

Ontologías para la terminología: por qué, cuándo, cómo

Antonio Moreno Ortiz
Universidad de Málaga

Resum: Aquest article tracta sobre la utilitat que pot tenir per a un terminòleg la integració d'una ontologia en la seva feina, els criteris que cal considerar en el cas de plantejar-se fer-ho, les pautes que cal seguir i les eines que té a la seva disposició. Ofereix una visió actualitzada de l'àmbit d'aplicació de les ontologies, des de la perspectiva del Web Semàntic.
Paraules clau: Terminologia, ontologia, Web Semàntic, representació del coneixement.

Resumen: Este artículo discute la utilidad que puede tener para el terminólogo la integración de una ontología en su trabajo, los criterios a considerar en el caso de plantearse hacerlo, las pautas a seguir y las herramientas a su disposición. Se ofrece una visión actualizada del ámbito de aplicación de las ontologías, desde la perspectiva de la Web Semántica.
Palabras clave: Terminología, ontología, Web Semántica, representación del conocimiento.

Abstract: This article discusses the usefulness of integrating an ontology in one's work when dealing with terminology, the criteria to take into consideration if this is a consideration, the guidelines to follow and the tools available. This article offers an updated review of the application of ontologies from the perspective of Web Semántica.

Key words: Terminology, ontology, Semantic Web, knowledge representation.

1. Ontología: ¿palabra de moda o necesidad?

La palabra "ontología", en el ámbito de las tecnologías de la información y la comunicación, ha suscitado un gran interés en la última década, sobre todo a raíz de la consideración por parte del *World Wide Web Consortium* (W3C) como la tecnología que está llamada a proporcionar la infraestructura de conocimiento a la denominada Web Semántica, a la que algunos ya han dado el sobrenombre de "Web 3.0".

Más aún, desde que el grupo WebOnt, designado por el W3C específicamente para desarrollar un lenguaje de representación de ontologías para la Web, finalizase su actividad en 2004 con la plasmación de su trabajo en el denominado Ontology Web Language (OWL), disponemos por primera vez de un formato estándar para la creación de ontologías, una de cuyas críticas más repetidas hasta entonces era la falta precisamente de un lenguaje consensuado que garantizase aquello que se suponía era una de las principales ventajas del uso de ontologías: la reutilización de recursos y la interoperabilidad de los mismos.

2. ¿Qué es (una) ontología?

Para ser una tecnología cuyo principal cometido es clarificar, explicitar y consensuar el conocimiento relativo a un determinado dominio, resulta paradójico que no exista un consenso claro sobre lo que una ontología es o debería ser (Shirky, 2007). También se ha dicho que el término "ontología" está sobrecargado (Lacy, 2004), en el sentido de que significa cosas diferentes según quién lo defina.

La definición tradicional establece una diferenciación preliminar con el concepto filosófico en el que, la "Ontología", con "O" mayúscula, es una rama de la filosofía que se ocupa de estudiar la naturaleza y organización de la realidad o la existencia, relacionado por tanto con la epistemología. En el ámbito de la ingeniería del conocimiento, la sucinta pero efectiva definición de Gruber (1993) sigue siendo la más comúnmente citada: "una ontología es una especificación explícita de una conceptualización".

La idea que subyace a ambas acepciones, es el estudio de lo que existe (en el caso de la filosofía diríamos más bien "de lo que puede existir"). Se trata de establecer clasificaciones o categorizaciones de las entidades relevantes en un campo de conocimiento concreto, y las relaciones que estas entidades mantienen entre sí.

Otra idea importante a tener en cuenta es que, al contrario que la concepción filosófica, una ontología, en el sentido que nos interesa, no debe ser considerada como una entidad natural

que se descubre, sino como un recurso artificial que se crea (Mahesh, 1996) con un objetivo determinado y para una aplicación concreta.

