

De repente, ¿todos hablamos de ontologías?

Por **Sonia Sánchez-Cuadrado**, **Jorge Morato-Lara**, **Vicente Palacios-Madrid**, **Juan Llorens-Morillo** y **José Antonio Moreiro-González**



Sonia Sánchez-Cuadrado es licenciada en Filología por la Universidad Complutense de Madrid y doctora en Documentación por la Universidad Carlos III. Desde 2002 es profesora ayudante en el Departamento de Informática de dicha Universidad con docencia en la licenciatura de documentación e informática sobre temas relacionados con modelado e ingeniería de la información. Actualmente se halla involucrada en distintos proyectos para la creación automática de ontologías.



Jorge Morato-Lara es licenciado en CC. Biológicas por la Universidad de Alcalá de Henares. A partir de 1991 empezó a trabajar en distintas empresas relacionadas con la documentación. Obtuvo el título de doctor en documentación por la Universidad Carlos III en 1999. Desde ese mismo año ha impartido clases en las licenciaturas de Documentación, ADE e Informática en asignaturas relacionadas con la ingeniería de la información, recuperación de información e ingeniería del software.



Vicente Palacios-Madrid es ingeniero superior por la Universidad Carlos III de Madrid. Es profesor de ingeniería del software e ingeniería de la información y miembro del servicio informático de la misma unidad. Actualmente se encuentra realizando la tesis en Documentación sobre interoperabilidad semántica de ontologías.

Resumen: Los artículos sobre ontologías llevan casi veinte años de gran actualidad en la literatura profesional, como se puede comprobar haciendo búsquedas en bases de datos. A pesar de esto, sigue sin haber consenso sobre el significado de este concepto. Probablemente la representación gráfica conocida como el espectro de las ontologías ha generado cierta confusión en algunos lectores. Por ello tratamos de argumentar aquí el origen de esta confusión, que puede deberse a la mezcla de varios tipos de sistemas de organización del conocimiento, cuya distinta finalidad llevó a definirlos de diferente forma. De hecho, construir una ontología formal para un sistema no siempre supone una mejora y en muchas ocasiones es suficiente con una representación menos compleja.

Palabras clave: Ontologías, Tesoros, Sistemas de organización del conocimiento, Espectro de las ontologías, Tipos de ontologías

Title: And suddenly, everybody is talking about ontologies?

Abstract: The number of papers written on ontologies has increased considerably over the last twenty years. This trend can be easily observed by searching words like “ontology” or “thesaurus” in databases. Despite this fact, there hasn’t been a consensus about the significance of this concept. The graphical representation known as ontology spectrum must have generated a lot of confusion amongst readers. In this paper we argue that this confusion is due to the mix of the various types of knowledge organization systems with distinct objectives in the same graphical representation. Thus, constructing a formal ontology for a system does not always presume an improvement, frequently it is adequate with less complex representations.

Keywords: Ontologies, Thesauri, Knowledge organization systems, Ontology spectrum, Types of ontologies

Sánchez-Cuadrado, Sonia; Morato-Lara, Jorge; Palacios-Madrid, Vicente; Llorens-Morillo, Juan; Moreiro-González, José Antonio. “De repente, ¿todos hablamos de ontologías?”. En: *El profesional de la información*, 2007, noviembre-diciembre, v. 16, n. 6, pp. 562-568.

DOI: 10.3145/epi.2007.nov.03



Juan Llorens-Morillo es ingeniero industrial y catedrático de Informática en la Universidad Carlos III de Madrid. Sus principales áreas de investigación se centran en la representación y procesamiento de la información para la reutilización del software.



José-Antonio Moreiro-González es catedrático de Documentación y Decano de la Facultad de Humanidades, Comunicación y Documentación de la Universidad Carlos III de Madrid. Ha escrito diferentes monografías y artículos sobre diversos aspectos de Bibliotecología y Documentación.

