, los productores de páginas y servicios web podrán intercambiar sus datos siguiendo estos esquemas comunes consensuados e, incluso, podrán reutilizarlos.

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

El término ontología se ha empleado desde hace muchos siglos en el campo de la filosofía y del conocimiento y hace ya varias décadas cobró especial relevancia en el campo de la biblioteconomía y la documentación. Hoy ha sufrido un nuevo impulso debido al desarrollo de la Web Semántica donde prima la idea de transformar la red no sólo en un espacio de información, sino también en un espacio de conocimiento. [6]

En el campo de la filosofía, la ontología -en singular- se ha considerado una rama de la filosofía que se ocupa de la naturaleza y organización de la realidad. En los años 90, dicho concepto se empezó a utilizar en el campo de la Inteligencia Artificial, la ingeniería, la representación del conocimiento y la lingüística computacional como modelos de representación del conocimiento. En el campo de la documentación, las ontologías son una herramienta para el intercambio y uso del conocimiento ya que proveen una comprensión compartida y consensuada de un dominio del conocimiento, que puede ser comunicada entre personas y sistemas heterogéneos. Dentro de los sistemas de organización del conocimiento, las ontologías, al igual que los tesauros, se encasillan dentro de las llamadas listas relacionadas.

Existen numerosas definiciones de ontologías, entre las que cabe destacar:

- ✓ "Una ontología es un vocabulario acerca de un dominio: términos + relaciones + reglas de combinación para extender el vocabulario". Neches, 1991.
- ✓ "Una ontología es la especificación de una conceptualización". Gruber, 1993. (Aquí el término conceptualización se refiere a un modelo conceptual). [7]

- ✓ "Una ontología es una especificación formal de una conceptualización compartida". Borst, 1997. (Aquí el término forma se refiere a que es procesable por ordenador).
- ✓ "Una ontología es una base de datos que describe los conceptos generales o sobre un dominio, algunas de sus propiedades y cómo los conceptos se relacionan unos con otros". Weingand, 1997.
- ✓ Una ontología necesariamente incluirá un vocabulario de términos y una especificación de su significado (definiciones e interrelaciones entre conceptos) que impone estructura al dominio y restringe las posibles interpretaciones. Uschold-Jasper.

Tanto los tesauros como las ontologías son herramientas que sirven para estructurar conceptualmente determinados ámbitos del conocimiento por medio de vocabularios controlados. La diferencia entre los tesauros y las ontologías radica en la complejidad ya que estas últimas introducen un mayor nivel de profundización semántica y proporcionan una descripción lógica y formal que puede ser interpretada tanto por las personas, como por las máquinas, mientras que los tesauros sólo pueden ser interpretados por humanos. Las ontologías permiten, además, la interoperabilidad entre sistemas distintos.

En resumen, una ontología es un sistema de representación del conocimiento que resulta de seleccionar un dominio o ámbito del conocimiento, y aplicar sobre él un método con el fin de obtener una representación formal de los conceptos que contiene y de las relaciones que existen entre dichos conceptos. Además, una ontología se construye en relación a un contexto de utilización. Esto quiere decir que una ontología especifica una conceptualización o una forma de ver el mundo, por lo que cada ontología incorpora un punto de vista. Además, una ontología contiene definiciones que proveen del vocabulario para referirse a un dominio. Estas definiciones dependen del lenguaje utilizado para describirlas. Todas las conceptualizaciones (definiciones, categorizaciones, jerarquías, propiedades, herencia, etc.) de una ontología pueden ser procesables por máquina. [8]

Según Gruber, las ontologías se componen de:

- ✓ **conceptos:** son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.
- ✓ **relaciones:** representan la interacción y enlace entre los conceptos de un dominio. Suelen formar la taxonomía del dominio. Por ejemplo: subclase-de, parte-de, parte-exhaustiva-de, conectado-a, etc.
- ✓ **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. Por ejemplo, pueden aparecer funciones como: asignar-fecha, categorizar-clase, etc.
- ✓ **instancias:** se utilizan para representar objetos determinados de un concepto.
- ✓ **reglas de restricción o axiomas:** son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: "Si A y B son de la clase C, entonces A no es subclase de B", "Para todo A que cumpla la condición B1, A es C", etc. Los axiomas, junto con la herencia de

conceptos, permiten inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos. [7]

