

Ontologías para relacionar contenidos educativos digitales

CLARA LÓPEZ GUZMÁN

*Dirección General de Servicios de Cómputo Académico
Universidad Nacional Autónoma de México*

FRANCISCO J. GARCÍA PEÑALVO

*Facultad de Informática y Automática
Universidad de Salamanca, España*

En los sistemas de enseñanza y aprendizaje en línea se pretende que los contenidos educativos tengan un alto nivel de interoperabilidad, con ello se puede explotar al máximo el uso de una infraestructura digital para comunicar, crear y ejecutar los programas académicos. Es necesario que los contenidos puedan utilizarse en diversas aplicaciones, como en las plataformas de aprendizaje, en las bibliotecas digitales, en los repositorios, en los sistemas de información e incluso en páginas web sencillas. Aunque hay estándares tecnológicos para que estas herramientas puedan comunicarse, los métodos de descripción utilizados requieren de mayor definición para emparar los significados de los descriptores y realizar un intercambio realmente eficiente de información y contenidos.

Por otra parte también se vislumbran las primeras aplicaciones de la Web Semántica en la educación y ésta requiere no sólo de la descripción de los contenidos, sino también de ubicarlos semánticamente en un dominio de conocimiento, para que las computadoras puedan ejecutar más procesos automáticos de los que realizan actualmente, y principalmente para mejorar la recuperación e indización de los recursos docentes dispersos en Internet, así como aquellos almacenados en las diversas aplicaciones de los entornos digitales de aprendizaje.

Los metadatos han coadyuvado en la tarea de organizar la información en la Web, al hacer descripciones genéricas y utilizar lenguajes apropiados para que las computadoras puedan indizarlos. Sin embargo, las etiquetas que se generan no proporcionan mayor descripción de los elementos mismos, más allá de datos que la computadora puede procesar pero no comprender. Con el uso de los metadatos y de los tesauros se pueden obtener jerarquías temáticas sobre los contenidos pero no es posible, de forma automatizada, que las computadoras puedan crear relaciones entre contenidos, así como tampoco se puede delimitar el dominio de conocimiento que comprenden.

Para obtener una mejor interoperabilidad entre contenidos y potenciar así las aplicaciones de los sistemas educativos se le está dando auge a la descripción y relación de contenidos a través de Ontologías, para indizar documentos e incrementar el desempeño de los servicios Web; otra aplicación relevante relacionada con la educación se encuentra en la ingeniería del conocimiento, que apoya fuertemente la tendencia hacia la reutilización de conocimiento para generar más conocimiento, y también hacia la construcción de cursos de forma colaborativa para desarrollar e integrar mejores sistemas de aprendizaje basados en la Web. Dentro de los sistemas de información, las Ontologías facilitan la interoperabilidad entre sistemas heterogéneos y la distribución de conocimiento dentro y entre las aplicaciones de software (García 2004).

LA WEB SEMÁNTICA

Los contenidos disponibles en la Web se encuentran codificados en HTML,¹ un lenguaje que, a través de etiquetas predefinidas, sólo permite codificar contenidos por su estructura y con elementos muy limitados. En este tipo de archivos, únicamente se puede indicar qué es un párrafo y qué es un encabezado pero ningún otro tipo de elementos como el nombre de un alumno y una calificación, ya que no hay etiquetas para ello. En las páginas creadas con HTML, estas etiquetas permiten que la información se despliegue en la pantalla de la computa-

1 Hyper Text Markup Language, <http://www.3w.org/html/>

dora con ciertas características y que sean entonces éstas legibles para el usuario, pero para las máquinas estas etiquetas son de muy poca utilidad ya que para procesar la información requieren de más elementos. Ante esta limitante aparece XML,² un lenguaje que permite incluir etiquetas específicas para diferentes componentes de un documento, y que tienen una sintaxis más detallada y manejable que permite definir documentos con estructuras más descriptivas.