Artículo recibido el 13-07-07

Aceptación definitiva: 26-09-07

Es de sobras conocido que las ontologías son un concepto bastante antiguo entre los mayores predicados de la Filosofía. Sin embargo, últimamente, el término “ontología” ha cobrado gran relevancia, siendo posiblemente uno de los que más definiciones y menciones acumula desde principios de los años noventa y, en especial, por parte de la comunidad científica de áreas de conocimiento ajenas a la Filosofía.

¿Qué son las ontologías?

Entre las definiciones más extendidas se encuentran las de **Gruber** y **Guarino**. Para **Gruber** (1993), las ontologías son una especificación de una conceptualización. Mientras, **Guarino** (1998) definió la ontología como un producto de ingeniería consistente en un vocabulario específico usado para describir una realidad más un conjunto de asunciones relacionadas con el significado del vocabulario.

Por otra parte, **Lassila** y **McGuinness** (2001) hablan del espectro de las ontologías, *ontology spectrum*, que no es más que un espacio en el cual se presentan diferentes formas de representar el conocimiento. El marco en el que se desarrolla la comparación toma como elemento principal la riqueza semántica, caracterizando los extremos por semántica débil y semántica fuerte. Aquello que se encuentra próximo al extremo débil representa una semántica simplificada, mientras que lo que se encuentra en el extremo fuerte representa una semántica más compleja. Posteriormente se han realizado revisiones como la de **Daconta** et al. (2003: 157).

En otras palabras, el espectro de las ontologías (Ilustración 1), considera como ontología a toda organización cognitiva que oscile desde la noción más simple de las taxonomías, pasando por los tesauros y modelos conceptuales hasta llegar a las teorías lógicas que representan la noción más compleja. Desde este punto de vista, una ontología define conceptos (significados) usados para describir y representar un área de conocimiento.

¿Por qué la comunidad científica no se pone de acuerdo en qué es una ontología y qué no?

En este punto los expertos se preguntan y discuten si las taxonomías, los tesauros o los mapas conceptuales se consideran también ontologías. En cualquier evento científico donde se hable de ontologías, son comunes dos frases: “[...] sí claro, eso también es una ontología [...]” y/o “[...] no, es que eso no es una ontología [...]”. Tal es así, que aunque **Studer** et al. (1998) incorporaron en la definición los términos conceptualización explícita, formal y compartida (esto último como conocimiento consensuado por un grupo o comunidad), aún hoy paradójicamente el término “ontología” sigue sin tener para muchos una significado consensuado.

Probablemente la comunidad científica discrepa a la hora de determinar qué es una ontología y qué no es debido a que las definiciones propuestas tienen una vertiente específica (cuando se definen las ontologías formales), pero también otra genérica (referida a las ligeras).

En consecuencia, cada disciplina (Documentación, Ingeniería lingüística, Ingeniería de software, Inteligencia artificial, etc.) ha adaptado la definición de las ontologías a sistemas que venían desarrollando antes (tesauros, redes semánticas, modelos conceptuales, etc.). Y así, el espectro de las ontologías alberga un considerable número de conceptos asociados a áreas de conocimiento que hasta el momento rara vez habían sido denominados así.

Por otra parte, es posible que la ilustración del espectro de las ontologías incite a cierta confusión al dibujar en la misma línea modelos de conocimiento y lenguajes para representar el conocimiento.