1.1. Posibles aplicaciones y usos de las Ontologías:

- ✓ repositorios para la organización del conocimiento
- ✓ servir de herramienta para la adquisición de información
- ✓ servir de herramientas de referencia en la construcción de sistemas de bases de conocimiento que aporten consistencia, fiabilidad y falta de ambigüedad a la hora de recuperar información
- ✓ normalizar los atributos de los metadatos aplicables a los documentos
- ✓ crear una red de relaciones que aporte especificación y fiabilidad
- ✓ permitir compartir conocimiento
- ✓ posibilitar el trabajo cooperativo al funcionar como soporte común de conocimiento entre organizaciones, comunidades científicas, etc.
- ✓ permitir la integración de diferentes perspectivas de usuarios
- ✓ permitir el tratamiento ponderado del conocimiento para recuperar información de forma automatizada
- ✓ permitir la construcción automatizada de mapas conceptuales y mapas temáticos
- ✓ permitir la reutilización del conocimiento existente en nuevos sistemas
- ✓ permitir la interoperabilidad entre sistemas distintos
- ✓ establecer modelos normativos que permitan la creación de la semántica de un sistema y un modelo para poder extenderlo y transformarlo entre diferentes contextos
- ✓ servir de base para la construcción de lenguajes de representación del conocimiento [9]

1.2. Algunas de las características:

- ✓ pueden existir ontologías múltiples: si el propósito de una ontología es hacer explícito algún punto de vista, en algunos casos, se necesita combinar dos o más ontologías. Cada ontología introduce conceptualizaciones específicas.
- ✓ se pueden identificar distintos niveles de abstracción estableciendo una topología de ontologías: se puede caracterizar una red de ontologías usando multiplicidad y abstracción. Al no poder realizar una descripción completa del mundo, se puede pensar una estrategia de construcción gradual que vaya de abajo hacia arriba.
- ✓ multiplicidad de la representación: un concepto puede ser representado de muchas formas, por lo que pueden coexistir múltiples representaciones del mismo concepto.
- ✓ mapeo de ontologías: se pueden establecer las relaciones entre los elementos de una o más ontologías para establecer generalizaciones, especializaciones, conexiones, etc.

Se pueden establecer distintos tipos de ontologías atendiendo a diversos aspectos. Se pueden destacar las siguientes clasificaciones, aunque existen otras muchas:

Según el ámbito del conocimiento al que se apliquen:

- ✓ **Ontologías generales:** son las ontologías de nivel más alto ya que describen conceptos generales (espacio, tiempo, materia, objeto, etc.)
- ✓ **Ontologías de dominio:** describen el vocabulario de un dominio concreto del conocimiento.
- ✓ **Ontologías específicas:** son ontologías especializadas que describen los conceptos para un campo limitado del conocimiento o una aplicación concreta.

Según el tipo de agente al que vayan destinadas:

- ✓ **Ontologías lingüísticas:** se vinculan a aspectos lingüísticos, esto es, a aspectos gramáticos, semánticos y sintácticos destinados a su utilización por los seres humanos.
- ✓ **Ontologías no lingüísticas:** destinadas a ser utilizadas por robots y agentes inteligentes.
- ✓ **Ontologías mixtas:** combinan las características de las anteriores.

Según el grado o nivel de abstracción y razonamiento lógico que permitan:

- ✓ **Ontologías descriptivas:** incluyen descripciones, taxonomías de conceptos, relaciones entre los conceptos y propiedades, pero no permiten inferencias lógicas.
- ✓ **Ontologías lógicas:** permiten inferencias lógicas mediante la utilización de una serie de componentes como la inclusión de axiomas, etc. [10]

1.3. Conceptos clave en relación a las Ontologías:

- ✓ **Clase:** Es un objeto que define una categoría. Describe conceptos en el dominio del discurso.
- ✓ **Subclase:** Es en sí misma una clase, pero que es hija de alguna otra clase.
- ✓ **Clase jerárquica:** La compuesta por una colección de clases conectadas por relaciones "es un tipo de" (classhierarchy).
- ✓ **Casos (instances):** Ejemplos específicos pertenecientes a alguna clase, esto es, objetos de una clase.
- ✓ **Roles o Propiedades (slots):** Propiedades de cada concepto que describen varias características y atributos del concepto. Ayudan a definir las características de las clases.
- ✓ **Facetas:** Se utilizan para definir qué tipo de valor puede contener un slot particular, valores permitidos, número de valores, etc. También se denominan restricciones de roles.
- ✓ **Valor:** Describe una propiedad que se aplica a alguna clase o instance.
- ✓ **Tipo:** Define el tipo de valor (como cadena de caracteres, número, booleano, etc.)
- ✓ **Cardinalidad:** Define cuántos valores puede tener un *slot* individual (máximo y mínimo).
- ✓ **Herencia (inherence):** Es el proceso por el cual las subclases e instancias de alguna clase heredan propiedades y valores definidos más arriba en la jerarquía.
- ✓ **Variable:** Espacio vacío que puede llenarse preguntando a clases e instancias. Cada variable comienza con un signo de interrogación.