El principal inconveniente es que estos lenguajes no incluyen semántica sobre esas estructuras o sobre el contenido global y que las computadoras no tienen la inteligencia para deducirlo. También se apuntan otros inconvenientes sobre cómo evitar ruido en las búsquedas, navegación jerárquica y poca flexibilidad (García & Sicilia, s/f).

Para atacar estas deficiencias, que repercuten principalmente en la calidad de las búsquedas y las recuperaciones de los contenidos, se está desarrollando una web complementaria en la que la información es definida por su significado y sus relaciones, y no por su estructura morfológica ni por su estilo gráfico. A esta nueva web se le conoce como Web Semántica.³

La Web Semántica se propone darle significado al contenido de los sitios web, y permitir un entorno en el que agentes de software puedan ejecutar tareas complejas para leer las páginas y extraer información más sofisticada, lo cual facilitaría las tareas de indización y de búsqueda pero principalmente las haría más precisas y las acercaría más a la forma en que los seres humanos buscamos (Bernes-Lee, Hendler, Lassila, 2001).

El objetivo más general de la Web Semántica es que la información en la Web sea definida y ligada de tal modo que pueda ser utilizada por máquinas no sólo para propósitos de despliegue sino también para la automatización, la integración y la reutilización de los datos entre diferentes aplicaciones.

Para lograr que los contenidos tengan un significado explícito para las máquinas, la Web Semántica utiliza XML como lenguaje que le da

2 eXtended Markup Language, <http://www.w3.org/XML/>

3 <http://www.w3.org/2001/sw/>

una sintaxis a la RDF⁴ para modelar las relaciones, y las Ontologías para las descripciones y los vocabularios.

Hoy día, las Ontologías se consideran un elemento clave en la arquitectura de la Web Semántica y se vislumbran como un método explícito para hacer la descripción de contenidos educativos que tengan el potencial para relacionarse y reutilizarse más fácilmente.

DEFINICIÓN DE LAS ONTOLOGÍAS

Las Ontologías comienzan a utilizarse a finales de los años 80, en el campo de la inteligencia artificial, aplicándose como un método para compartir y reutilizar el conocimiento. A finales de los años 90, comienzan a utilizarse ya en un contexto web con la intención de adjuntarle descripciones semánticas explícitas a contenidos y servicios.

La definición más citada es la que da Gruber (1993), proveniente de la filosofía, que dice: “Una Ontología es una especificación explícita de una conceptualización”. Una conceptualización es una abstracción de algo que existe; es una forma simple de representación. Una especificación explícita es un conjunto de declaraciones sobre esa representación de lo existente. Con los sistemas informáticos, estas declaraciones se llevan a cabo a través de métodos y lenguajes de computadora.

Uschold y Gruninger (1996) proponen que una Ontología es el conjunto de términos utilizado para referirse a un entendimiento compartido en algún dominio de interés, lo cual se utiliza como un marco unificador para resolver problemas de un mismo dominio. Los mismos autores conciben la conceptualización como la definición de conceptos y sus interrelaciones, y concluyen diciendo que asumen una Ontología como un recuento explícito o la representación de una conceptualización, o de parte de ésta.

ONTOLOGÍAS, TESAUROS Y LOM

La manera más simple de entender una Ontología es considerarla como un vocabulario de términos, y la especificación del significado de

4 RDF. *Resource Description Framework* (Estructura para la Descripción de Recursos). <http://www.w3.org/RDF/>

éstos y de sus interrelaciones en un dominio de conocimiento, para hacer búsquedas y otros usos entre agentes. Esta definición se acerca mucho a lo que hemos definido como un Tesouro. García (2004) hace una interesante revisión de diversos autores que analizan las diferencias y coincidencias entre Ontologías y Tesauros, y resalta que por un lado hay Ontologías *Lightweight* (de peso ligero), que incluyen conceptos, las taxonomías de los conceptos, las relaciones entre los conceptos y las propiedades que describen los conceptos; y que este tipo de Ontologías son similares a los tesauros. Pero por otro lado están las ontologías *heavyweight* (de peso pesado) que incluyen axiomas, y éstas superan el alcance de los tesauros.