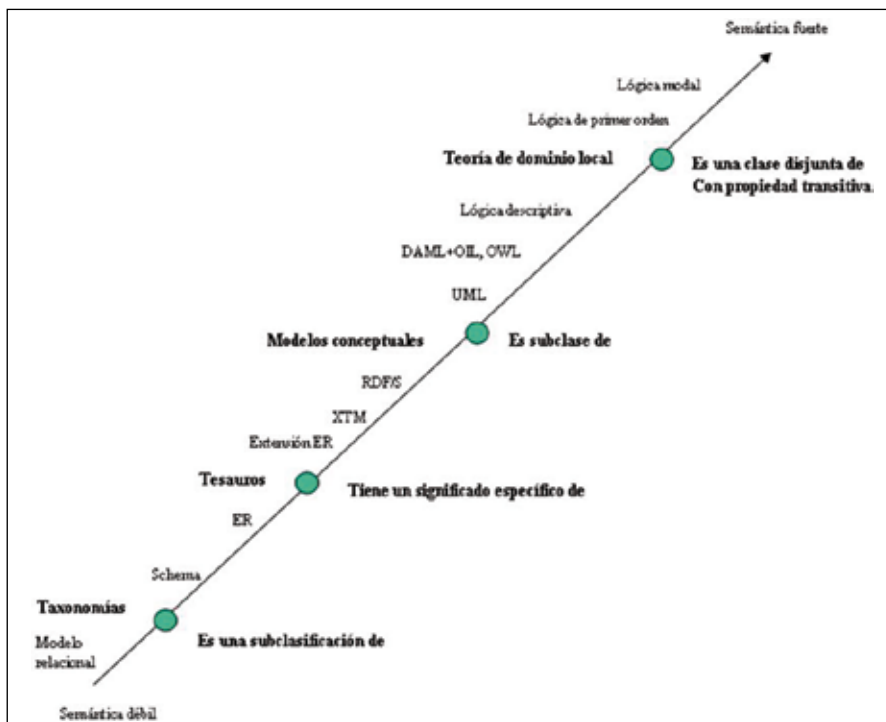


Ilustración 1: Espectro de las ontologías

¿Cuántos tipos de ontologías hay?

Cuando la comunidad científica diferencia entre tipos de ontologías suele clasificarlas siguiendo diferentes criterios: según el contenido (Mizoguchi et al., 1995); según el volumen y el tipo de estructura, pero también por la conceptualización específica del conocimiento (Heijst et al., 1997); según su grado de dependencia -ontologías de alto nivel, de dominios genéricas, de dominio, de aplicación, y de tareas- (Guarino, 1998); o por la riqueza semántica de su estructura interna (Lassila y McGuinness, 2001).

“Se considera ontología toda organización cognitiva”

Siguiendo a otros autores, habría que añadir además las denominadas ontologías lingüísticas (Gómez-Pérez et al., 2004) que consideran las palabras como unidades gramaticales y están pensadas para describir construcciones semánticas, más que para proponer modelos específicos de un dominio. La finalidad de las ontologías lingüísticas, por ejemplo *WordNet* (Miller et al., 1990) o *Sensus* (Swartout et al., 1997), ha sido conformar bases de datos léxicas, aplicándolas a las máquinas de traducción y a la generación de lenguaje natural. Sin embargo, no toda la comunidad científica coincide en otorgar a recursos como *WordNet* la categoría de ontología, a pesar de que a partir suyo se hayan generado ontologías como *Sensus* o *webKB*.

¿Estamos frente a una burbuja ontológica?

La respuesta a esta pregunta es afirmativa, aunque sólo sea parcialmente. Sin duda es un término de moda, del que se han escrito infinidad de documentos que vinculan las ontologías con múltiples áreas de cono-

cimiento, provocando incluso cierta descompensación entre la producción literaria y el número de ontologías existentes.

Una revisión de diferentes bases de datos demuestra esta casuística. Se seleccionan como ejemplo los términos “ontología” y “tesauro”, si bien puede comprobarse que se produce el mismo comportamiento con otros términos referidos a los sistemas de organización del conocimiento. Para ello se realizó una búsqueda con los términos “ontology” y “thesaurus” (junto con sus variaciones gramaticales) para los años 1985, 1995, 2000 y 2003 en las bases de datos *COS Scholar universe: Social science e Ieee* (esta última especializada en artículos de ingeniería a texto completo). Los resultados muestran una ligera tendencia incremental en el uso del término “tesauro” y una fuerte tendencia exponencial en el de “ontología” (tabla 1).

	COS Scholar social		Ieee	
	thesaurus	ontología	thesaurus	ontología
1985	27	159	0	0
1995	44	263	7	11
2000	48	326	11	50
2003	46	417	14	205

Tabla 1. Utilización de los términos ontología y tesauros en las bases de datos COS de CSA e Ieee full-text.