- ✓ **Relación:** Nuevo conocimiento que se obtiene por deducción, partiendo del conocimiento que se encuentra en la ontología. Las relaciones utilizan variables. [11]

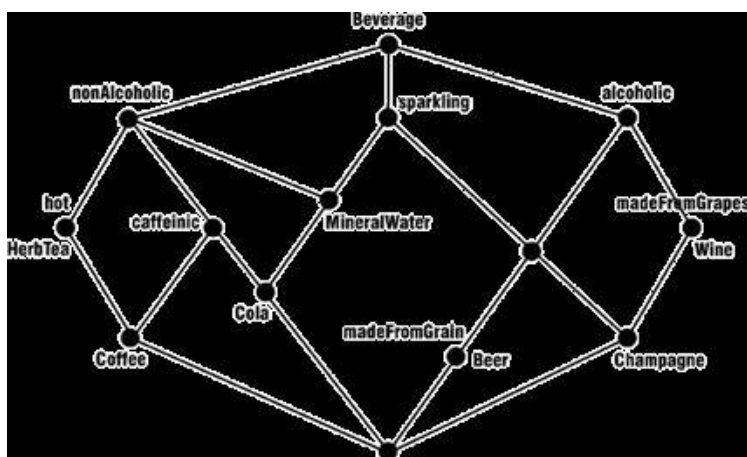


Fig. 1.1. Ejemplo del concepto de relación para una ontología

1.4. Otros conceptos destacables a la hora de hablar de Ontologías:

- ✓ **primitiva:** categoría de una ontología que no puede ser definida en términos de otras categorías en la misma ontología. Un ejemplo de una primitiva es el concepto del tipo Punto en la geometría de Euclides. El significado de una primitiva no está determinado por una definición con una forma cerrada (closed-form), sino por axiomas que especifican cómo se relaciona a otras primitivas. Una categoría que es una primitiva en una ontología debe no ser primitiva en un refinamiento (refinement) de aquella ontología.
- ✓ **base de conocimiento:** término informal para referirse a una colección de información que incluye una ontología como un componente. Además de una ontología, una base de conocimiento debe contener información especificada en un lenguaje declarativo tal como reglas lógicas o sistemas expertos, aunque también incluye información no estructurada o formalizada expresada en lenguaje natural o en lenguaje de procesado.
- ✓ **refinamiento (refinement):** Un refinamiento (matización) de cada categoría de una ontología A hacia alguna categoría de otra ontología B, lo cual se denomina un refinement de A. Cada categoría en A debe corresponder a una categoría equivalente en B, pero algunas de A deben ser equivalentes a no primitivas en B. El refinamiento define un orden parcial de ontologías: si B es un refinamiento de A, y C es un refinamiento de B, entonces C es un refinamiento de A; si dos ontologías son refinamientos una de la otra, entonces deben ser isomórficas. [12]

En el campo de las aplicaciones actuales y de la Web Semántica, las ontologías capturan conocimiento consensuado de un modo genérico y formal de tal manera que pueda ser compartido y reutilizado por distintos grupos de personas y aplicaciones de software. Una de las condiciones para que funcione la Web Semántica es que el contenido de los documentos se presente por medio de la utilización de ontologías que sean públicas y accesibles, de uso común y, a ser posible, normalizadas. Nace así el concepto de documento multifacetado en forma y

contenido, necesario para que la recuperación de información pueda realizarse mediante softbots, esto es, robots y agentes de software que realicen esta labor de forma automatizada. Sin embargo, este proceso de recuperación no se produce de la misma forma en un entorno homogéneo que en un entorno heterogéneo.

Entorno homogéneo	Entorno heterogéneo
<ul style="list-style-type: none"> • Tarea delegada • Consulta de la ontología • Selección de categorías • Identificación de documentos pertinentes 	<ul style="list-style-type: none"> • Punto crítico de inicio • Uso de intermediarios de ontologías • Consulta de ontologías • Selección de categorías • Identificación de documentos • Consulta de ontología de estructura • Identificación de estructura de documento

Fig. 1.2. Entornos homogéneo y heterogéneo respectivamente

La Web Semántica, es una ampliación de la Web actual, se trata de un espacio donde la información tiene un significado bien definido, de manera que pueda ser interpretada tanto por agentes humanos como por agentes computarizados.

Dentro de la Web Semántica puede observarse el papel que cumplen las ontologías en el siguiente esquema:

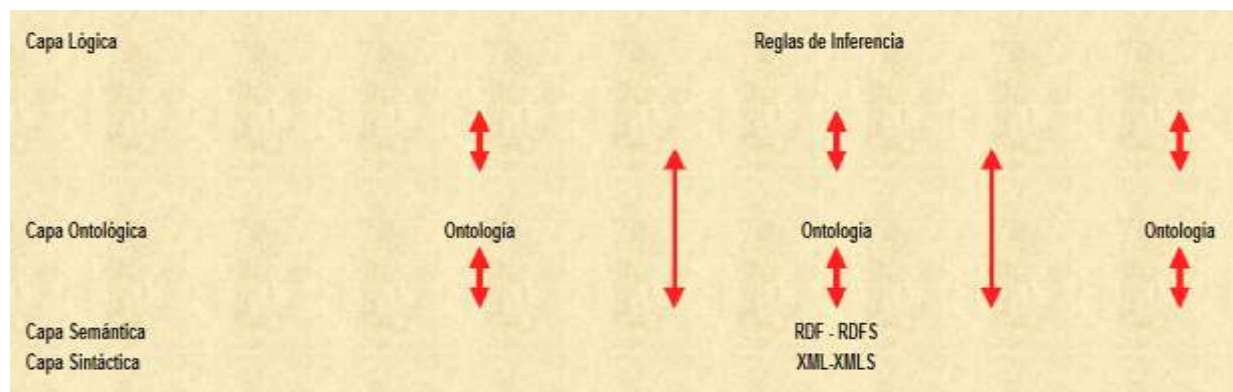


Fig. 1.3. Papel que cumplen las ontologías dentro de la Web Semántica

El lenguaje XML proporciona una forma de escribir datos que es independiente de lenguajes, plataformas y herramientas, y que proporciona una estructura sintáctica para que los datos puedan ser interpretados por computadoras. XMLS (el lenguaje de esquemas de XML) permite la definición de gramáticas y etiquetas significativas para los documentos a través de *namespaces* o espacios de nombre. Sin embargo, XML o XMLS no son suficientes ya que aportan una estructura, pero no una semántica. La semántica es aparente para los humanos, pero no para las máquinas. La semántica estudia cómo los símbolos se refieren a los objetos. Es necesaria más expresividad para el procesamiento semántico y de esta forma se creó el lenguaje RDF, como un lenguaje para modelar los datos. El lenguaje RDF mediante recursos, propiedades (atributos y relaciones para describir recursos) y sentencias (combinación de recursos y propiedades) permite una representación explícita de la semántica de los datos. [13]

RDF carece de poder expresivo (negación, implicación, cardinalidad, etc). Por ejemplo, no es posible especificar las condiciones necesarias y suficientes para definir la pertenencia a una clase. Para lograr una mayor expresividad para el procesamiento semántico, se han desarrollado nuevos estándares para la representación de ontologías que constriñen los vocabularios de descripción de recursos basados en RDF y RDF Schemas (RDFS). Tales ontologías permitirán, entre otras cuestiones, distribuir definiciones autorizadas de vocabularios que soporten referencias cruzadas como los tesauros. Por lo tanto, las ontologías de representación están pensadas para que tomen el papel que hasta ahora ocupaban los tesauros normalizados, pero es preciso un lenguaje estándar que especifique dichas ontologías con mayor precisión que los estándares ISO sobre tesauros. Con ese fin se ha creado el lenguaje OWL, una especificación del W3C para especificar ontologías. Por ejemplo, OWL sí permite definir las condiciones necesarias y suficientes para definir la pertenencia a una clase, luego, aunque basado en el lenguaje RDF, va más allá que RDF Schema.

Las ontologías requieren de un lenguaje lógico y formal para ser expresadas. En la inteligencia artificial se han desarrollado numerosos lenguajes para este fin, algunos basados en la lógica de predicados, como KIF y Cycl que ofrecen poderosas primitivas de modelado, y otros basados en *frames* (taxonomías de clases y atributos), que tienen un mayor poder expresivo, pero menor poder de inferencia; e incluso existen lenguajes orientados al razonamiento como DescriptionLogic y Classic. Todos estos lenguajes han servido para desarrollar otros lenguajes aplicables a la Web. En un lenguaje de ontologías se pretenderá un alto grado de expresividad y uso.

Dentro de los principales **lenguajes de ontologías** pueden destacarse los siguientes:

- ✓ **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.
- ✓ **OIL**: *OntologyInferenceLayer*. Este lenguaje, 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.
- ✓ **DAML y OIL**: Este lenguaje nació fruto de la cooperación entre OIL y DARPA y unifica los lenguajes DAML (*DARPA'sAgentMarkupLanguage*) y OIL (*OntologyInferenceLayer*). Se basa ya en estándares del W3C. El lenguaje DAML se desarrolló como una extensión del lenguaje XML y de Resource

Description Framework (RDF) y para extender el nivel de expresividad de RDFS. DAML- OIL hereda muchas de las características de OIL, pero se aleja del modelo basado en clases (*frames*) y potencia la lógica descriptiva. Es más potente que RDFS para expresar ontologías. Sin embargo, este lenguaje presenta algunas carencias debido a su complejidad conceptual y de uso, complejidad que se intentó solventar con el desarrollo de OWL. No obstante, se desarrollaron muchas aplicaciones que utilizan DAML-OIL y también existen herramientas para convertir DAML a OWL.