Esto último tiene importantes repercusiones sobre la anterior, ya que la aplicación determina en muchos casos la categorización que se hace. Una clasificación o taxonomía se hace siempre desde un punto de vista, es decir, teniendo un criterio que es el que determina la jerarquía resultante, y lo mismo ocurre con el resto de relaciones que definen y dotan de expresividad al conjunto. Por tanto, crear una ontología conlleva realizar elecciones y seleccionar criterios y, puesto que el objetivo de una ontología no es otro que servir de referencia a personas y aplicaciones, es necesario que estas elecciones, estas categorizaciones estén consensuadas y aceptadas entre ellos. De este modo, las ontologías son descripciones conceptuales y terminológicas de un entendimiento compartido sobre un dominio específico.

Si dejamos de lado los aspectos relativos a la formalización y la interoperabilidad de aplicaciones, nos damos cuenta de que esto no es sino la principal competencia de la terminología: mejorar la comunicación mediante la utilización de un mismo sistema conceptual y terminológico.

3. La Web Semántica y sus tecnologías

En una entrevista concedida a la revista Business Week en octubre de 2004, Tim Berners Lee, considerado el "padre" de la Web y actual Director del W3C, afirmaba que el tiempo necesario para llegar a tener una Web Semántica completa era de entre 5 y 10 años y, a juzgar por el desarrollo de estas tecnologías y por la expectación generada en empresas, organizaciones e investigadores, su predicción parece correcta.

El objetivo de la Web Semántica no es, como podría pensarse, que las máquinas sean capaces de entender el significado de las lenguas naturales. Como el propio Lee afirmaba en la ya famosa conferencia plenaria en el XML 2000, "semántica" quiere decir, "procesable por la máquina", descartando de forma explícita la acepción que se refiere al procesamiento semántico del lenguaje humano, algo que aún queda lejos. La Web Semántica es la siguiente evolución de la Red, que tiene como característica diferenciadora la inclusión de una semántica explícita comprensible para la nueva generación de aplicaciones Web. Las aplicaciones prácticas de este tipo de recursos "inteligentes" no han hecho más que comenzar a aparecer, pero todo hace presagiar que el cambio será mucho más significativo que el presenciado ante las tecnologías colaborativas de la Web 2.0. Las tecnologías

necesarias para llegar a esta nueva Web ya están disponibles, y las ontologías y sus lenguajes conforman el eje sobre el que éstas giran, pues aportan precisamente la semántica que hasta ahora ha estado ausente en la Web.

OWL no es el único lenguaje ontológico existente. Previamente han existido un número de propuestas con diversa aceptación. Entre éstas son destacables KIF (Knowledge Interchange Format), DAML+OIL, al que se puede considerar precursor de OWL. Sin embargo, OWL es la propuesta de lenguaje ontológico del W3C, con lo que se ha convertido en el estándar de facto para el desarrollo de ontologías y el punto de referencia para todos los que de un modo u otro, ya sea como usuarios o como desarrolladores de tecnología, trabajan con ellas. La sintaxis de OWL está basada en XML (eXtensible Markup Language), lo que también es cierto del resto de las tecnologías de la Web Semántica. Por utilizar un símil lingüístico, XML es algo así como el alfabeto de la Web Semántica, mientras que RDF (Resource Description Format) aporta el léxico (sobre el que OWL construye significados más complejos).

¿Significa esto que es necesario conocer a fondo estas tecnologías para crear una ontología? No, de hecho XML no está pensado para ser utilizado directamente por humanos, es un alfabeto para la máquina y todo lo que se construye sobre él sigue siendo, cuando menos, no muy amigable para el usuario humano. La buena noticia es que no es necesario conocer OWL para crear una ontología en OWL, al igual que no es necesario conocer HTML para crear un documento HTML, aunque como en este último, también es conveniente conocerlo para expresar todas sus posibilidades (en el caso de HTML, sus posibilidades de composición y formato; en el caso de OWL, su potencial expresivo). Más abajo se muestran algunas aplicaciones que nos facilitan las tareas de creación y visualización de ontologías.