Un análisis de la base de datos de referencia *Research index* muestra un fenómeno análogo.

Si bien la elaboración de una ontología presenta una serie de etapas comunes con la propia de los tesauros, resulta sugestivo comparar cuándo se utiliza un término y no el otro, y cuándo se utilizan de forma conjunta. Dicha comparación se ha realizado con el buscador web de documentación científica *Google Scholar*, y los resultados se pueden ver en la siguiente tabla (tabla 2).

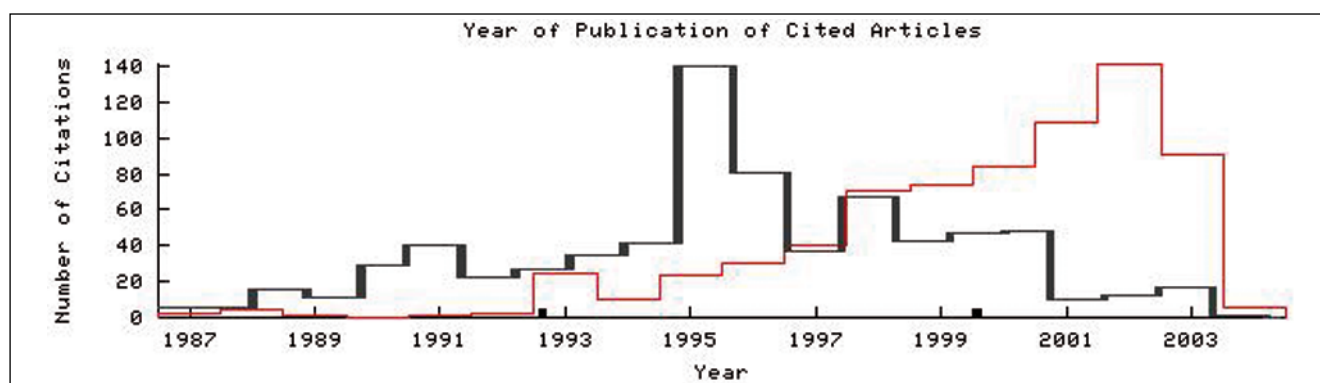


Ilustración 2: Comparación del uso del término “ontologías” (trazo fino) y “tesauros” (trazo grueso) durante el período 1987-2003 según el análisis de citas de ResearchIndex

	Ontologías sin tesauros	Tesauros sin ontologías	Ontologías y tesauros
1985	614	371	4
1995	2.480	1.780	48
2000	6.580	4.180	265
2003	12.800	3.780	608
2006	15.000	3.010	657

Tabla 2. Uso de los términos "ontologías" y "tesauros" en Google Scholar

En esta tabla se observa el uso creciente del término "ontología", y la tendencia inversa para "tesauro", y cómo algunas instancias del término "ontologías" están hablando también de tesauros.

Por otra parte, no es extraño que se escriba tanto sobre las ontologías, no sólo por la controversia del concepto, sino por tratarse de un tema complejo sobre el que existen infinidad de cuestiones pendientes de resolver.

¿Por qué han cobrado tanta importancia las ontologías?

La notable presencia alcanzada por las ontologías puede deberse a su consideración como recurso fundamental para la web semántica (Tim Berners-Lee, 1998). Posiblemente estemos ante otro término de moda, cuya función sea dar respuesta a la necesidad de describir la semántica de un dominio en el entorno web de modo que pueda ser interpretada por las aplicaciones informáticas de forma inteligible para los usuarios.

Asimismo, parte de la relevancia alcanzada puede atribuirse a que durante décadas se ha estado investigando en lo que algunos ahora denominan ontologías y previamente eran tipos de sistemas de organización del conocimiento o *KOS* (*knowledge organization systems*) (Zeng; Chan, 2004).

¿Por qué hay ontologías en todos los ámbitos?