- ✓ **OWL:** *OWL Web OntologyLanguage* 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, 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 (ontologías). En realidad, OWL es una extensión del lenguaje RDF y emplea las tripletas de RDF, aunque es un lenguaje con más poder expresivo que éste. Se trata de un lenguaje diseñado para usarse cuando la información contenida en los documentos necesita ser 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 (motores de búsqueda, agentes, etc.).
- ✓ **KIF:** *KnowledgeInterchangeFormat* es un lenguaje para representar ontologías basado en la lógica de primer orden. KIF está basado en la lógica de predicados con extensiones para definir términos, meta conocimiento, conjuntos, razonamientos no monotónicos, 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.
- ✓ **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, etc.

Las ontologías, son pues, vocabularios comunes que, junto con otras tecnologías que proveen de herramientas y lenguajes para generar marcado y procesamiento semántico, harán posible la Web Semántica. Es, pues, necesario, que los documentos generados para la Web, posean una semántica formalizada en ontologías con el fin de que este conocimiento, sea intercambiado por los agentes de *software*.

También existen herramientas y programas para realizar anotaciones en páginas *web* con lenguajes de marcado propios. La mayoría de estos programas permiten describir el contenido de los documentos en forma de metadatos, soportados sobre

una ontología representada en RDF Schema (RDFS) o basados en grafos conceptuales.

Además, para potenciar el uso de ontologías, se han desarrollado aplicaciones específicas de búsqueda de ontologías en la Web, tales como **OntoAgent**, para que indiquen a los usuarios las ontologías ya existentes y sus características para poder utilizarlas en su sistema; o herramientas como **OntoJava**, un compilador que traslada ontologías realizadas con Protégé a bases de datos de objetos Java, **OntoSQL** que permite usar una base de datos relacional como una base de datos deductiva o **RDFCrawler** para buscar y escanear datos RDF en la Web. [14]

2. Sobre Integración de Datos

La arquitectura típica de sistemas de mediación, basada en wrappers, presenta un alto grado de acoplamiento entre los componentes. Ese acoplamiento no permite compartir servicios entre organizaciones o la integración dinámica de nuevas fuentes de datos. Al mejorar la implementación de wrappers publicándolos como servicios Web, haciendo accesible su semántica, estos pueden ser reutilizados por otros mediadores.

Los nuevos trabajos en el área de la integración de datos se han enfocado hacia la adición de nuevos mecanismos para explicitar la semántica de las fuentes de datos [15]. En este sentido, los esquemas de integración basados en esquemas como el de una base de datos relacional se han sustituido por esquemas de integración basados en ontologías.

Para poder integrar datos de varias fuentes XML mediante un sistema de mediación cuyo esquema de integración está basado en ontologías, las asociaciones se deben establecer entre los elementos del esquema de las fuentes y los recursos de las ontologías utilizadas como esquema de integración.

Dado que el esquema de integración está basado en ontologías, tanto las consultas de los usuarios como los resultados que se deben presentar a éstos deben estar descritos en términos de las ontologías usadas como esquema de mediación. Para presentar a los usuarios de un sistema de integración que integra datos de fuentes XML los resultados en términos de las ontologías usadas como esquema de integración, la librería desarrollada para convertir el contenido de un documento XML a instancias de una ontología es un componente imprescindible.

Esta librería permite obtener las instancias presentes en un documento XML para el posterior proceso de filtración para satisfacer las restricciones establecidas por la consulta del usuario.

Al desvincular la información sobre las relaciones existentes entre los recursos y ontologías del sistema de integración, se logra su descripción en los directorios semánticos. Los directorios semánticos ofrecen una infraestructura genérica que puede ser aplicada para el desarrollo de diferentes tipos de aplicaciones, como es el caso del desarrollo de un sistema de integración basado en ontologías.

Gracias a la desvinculación de un sistema de integración específico se proporciona a los desarrolladores un entorno ideal para el desarrollo de diferentes sistemas de integración basada en ontologías con un conjunto de datos común, produciendo soluciones diferentes. Al desvincular los wrappers de los sistemas de integración, publicándolos como servicios Web, esta característica permite que no sólo se puedan compartir descripciones semánticas de recursos, sino que los recursos en sí mismos podrán ser reutilizados entre diferentes aplicaciones. [15]

El desarrollo de una arquitectura como una infraestructura disponible para su uso local o remoto, posibilita que soluciones de integración existentes se adapten al uso de ontologías sin necesidad de incluir la gestión de las mismas como parte de su código interno. Además, soluciones existentes que hagan uso de ontologías podrían delegar estas tareas en otras infraestructuras, centrándose en la mejora de los algoritmos de planificación/reescritura y su posterior evaluación. Una solución de este tipo ofrece nuevas posibilidades interesantes en el desarrollo de sistemas de integración basados en ontologías.

3. Integración de Ontologías y Folksonomías

Una folksonomía es una taxonomía abierta, generada y ampliada por sus usuarios, y cuyos elementos (en lo que sigue, “símbolos”) se utilizan para etiquetar recursos (documentos, servicios, etc.). Típicamente, las folksonomías carecen de estructura jerárquica, y se reducen a colecciones de símbolos no documentados.