El principal potencial de una Ontología es contar con reglas de inferencia que pueden facilitar la generación automática de conocimiento, propiedad de la que carecen los tesauros. Además al describir contenidos educativos difícilmente se recurre a los tesauros para hacer una descripción general; se podrían utilizar tesauros pero como un tipo de vocabulario para el correspondiente elemento de sus metadatos.

Por otra parte, el modelo de datos para contenidos educativos, usualmente utilizado LOM (*Learning Object Metadata*-Metadatos para Objetos de Aprendizaje) (IEEE, 2001), en la Web tradicional, carece de cualquier formalidad en el vocabulario del contenido de cada elemento del estándar, y aunque hay grupos categorizados de tipos de metadatos no se tiene clara la relación entre éstos; y además, no se hace explícito el tipo de convenciones que se siguen para hacer el llenado de los metadatos.

Algunas iniciativas han desarrollado LOM en RDF (RDF Resource Description Framework-Marco de Descripción de Recursos),⁵ con la intención de comenzar a probar este tipo de metadatos en ambientes semánticos y echar a andar el desarrollo de aplicaciones educativas sobre la Web Semántica, basándose en ontologías y en RDF (Lee, Hua & Wang, 2007).

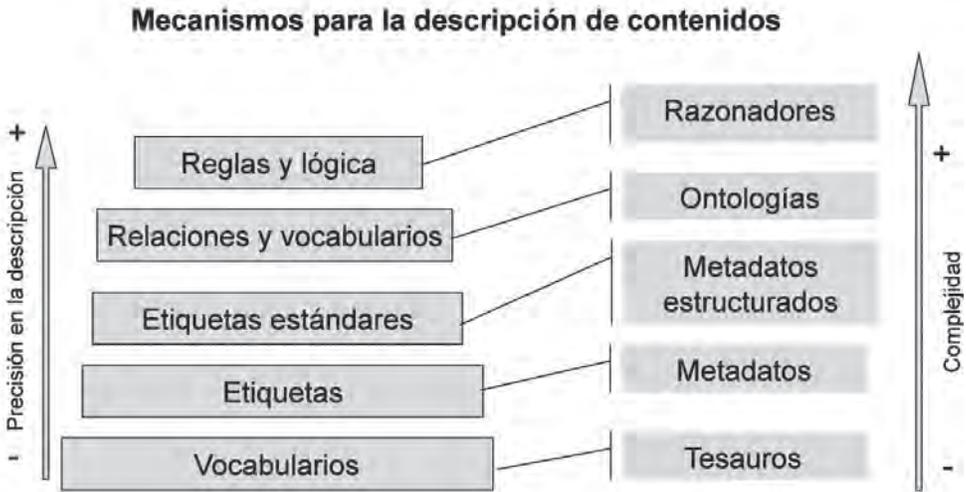
Por encima de los Tesauros y los Metadatos, las Ontologías garantizan consistencia en dominios de conocimiento y hacen explícitos sus

5 Busque "LOM" en <http://swoogle.umbc.edu/> y recuperará diferentes versiones y ejemplos.

vocabularios, pero principalmente lo que aportan son las reglas que les permiten a las máquinas interpretar las relaciones entre contenidos. Los Tesauros y los Metadatos estructuran contenidos; las Ontologías estructuran la semántica de los contenidos.

De esta manera la organización o descripción de contenidos puede llevarse a diferentes niveles de complejidad (figura 1); y por supuesto que una mayor precisión en la descripción de un recurso siempre implicaría mayor complejidad, y aunque los Tesauros han jugado un papel fundamental para el almacenamiento y la recuperación de la información ahora se auxilian de los Metadatos para tener estructuras más consistentes que amplían la descripción de un recurso, sobre todo si se hace uso de metadatos estándares. Las Ontologías tenían aún más estas estructuras y la gran ambición es lograr obtener razonadores automáticos que le simplifiquen el trabajo al humano y descarguen en las computadoras las tareas de organizar, pero sobre todo las de relacionar contenidos.