Resulta evidente la necesidad de estructurar semánticamente los conceptos en determinado sistema de organización del conocimiento. El fundamento de las ontologías es representar el conocimiento teniendo como punto de partida la organización taxonómica. Una taxonomía puede ser descrita como un modo de clasificar o categorizar un conjunto de conceptos de forma jerárquica, siguiendo básicamente una estructura en forma de árbol que establece entre los conceptos una relación de generalización-especialización (donde las relaciones son "subclasificación-de" o "subclase-de") y dentro, por tanto, de una semántica simple si seguimos algunas de las propiedades que caracterizan a

las ontologías (Daconta et al., 2003: 145; Z3919:2005: 9). Por este motivo, se trata de clasificar o categorizar un conjunto de conceptos con semántica más compleja que varían desde las taxonomías a las ontologías más completas y formalizadas.

La semejanza de estos sistemas a lo que en inteligencia artificial y la web semántica se denominaba "ontologías" resultaba evidente, siendo importado el término por otras áreas de conocimiento. Por tanto, uno de los motivos que justifica el uso de ontologías en diferentes disciplinas es precisamente que ya existía una forma de representar el conocimiento de un dominio específico, pero recibiendo hasta ahora otra denominación. De otra parte, ciertamente las ontologías suponen un recurso útil para multitud de sistemas de los que se espera que cuenten con relaciones entre sus elementos y cierto grado de inferencia o razonamiento.

Además, hay que tener en cuenta que la ausencia de un estándar que clarifique cuál es el mínimo conjunto de elementos que una ontología debe contener o restringir y la finalidad a la que se debe aplicar, convierten a este término en un "cajón de sastre".

Y entonces...

Una ontología es una organización cognitiva que conforma un sistema de organización del conocimiento. Sin embargo, uno de los principales problemas para representar el conocimiento es el consenso sobre qué representar y cómo hacerlo. Para ello, desde diferentes disciplinas (Biblioteconomía y Documentación, Inteligencia artificial, Ingeniería del software, Lingüística, Ingeniería ontológica, etc.) se ha propuesto una serie de modelos para su representación (Llorens et al., 2004: 221-253).

El grado de representación semántica y la finalidad que se requiera plasmar condicionará los modelos y lenguajes a la hora de construir un sistema de organización del conocimiento, entendiendo como tales aquellos que están englobados en el espectro de las ontologías.

Las ontologías como sistemas de organización del conocimiento tratan de representar tanto información genérica como información concreta. En cuanto sistema de organización, las ontologías pueden ser configuradas siguiendo distintas técnicas de modelado del conocimiento y pueden ser puestas en funcionamiento con diversos lenguajes formales. En muchos casos, los lenguajes para los modelos de representación del conocimiento representan un paradigma completo y un lenguaje de soporte.

Dentro de la web semántica, las ontologías constituyen una pieza clave para el modelado de conocimiento (Antoniou; Harmelen, 2004: 10-18). Las principales

adaptaciones que han sufrido éstas al entrar en contacto con internet han sido:

- la habilitación de elementos para mejorar la interoperabilidad y reutilización en un entorno heterogéneo (como es internet);
- la expresión mediante lenguajes web (como xml o basados en éste).

El potencial de las ontologías dentro de la web semántica está determinado por la capacidad desarrollada a la hora de favorecer la interoperabilidad y la capacidad de reutilización. Estas propiedades se fundamentan en:

- La adopción de lenguajes comunes y compatibles (con un lenguaje de sintaxis, xml, y la expresión

Disciplina	Objetivos	Elementos máximos	Ejemplos de representación del conocimiento	Carga semántica
Documentación	Representar el conocimiento de un dominio	términos	Listados de términos	- complejo
	Clasificar objetos de información	conceptos	Taxonomías	
	Recuperar información	sinonimias	Tesauros	
		taxonomías	Topic Maps	
		relaciones asociativas		+ complejo
Ingeniería del software	Modelado de bases de datos	conceptos	Modelos ER	- complejo
	Modelado de aplicaciones	taxonomías	Diagrama ER extendido	
	Ayudas gráficas para comunicación entre clientes, analistas y desarrolladores	relaciones asociativas	Diagramas de clases de UML	+ complejo
		funciones		
		restricciones básicas		
Inteligencia artificial	Imitar la mente humana	conceptos	Redes semánticas	- complejo
	Almacenar conocimiento común	sinonimias	OWL DL (lógica descriptiva)	
	Con mecanismos para realizar inferencias	taxonomías	SCL (lógica de primer orden)	
	Con mecanismos de aprendizaje	relaciones asociativas		+ complejo
	Con capacidad de operar en sistemas informáticos	funciones		
		restricciones formales		
	reglas de inferencia			