Las ontologías parecen encontrarse a sus anchas en el ámbito de las ciencias de la salud (medicina, biología, genética, etc.), en el cual se han construido enormes ontologías. En cambio, en dominios de tipo menos vertical y ligados a aplicaciones colaborativas en Internet, las folksonomías han florecido en los últimos años para etiquetar vídeos, fotografías, enlaces, etc. Esto puede deberse a que existe un condicionante importante a la hora de considerar la adaptación de ontologías y folksonomías a dominios concretos: cuanto mayor sea el grado de especialización del dominio, menor será el número potencial de personas con capacidad de etiquetar recursos, por lo que las folksonomías encontrarán mayores dificultades para implantarse. Resulta evidente que el número de personas capacitadas para etiquetar un vídeo o una fotografía es muy superior al de científicos capaces de etiquetar un gen. Por otra parte, cuanto mayor sea la necesidad de precisión y calidad, mayor será también la motivación para invertir recursos en la construcción de una ontología. En cambio, si el dominio es demasiado amplio, resulta impensable construir una ontología, debiéndose optar por una folksonomía.

El objetivo fundamental del etiquetado de recursos es facilitar su posterior recuperación. La llegada de las folksonomías ha supuesto una revolución en la manera de anotar recursos en la web. En los últimos años han aparecido en la red servicios muy populares basados en la anotación mediante folksonomías (Youtube, del.icio.us y Technorati entre otros muchos). Su irrupción exitosa llevó a algunos a plantearse que las ontologías habían quedado en cierta medida obsoletas [Shirky05]. La realidad es más compleja. Ontologías y folksonomías no deberían excluirse mutuamente, pues presentan algunas características complementarias que, explotadas de forma conveniente, podrían generar sinergias productoras de más valor. Si bien en algunos dominios parece que se adaptan mejor las ontologías (p.e.: biología) y en otros se adaptan mejor las folksonomías (p.e.: portales comunitarios de compartición de medios audiovisuales), en general puede decirse que hay multitud de dominios donde ambas tecnologías podrían apoyarse mutuamente y sumar sus ventajas. [16]

Las aplicaciones basadas en folksonomías se benefician, por ejemplo, de la naturaleza dinámica y extensible de las mismas, unida a su mayor capacidad para ser aceptadas por los usuarios (generalmente los interfaces de anotación basados en folksonomías son más sencillos que sus equivalentes para ontologías), y su potencial para canalizar la colaboración de los usuarios. Por tanto, las folksonomías

resultan muy adecuadas en la interfaz de la aplicación porque su uso es más sencillo.

Por otro lado, las aplicaciones basadas en ontologías explotan su rigor, siendo capaces de ofrecer respuestas basadas en el razonamiento lógico y una estructuración bien definida del conocimiento. Las ontologías resultan idóneas para aportar formalismo en el “back-end” de la aplicación. Unas y otras serían más valiosas si adquirieran las características que posee el otro tipo. Esta es, por tanto, la motivación para integrar folksonomías y ontologías.

Pueden distinguirse distintos escenarios en los que resulta interesante una combinación de las tecnologías ontológicas y folksonómicas.

3.1. Integración de una Ontología y una Folksonomía existentes

La primera categoría comprende aquellas situaciones en las que se pretende integrar una ontología y una folksonomía que existen previamente de manera independiente. Tal sería el caso, por ejemplo, de un portal de compartición de vídeos basado en folksonomías que pretendiera mejorar sus capacidades de recuperación de información y catalogación de contenidos utilizando una ontología de medios audiovisuales (como la que proporciona el estándar MPEG-7). En este caso se trataría de encontrar la forma de alinear los dos artefactos, tendiendo puentes que relacionen los componentes de uno y otro lado. Es posible que, como parte del proceso, sea necesario extender uno o ambos artefactos. [17]

3.2. Extensión de una Ontología existente mediante una nueva Folksonomía

La construcción de ontologías es costosa, por lo que su alcance tiende a ser limitado (en otras palabras: se concentran en un área específica), y una vez construidas, resulta del mayor interés sacar el máximo partido posible a la inversión. Las folksonomías pueden ser una forma de llevar a cabo esta extensión. Una ontología puede ser utilizada como base para generar una folksonomía, o para dirigir su construcción.

Por ejemplo, mientras los portales basados en folksonomías suelen incluir un único campo de texto para que el usuario introduzca una serie de símbolos separados por comas, una ontología podría organizar la información en distintos campos (autor, temática, localización, etc.), adaptándose a las características del contenido. De esta manera se establecería, con sólo un pequeño esfuerzo adicional, una clasificación de los símbolos que resultaría extremadamente útil para mejorar la recuperación de información para habilitar la navegación facetada. Se podrían, además, resolver algunas ambigüedades (ya no habría dudas entre si la foto representa a un “león” o fue tomada en “León”). Un paso en este sentido son las “machine tags”: símbolos extendidos con prefijo y valor, como “medium:paint=oil”, “place:city=Paris”. El popular portal Flickr ha adoptado ya el uso de “machine tags”². Un problema del enriquecimiento de las folksonomías viene dado por la naturaleza dinámica y siempre en expansión de éstas. El enriquecimiento no puede ser, por tanto, llevado a cabo de manera puntual, sino continua.