Figura 1. Niveles de descripción de contenidos



DISEÑO DE ONTOLOGÍAS

El diseño de Ontologías no es una tarea simple, a diferencia de otros mecanismos de descripción, éstas requieren de un profundo conocimiento del dominio que se describe, ya que también hay que definir relaciones; y se pretende que las Ontologías puedan ser reutilizadas y que además puedan ser la base para generar más conocimiento. Dentro de un contexto de conceptualizaciones compartidas, Gruber (1993) dice que los criterios a considerar para el diseño de Ontologías son:

Claridad. La definición debe ser objetiva y comunicar de forma eficiente el significado. También debe ser independiente del contexto informático o social.

Coherencia. Las inferencias deben ser consistentes con las definiciones.

Extensibilidad. El diseño debe permitir diversos usos del vocabulario compartido. Principalmente, una definición debe servir para definir otra.

Codificación mínima. La conceptualización debe especificarse sin depender de símbolos de codificación, ya que los agentes que los pueden utilizar pueden ser llevados a cabo en diferentes estilos de representación.

Mínimo compromiso ontológico. Una Ontología debe tener el mínimo de casos especiales o pretensiones sobre lo que está representando, con la intención de que la propia Ontología pueda después ser más especificada para usos particulares.

Un factor de relevancia es que para que una ontología resulte de utilidad en la Web Semántica tiene que ser aceptada universalmente, y para que esto sea posible es necesario definirla colaborativamente por la comunidad en la que se quiera utilizar (García & Sicilia, s/f).

METODOLOGÍAS PARA CONSTRUIR ONTOLOGÍAS

García (2005) define que para desarrollar una Ontología se requiere, de manera general, definir las clases que forman un dominio, organizar las clases en una jerarquía taxonómica, definir las propiedades

de cada clase e indicar las restricciones de sus valores, y asignarle valores a las propiedades para crear instancias. Sin embargo, no existe una metodología única para construir Ontologías dado que el diseño de ellas no es algo reciente como alternativa para la organización del conocimiento. Entre las más destacadas se encuentran las metodologías de Uschold & King, Methontology y Onto-knowledge, aunque también hay otras, como la IDEF5, Gruninger y Fox y The reference ontology, entre otras.

Es importante resaltar que aun cuando se siga una misma metodología para un determinado dominio, el resultado de cada grupo de diseño es diferente porque se puede interpretar y representar de diferente manera el conocimiento de cada individuo o de cada comunidad.

LINGÜAJES PARA LA DEFINICIÓN DE ONTOLOGÍAS

Para definir una Ontología se requiere de lenguajes que traducen lo que se anotó en el lenguaje lógico a lenguaje de máquina. Estos lenguajes permiten el manejo de taxonomías complejas y de relaciones lógicas, que además se estructuran mediante una semántica definida. Dentro de estos lenguajes, entre los más utilizados, se encuentran OIL (Fensel, 2000) y DAML (Hendler & MacGuinness, 2000), actualmente unidos en uno solo llamado DAML+OIL (Connolly, 2001), y OWL (Lacy, 2005).⁶ Dado que la incorporación de las Ontologías a la web es relativamente reciente estos lenguajes se encuentran en constante evolución, pero pueden utilizarse sin que surjan problemas de inconsistencia por renovación de versiones.

EDITORES DE ONTOLOGÍAS

Cuando no se quiera trabajar la programación de Ontologías directamente en un lenguaje de definición, se puede recurrir a software de autor para el diseño, el cual ofrece interfaces gráficas y amigables.

⁶ Buscar todas estas en la WEB,C.E.