Tabla 3. El espectro de las ontologías desglosado según la finalidad de las disciplinas de Documentación, Ingeniería del software e Inteligencia artificial

normalizada del conocimiento mediante las tripletas recurso-atributo-valor, esto es, rdf).

- La referencia a vocabularios de metadatos para desambiguar conceptos (p. e.: *DC*, *Skos*).
- La creación de ontologías con capacidad de reutilización, como las de alto nivel (*top ontologies*) y las de amplio uso (*generic ontologies*).
- La adopción de paradigmas comunes para expresar el conocimiento (p. e.: *OWL DL* o *SCL*).

Por lo tanto, y a modo de conclusión, el espectro de las ontologías sería más clarificador y completo añadiendo otros parámetros complementarios como la finalidad para la que se construye el recurso. Estos objetivos, en función del dominio de aplicación, se han esquematizado en las tablas 3 y 4.

Las tablas 3 y 4 presentan ejemplos de sistemas de organización del conocimiento propuestos desde diferentes disciplinas. Se observa como los elementos máximos que los caracterizan están condicionados por el área de conocimiento y el objetivo para el que habían sido propuestos.

Para la ingeniería ontológica, las ontologías pueden estar organizadas con distintas técnicas de modelado de conocimiento y ser implementadas en diversos lenguajes (**Uschold; Grüninger**, 1996). Sin embargo, los diferentes modelos y técnicas no permiten representar el mismo conocimiento y, por tanto, tampoco tienen el mismo grado de especificación semántica.

En resumen, una ontología puede tener mayor o menor grado de complejidad y puede ser visualizada

mediante diversos criterios gráficos. No obstante, los diferentes tipos de ontologías o *KOS* que se deban implementar estarán en función de los objetivos de aplicación (tablas 3 y 4). Bajo este principio las ontologías se pueden clasificar según su finalidad, e incluso pueden denominarse de distinta manera dependiendo de la disciplina. Así, no siempre resulta rentable el esfuerzo de realizar una ontología formalizada con axiomas y reglas de inferencia si la finalidad del recurso no lo justifica.

Bibliografía

Antonioni, G.; Harmelen, F. van. *A semantic web primer*. London: The MIT Press, 2004.

Berners-Lee, Tim. *Semantic web road map. Personal notes*. 1998. <http://www.w3.org/DesignIssues/Semantic.html>

Daconta, Michael C.; Obrst, Leo J.; Smith, Kevin T. *The semantic web. A guide to the future of XML, web services, and knowledge management*. Indianapolis: Wiley, 2003.

Gómez-Pérez, A.; Fernández-López, M.; Corcho, O. *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. London: Springer, 2004. p. 403.

Gruber, T. R. "A translation approach to portable ontology specifications". En: *Knowledge acquisition*. Jun 1993, vol. 5, n. 2, pp. 199-220.

Guarino, N. "Formal ontology and information systems". En: *Proceedings of the 1st International conference on formal ontologies in information systems*, Trento, Italy, 6-8 Jun 1998, Amsterdam: IOS Press, 1998, pp. 3-15.

ISO/IEC 13250: 2000. Information technology—SGML applications—topic maps. Geneva: ISO, 2000. <http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0129.pdf>

ISO/IEC 13250:2000. Topic maps: information technology - Document description and markup languages. **Michel Biezunski, Martin Bryan, Steven R. Newcomb** (eds.), 3 Dec 1999.