La conclusión general es que la integración de folksonomías y ontologías es interesante para ciertos tipos de aplicaciones por los beneficios que reporta, es también representable utilizando formalismos como SKOS, pero sin embargo, es difícil de implementar en la práctica debido a la magnitud de la tarea, que no es fácilmente mecanizable.[18]

4. Integración de Esquemas

En trabajos de integración de esquemas de fuentes biomédicas se sugiere la creación de esquemas virtuales, basados en ontologías, enlazados a fuentes de datos físicas, que se unifican para integrar la información de los repositorios físicos originales, aunque existen numerosos datos que deben ser pre-procesados para evitar inconsistencias y errores en los resultados.

En otras investigaciones realizadas la integración de instancias se afronta junto con el preprocesamiento de datos debido a su similitud funcional, utilizando una ontología como estructura de soporte formal. Una vez que la información correspondiente a las inconsistencias detectadas en los datos es almacenada en dicha ontología—proceso supervisado por un experto—los datos pueden ser transformados automáticamente cuando son requeridos por el usuario, de esta manera el análisis estadístico de los resultados muestra datos correctamente integrados y pre-procesados. La posterior aplicación de algoritmos de minería sobre los resultados obtenidos ofrece también mejores resultados finales que las fuentes de datos por separado o utilizando métodos de integración tradicionales en los que no se transforman las instancias. Una ventaja adicional de este enfoque es poder acceder de forma homogénea a información de varias fuentes heterogéneas unificadas y semánticamente coherentes.

Tradicionalmente los métodos disponibles habían abordado sólo el tratamiento de inconsistencias de esquemas. El modelo propuesto se enmarca en el área de la informática—integración de bases de datos y KDD—, y en la biomedicina como dominio de aplicación. [19]

Resultados y Discusión

5. Aplicación de Ontologías a la Integración de Datos

Las ontologías son particularmente idóneas para la tarea de integración, consistente en cubrir los huecos semánticos y sintácticos existentes entre las distintas fuentes de datos (Silvescu et al., 2001; Wache et al., 2001). De hecho, es aquí donde la comunidad científica ha encontrado la mejor asociación entre ontologías y KDD.

En este escenario, la integración de bases de datos está evolucionando actualmente hacia los enfoques híbridos y concretamente en aquellos basados en ontologías. En estos sistemas, se utilizan vistas basadas en ontologías que a su vez pueden realizar parte de la selección de los datos. Estas ontologías facilitan el enlazado entre elementos pertenecientes a una base de datos y conceptos de un vocabulario compartido. Si dos o más bases de datos contienen el mismo concepto, pero está representado por distintos nombres, las ontologías se utilizan para enlazar estos nombres al mismo concepto. Mientras estos enfoques son muy prometedores en el campo de la integración de bases de datos, todavía quedan asuntos importantes que requieren una investigación más extensa. En concreto, los relativos al rediseño virtual de los modelos de datos o la unificación de los mismos. Además, estas investigaciones se han centrado generalmente en la integración de esquemas, mientras que la integración de instancias ha quedado muchas veces olvidada. [19]

Los pocos ejemplos de integración de instancias que existen entre este tipo de sistemas enlazan las instancias de las bases de datos a ontologías de dominio para poder así estandarizarlas (Ceusters, et al., 2004). Sin embargo, esta estrategia puede presentar problemas debido a la gran cantidad de datos que se deben

enlazar, ya que generalmente este ha de ser supervisado, e incluso completamente manual. Otro enfoque existente, aunque mucho menos investigado, es el de la utilización de una ontología como marco que presente una visión general de las posibilidades de preprocesamiento de los datos, y que defina las distintas transformaciones que se pueden llevar a cabo para la integración y preprocesamiento. La utilización de ontologías en esta fase del KDD, aporta una mayor claridad y facilita el entendimiento de la fase de limpieza por un experto en los datos.