4. Relevancia para la terminología

Con lo dicho hasta ahora, parece bastante obvio que el interés de las ontologías para el terminólogo debería ser grande. Una ontología es una formalización consensuada de un sistema conceptual, precisamente el campo de acción del terminólogo, cuyo ámbito de trabajo va mucho más allá de la mera recopilación de palabras pertenecientes a un dominio especializado. El terminólogo trabaja con conceptos, y de forma sistemática elabora y define las relaciones conceptuales que existen entre esos conceptos, crea taxonomías y especifica cuáles son las unidades léxicas que se emplean en las diversas lenguas objeto de su estudio para hacer referencia a esos conceptos. Por tanto, a primera vista parece que la disponibilidad de un sistema de representación conceptual formalizado, explícito y, lo que quizás sea más importante, estandarizado y consensuado, es algo que debería facilitar parte de la labor del terminólogo.

La estructuración conceptual no es nuevo en terminología. De hecho, encontramos referencias a la necesidad de desarrollar esta actividad en cualquier manual de terminología (Sager, 1990; Cabré, 1993; Wright & Budin, 1997). Como nos recuerda Sager:

«A theory of terminology is usually considered as having three basic tasks: it has to account for sets of concepts as discrete entities of the knowledge structure; it has to account for sets of interrelated linguistic entities which are somehow associated with concepts grouped and structured according to cognitive principles; it has, lastly, to establish the link between concepts and terms, which is traditionally done by definitions.» (Sager, 1990: 21).

Luego, la estructuración del conocimiento aparece como el objetivo primero del terminólogo y determinante del resto del proceso. Otra cosa es lo que en la práctica se ha hecho

efectivamente, pues el contexto de trabajo del terminólogo determina en gran medida su metodología y sus resultados, como también lo hace el tipo de herramientas que emplea.

Por otro lado, la definición de “término” de ISO, no deja lugar a dudas que los sistemas conceptuales juegan en el ámbito de la terminología (ISO 1087-1, 2000): “verbal designation of a general concept in a specific subject field”

Existen varias clasificaciones de ontologías dependiendo del criterio que se use (ejemplo claro de lo que mencionábamos más arriba). Dependiendo de su estructuración, una de las clasificaciones que se suelen citar es la que considera un continuum (McGuinness 2002) que iría desde simples listas de palabras hasta estructuras lógicas:

[insertar imagen: “expresividad.tiff”]

Si aceptamos esta gradación, los terminólogos han estado desarrollando ontologías desde siempre, pues, en el peor de los casos, han elaborado listas alfabéticas de términos que,

bajo esta perspectiva, vendrían a ser consideradas como un tipo de ontología poco rigurosa o poco precisa. La elaboración de taxonomías y otros sistemas conceptuales con un grado de formalidad variable también han formado parte del quehacer tradicional del terminólogo sistemático, normalmente como punto de partida para la organización del sistema conceptual de un dominio nuevo. Por supuesto, esto no quiere decir que todos estos tipos de recursos sean lo mismo. En Rees (2003), por ejemplo, se puede encontrar una definición de cada uno de ellos, con atención a las características que los diferencian.

Algo importante a considerar es que la construcción de una ontología, además de aportar rigor y sistematicidad al estudio terminológico, generará un recurso que podrá ser empleado para otros fines además de la motivación principal de la terminología (la traducción). Como recuerdan (Gillam, Tariq & Ahmad, 2005), este aspecto puede resultar una importante motivación, ya que la ciencia terminológica puede contribuir de forma significativa al actual debate sobre las ontologías.

Lo que merece la pena resaltar es que subir uno (o dos) peldaños en esta escala hasta llegar a sustituir nuestro sistema conceptual, más o menos rudimentario, por un sistema de representación conceptual formalizado, no debería resultar especialmente traumático para el terminólogo.

5. ¿Necesito una ontología?

La relevancia de las ontologías parece, por tanto, bastante evidente, como también se puede deducir que crear una ontología, en lugar de una representación conceptual “informal” es una tarea relativamente compleja que repercutirá en el tiempo necesario para completar el trabajo terminográfico. La cuestión es hasta qué punto es rentable o aconsejable esta inversión “extra” de tiempo y esfuerzo.