ISO-2788: 1986. Guidelines for the establishment and development of monolingual thesauri. International Organization for Standardization, Second edition 11-15 UDC 025.48. Geneva: ISO, 1986.

Disciplina	Objetivos	Elementos máximos	Ejemplos de representación del conocimiento	Carga semántica
Web semántica	Representar el conocimiento	conceptos	Vocabularios de metadatos	- complejo
	Clasificar objetos de información	sinonimias	rdf (<i>S</i>), <i>Topic maps</i>	
	Recuperar información	taxonomías	OWL Lite	
	Operar en sistemas heterogéneos con fines similares a la documentación, I.s. e I.a.	relaciones asociativas	OWL DL	+ complejo
	Formalización. Establecer Expresar de forma normalizada el conocimiento	funciones		
	Interoperabilidad	restricciones formales		
		reglas de inferencia		

Tabla 4. Las ontologías según la Web semántica

Lassila, Ora; McGuinness, Deborah. "The role of frame-based representation on the semantic web". KSL tech report number KSL-01-02. Jan 2001. <http://www.ksl.stanford.edu/people/dlm/etai/lassila-mcguinness-fbr-sw.html>

Llorens, Juan; Morato, Jorge; Génova, Gonzalo. "RSHP: An information representation model based on relationships". En: **Damiani, Ernesto; Jain, Lakhmi C.; Madravio, Mauro** (eds.). *Soft computing in software engineering*. Berlin: Springer, 2004. pp. 221-253. (Studies in fuzziness and soft computing series, vol. 159).

Miller, George; Beckwith, Richard; Fellbaum, Christiane (et al). "Introduction to WordNet: An on-line lexical database". En: *International journal of lexicography*, 1990, v. 3, n. 4, pp. 235-244.

Mizoguchi, R.; Vanwelkenhuysen, J.; Ikeda, M. *Task ontology for reuse of problem solving knowledge*. En: **Mars, N.** (ed.). *Towards very large knowledge bases: knowledge building and knowledge sharing (KBKS'95)*. University of Twente, Enschede, The Netherlands. Amsterdam: IOS Press, 1995, pp. 46-57.

Niso. *Ansi/Niso. Z39.19-2005. Guidelines for the construction, format, and management of monolingual controlled vocabularies*. [En línea] <http://www.niso.org/standards/index.html>

Studer, R.; Benjamins, V. R.; Fensel, D. "Knowledge engineering: principles and methods". En: *Data & knowledge engineering*. Mar 1998, v. 25, n. 1-2, pp. 161-197.

Swartout, B.; Patil, R.; Knight K. (et al). "Toward distributed use of large-

scale ontologies". En: *AAAI-97 Spring. Symposium series on ontological engineering*, 1997.

Uschold, M.; Gruninger, M. "Ontologies: principles, methods and applications". En: *Knowledge engineering review*, 1996, v. 11, n. 2, pp. 93-136.

Van Heijst, G.; Schreiber, A. T.; Weilinga, B. J. "Using explicit ontologies in KBS development". En: *International journal of human-computer studies*. 45:183-292. 1997.

Zeng, M. L.; Chan, Lois M. « Trends and issues in establishing interoperability among knowledge organization systems". En: *Journal of the American Society for Information Science and Technology*, 55, 5, 2004. pp. 377-395.

Sonia Sánchez-Cuadrado, Jorge Morato-Lara, Vicente Palacios-Madrid, Juan Llorens-Morillo, Departamento de Informática, Universidad Carlos III.

José-Antonio Moreiro-González, Departamento de Documentación. Universidad Carlos III.

ssanche@ie.inf.uc3m.es

jorge@ie.inf.uc3m.es

palacios@di.uc3m.es

llorens@ie.inf.uc3m.es

jamore@bib.uc3m.es

nature.com
es física

nature.com
es química



Todo está en nature.com

Los nuevos títulos de ciencia de Nature incluyen Nature Nanotechnology y Nature Photonics. Contacte a su representante en Nature para más detalles.

T: +44 (0)20 7843 4759 | E: institutions@nature.com | W: www.nature.com/libraries