Conclusiones

1. El uso de ontologías para la integración de datos tiene algunas ventajas sobre los sistemas basados en palabras clave. Las ontologías proveen un vocabulario compartido común (conceptos) para representar la información incluida en los documentos (contenidos). Además las ontologías permiten definir relaciones entre los conceptos (roles). Tanto conceptos como roles pueden ser usados para realizar consultas más complejas y recuperar de forma precisa la información en la que el usuario está interesado.
2. Tanto los tesauros como las ontologías son herramientas que sirven para estructurar conceptualmente determinados ámbitos del conocimiento por medio de vocabularios controlados. La diferencia entre los tesauros y las ontologías radica en la complejidad ya que estas últimas introducen un mayor nivel de profundización semántica y proporcionan una descripción lógica y formal que puede ser interpretada tanto por las personas, como por las máquinas, mientras que los tesauros sólo pueden ser interpretados por humanos. Las ontologías permiten, además, la interoperabilidad entre sistemas distintos.
3. Una ontología especifica una conceptualización o una forma de ver el mundo, por lo que cada ontología incorpora un punto de vista. Además, una ontología contiene definiciones que proveen del vocabulario para referirse a un dominio. Estas definiciones dependen del lenguaje utilizado para describirlas. Todas las conceptualizaciones (definiciones, categorizaciones, jerarquías, propiedades, herencia, etc.) de una ontología pueden ser procesables por máquina.
4. Pueden distinguirse distintos escenarios en los que resulta interesante una combinación de las tecnologías ontológicas y folksonómicas.
5. La integración de folksonomías y ontologías es interesante para ciertos tipos de aplicaciones por los beneficios que reporta, es también representable utilizando formalismos como SKOS, pero sin embargo, es difícil de implementar en la práctica debido a la magnitud de la tarea, que no es fácilmente mecanizable.
6. Las ontologías son particularmente idóneas para la tarea de integración, consistente en cubrir los huecos semánticos y sintácticos existentes entre las distintas fuentes de datos.

Referencias Bibliográficas

- [1] D. Woelk, P. Cannata, M. Huhns, W. Shen, and C. Tomlinson. (1993): "Using Carnot for enterprise information integration." Second International Conf. on Parallel and Distributed Information Systems.
- [2] Yigal Arens, Chun-Nan Hsu, and Craig A. Knoblock. (1997): "Query processing in the SIMS information mediator". In Michael N. Huhns and Munindar P. Singh,

- editors, *Readings in Agents*, pages 82--90. Morgan Kaufmann, San Francisco, CA, USA.
- [3] D. Dou, D. McDermott, and P. Qi. (2003): "Ontology translation on the semantic web. In *International Conference on Ontologies, Databases and Applications of Semantics*.
- [4] D. Calvanese, G. Giacomo, and M. Lenzerini. (2001): "Ontology of integration and integration of ontologies". In *Description Logic Workshop*.
- [5] E. Mena, V. Kashyup, A. Seth, and A. Illarramendi. (1996): "Observer: An Approach for Query Processing in Global Information Systems based on Interoperation across Preexisting Ontologies". In *Proc. COOPIS*.
- [6] Arano, Silvia. (2003): "La ontología: una zona de interacción entre la Lingüística y la Documentación". *Hipertext.net*, núm. 2.
- [7] Gruber, Tom R. (1993): "A translation approach to portable ontologies". *Knowledge Acquisition*, 5(2).
- [8] Baker, Tom. (2002): "Cores: A Forum on Shared Metadata Vocabularies". *ERCIM News*, No. 51, October. Special Semantic Web.
- [9] Genesereth, Michael R. (_) "KIF Knowledge Interchange Format".
- [10] Gruber, Tom R. (1993): "Toward Principles for the Design of Ontologies Used for Knowledge Sharing". Technical Report KSL-93-04, Knowledge Systems Laboratory, Stanford University, CA.
- [11] Matthews, Briam. Wilson, Michael. Brickley, Dam. (2002): "Semantic Web Avanced Development in Europe". *ERCIM News*, No. 51, October. Special Semantic Web.
- [12] Sowa, John F. (_) "Knowledge Representacion: Logical, Philosophical and Computational Foundations".
- [13] Studer, S. Benjamins, R. Fensel, D. (1998): "*Knowledge Engineering: Principles and Methods*", Data and Knowledge Engineering.
- [14] Tzitzicas, Yannis. Spyratatos, Nicolas. et. al. (2002): "Extended Faceted Taxonomies for Web Catalogs". *ERCIM News*, No. 51. Special Semantic Web.
- [15] Alexander Woehrer, Martin. Bloech Brezany, Peter. Guenther Gruber. Langegger, Andreas. Schentz, Herbert. And Woess, Wolfram. (2005): "Towards Semantic Data Integration for Advanced Data Analysis of Grid Data Repositories". Accepted for the First Austrian Grid Symposium. Schloss Hagenberg, Austria.
- [16] L. McGuinness, Deborah, van Harmelen Frank. (_) " Web Ontology Language ..".
- [17] Golder, Scott and A. Huberman, Bernardo. (_) "Usage Patterns of Collaborative Tagging Systems".
- [18] Miles, Alistair and Brickley, Dan. (2004): "*SKOS Mapping Vocabulary Specification*".
- [19] Pérez del Rey, David. (2007): "Un modelo de integración y preprocesamiento de información distribuida basado en ontologías". Madrid.