Algunos de estos productos⁷ son Ontolingua y Ontosaurus que fueron las iniciales, y entre las más actuales se encuentran Protégé, WebODE, Ontoedit y Oiled, además de algunas otras como WebOnto como KADS22, Link Factory, OpenknowME y SymOntoX. (Buscar en la WEB,C.E.)

REPOSITARIOS DE ONTOLOGÍAS

Se han realizado ya muchos desarrollos para el uso de Ontologías en diversos ámbitos y existen iniciativas que concentran esos desarrollos en un repositorio, usualmente de acceso libre, que ponen de forma abierta las Ontologías desarrolladas para ser reutilizadas por aplicaciones con objetivos similares. Incluso hay metodologías que proponen el desarrollo de una nueva Ontología a partir de una existente, y entonces se recurre a su búsqueda en algún repositorio. Dentro de los repositorios se encuentran SIMILE, DAMLontology library, OWL Ontology Library de Protégé, OWL SEC, Schema Web, RDFdata.or y SemWeb Central.

MOTORES DE BÚSQUEDA

Para la búsqueda de páginas semánticas, documentos RDF y ontologías, existen motores que “escarban” en la red y localizan este tipo de documentos. Algunos Motores de búsqueda son Semantic Web Search, Onto Search y Swoogle. Dentro de este género también se pueden incluir los *crawlers*, robos que indizan las páginas web con tecnologías semánticas, siguiendo los enlaces que en ésta aparecen, algunos ejemplos son DAML crawler, RDF crawler, OCRA y Scutter.

ONTOLOGÍAS Y CONTENIDOS EDUCATIVOS

Un contenido educativo se encuentra inmerso en un ambiente complejo: usuarios con intereses diferentes, aplicaciones con tecnologías

⁷ Una revisión detallada de más de cien editores se puede encontrar en el trabajo de Denny (2002), disponible en http://www.xml.com/2002/11/06/Ontology_Editor_Survey.html

heterogéneas, contextos variables y usos diversos. El desarrollo de Ontologías en este ámbito comienza a ser una necesidad, porque los sistemas tradicionales de clasificación y organización de la información y del conocimiento tienen limitantes en su diseño que les restringen el trabajo de tareas automáticas a los sistemas informáticos.

En general en el ámbito de la educación es fácil identificar que los autores utilizan diferentes terminologías para conceptos similares. En el campo de la enseñanza en línea esta situación es aún más crítica, y al tratar de utilizar términos en diversas plataformas o en objetos de aprendizaje que tienen movilidad entre aplicaciones el problema se hace más evidente. Por ello en los ambientes digitales educativos tiene sentido el uso de las Ontologías, específicamente para distinguir entre tres tipos de conocimiento: sobre contenido, pedagógico y estructural (Antoniou & Harmelen, 2004).

Una Ontología de contenido describe los conceptos básicos del dominio en el que se lleva a cabo el aprendizaje. Esto incluye las relaciones entre los conceptos y algunas propiedades básicas. Las Ontologías de contenido pueden utilizar relaciones para identificar sinónimos y abreviaturas.

Las Ontologías pedagógicas deben describir los recursos pedagógicos como: tutoriales, ejemplos, ejercicios, evaluaciones, etcétera.

Las Ontologías estructurales se proponen definir una estructura lógica entre los contenidos, según las relaciones de jerarquía, secuencia o navegación, como: previo, siguiente, requiere, es parte de, está basado en, etcétera.

Moral y Cernea (2005) dicen que para los Objetos de Aprendizaje, son necesarias Ontologías de dominio, del alumno, de las teorías de aprendizaje y del diseño instruccional, de la estructura y de los mismos metadatos, lo que resumen como: dominio, estructura, pedagogía y metadatos.

CONCLUSIONES

Es notable la existencia de una constante búsqueda por encontrar mecanismos más precisos para organizar y describir el conocimiento, la información y ahora, más específicamente, los recursos y los conteni-

dos educativos, ya que las aplicaciones requieren descripciones más completas y precisas para poder procesarlos de forma más efectiva.

La complejidad aumenta con la precisión y esto demanda un mayor análisis de parte de los diseñadores y un conocimiento más profundo del dominio que se esté estudiando.

El uso de las Ontologías dentro del ámbito educativo tiene un gran potencial, pero el nivel de desarrollo es incipiente y deben comenzarse a investigar estas líneas de aplicación a fin de dotar a los contenidos educativos de información que los haga más interoperables en sus diferentes perfiles de uso y de sus usuarios. En este trabajo se ha mostrado que ya existen tecnologías, metodologías, herramientas y repositorios para echar a andar aplicaciones específicas, y no debe perderse de vista que para desarrollar Ontologías se requiere de un buen entendimiento con la tecnología y con los conceptos documentales formales.

Para lograr una mejor integración de los contenidos educativos en los ambientes de aprendizaje en línea debe comenzarse a trabajar en el estudio de metodologías específicas para este sector, partiendo de la base de que se deben desarrollar Ontologías de contenido que describan los conceptos del dominio de las diferentes áreas de aprendizaje. Lo que se pretende no es estandarizar el conocimiento, sino las prácticas y las tecnologías que permitan que los contenidos en línea sean más fácilmente manipulados por los sistemas informáticos, y se cuente entonces con aplicaciones más eficientes.

REFERENCIAS

Antoniou, G. & Harmelen van, F. (2004), *A Semantic Web Primer*, Massachusetts Institute of Technology.

Berners-Lee, T., Hendler, J., Lassila, O., "The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities", en *Scientific American*, Mayo, 2001.

Primer Simposio Internacional sobre Organización...

- Castells, P. & Saiz, F., “Ontologías para la web semántica”, en *La web semántica: tecnologías y aplicaciones*, 2003.
- Connolly, D. *et al.* (2001), “DAML+OIL Reference Description”, disponible en <http://www.w3.org/TR/daml+oil-reference>.
- Fensel, D. *et al.* (2000), OIL in a Nutshell, disponible en <http://www.cs.vu.nl/~ontoknow/oil/download/oilnutshell.pdf>.
- García, E., *Aplicación de las ontologías para la representación del conocimiento*, ISKO, Capítulo español Barcelona, 2005.
- García, A. (2004), “Instrumentos de representación del conocimiento: tesauros versus ontologías”, en *Anales de documentación*, núm. 7.
- García, E., Sicilia, M.A., (s/f) “Una propuesta para la búsqueda semántica de recursos web de nuevas tecnologías aplicadas a la educación”, en línea <http://tecnologiaedu.us.es/ticsxxi/comunic/egb-masu.htm>
- Gruber, T. R. (1993), “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”, en Guarino, N., *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Kluwer Academic Press, Boston.
- Hendler, J. & MacGuinnes, D.L. (2000), “The Darpa Agent Markup Language”, en *IEEE Intelligent Systems*, 16(6).
- IEEE (2001), IEEE Learning Technology Standards Committee (LTSC), IEEE P1484.12, Learning Object Metadata Working Group, http://ltsc.ieee.org/wg12/s_p.html/
- Lacy, L (2005), *OWL: representing information using the Web Ontology Language*, Trafford. Victoria, Canadá.

- Lee, M., Hua K., Wang, T. (2007), "A practical ontology query expansion algorithm for semantic-aware learning objects retrieval", en *Science Direct, Computers & Education*, article in press, disponible en <http://simile.mit.edu/papers/iswc05.pdf>.
- Moral, M. E. & Cernea D. A. (2005), "Diseñando Objetos de Aprendizaje como facilitadores de la construcción del conocimiento", en SPDECE 2005, Universitat Oberta de Catalunya, <http://www.uoc.edu/symposia/spdece05/pdf/ID16.pdf>, consultado 7 de abril 2007.
- Uschold, M. & Gruninger, M. (1996), "Ontologies: principles, methods and applications", en *Knowledge Engineering Review*, 11(2).