En general, cuanto mayor es el alcance del trabajo terminológico, mayor debería ser el rigor empleado para desarrollarlo. Esto es cierto no sólo en cuanto a la estructuración conceptual, sino también en cuanto a la adopción de estándares de intercambio de datos. Idealmente, el sistema conceptual para un proyecto terminográfico a gran escala debería definirse a priori, ya que habrá de conformar el esqueleto sobre el que se construirán las descripciones lingüísticas.

El proyecto GALEN es un ejemplo de sistema basado en ontologías. Las aplicaciones que se están creando a partir del sistema ontológico y terminológico de GALEN van mucho más allá de un sistema de consulta y referencia, precisamente gracias a su enfoque de conocimiento formalizado (OpenGALEN).

En España, el proyecto OncoTerm (López-Rodríguez, Tercedor-Sánchez & Faber, 2006) es un buen ejemplo de un proyecto diseñado desde el principio para estar basado en una estructuración conceptual formalizada mediante una ontología, además de emplear los estándares de intercambio de información terminológica de ISO.

La estructuración conceptual se puede construir también de abajo arriba, partiendo de una terminología existente, a la que se le dota con posterioridad de un sistema ontológico. Los recursos terminológicos de la FAO (Food and Agriculture Organization) de Naciones Unidas es un buen ejemplo de este caso, para el que se construyó una ontología explícita a partir de un tesoro controlado (Soergel, 2004).

La utilidad de este tipo de recursos para la terminología ad hoc o a pequeña escala es limitada, pero no despreciable. La formulación de un dominio de conocimiento en sus más pequeños componentes de conocimiento de una manera sistemática, que no es sino la metodología impuesta por la construcción de una ontología formalizada, nos obliga a pensar detenidamente en la verdadera naturaleza cognitiva de las entidades con las que estamos tratando, nos conduce hacia una comprensión más profunda del dominio en su globalidad y nos ayuda a descubrir inconsistencias en nuestra estructuración.

6. Herramientas de gestión ontológica

Lo que hoy en día está al alcance del terminólogo es una serie de recursos para dotar de expresividad al sistema conceptual subyacente. El número de aplicaciones y utilidades para la estructuración conceptual ha crecido conforme el interés en este tipo de recursos ha ido aumentando. En Denny (2004) se puede encontrar un repaso pormenorizado de todas las aplicaciones existentes, sus funcionalidades, disponibilidad, etc.

En el ámbito de la terminología, de forma específica y obviando los rudimentarias funcionalidades de estructuración conceptual de la mayoría de los sistemas comerciales, existen pocas aplicaciones disponibles. OntoTerm (Moreno-Ortiz 2000a; 2000b; Moreno-Ortiz & Pérez-Hernández 2000), es una aplicación de gestión terminológica para MS Windows, disponible de forma gratuita para fines no comerciales. Esta aplicación fue la utilizada para el desarrollo del proyecto Oncoterm, anteriormente mencionado y el resultado de este trabajo, generado por OntoTerm, se puede consultar en <http://www.ugr.es/~oncoterm/alpha-index.html>. OntoTerm es capaz de generar páginas web a partir de la información contenida en la ontología y en las bases de datos terminográficas, integrando así en el resultado de referencia tanto las estructuras conceptuales como la información terminográfica. Sin embargo, no utiliza un lenguaje estándar de representación para la ontología (sí para la terminología).

La aplicación genérica de edición de ontologías más utilizada y de mayor aceptación es sin duda Protégé, desarrollada por el Stanford Center for Biomedical Informatics Research. Protégé es gratuito y de código abierto, cuenta con una sólida base de usuarios y desarrolladores, y existen multitud de extensiones (*plug-ins*) y listas de correo de apoyo. Existen dos versiones: Protégé-Frames y Protégé-OWL, que son en realidad dos modos muy diferentes de codificar conocimiento.

La curva de aprendizaje de esta aplicación puede ser bastante inclinada, pero utilizar sus funcionalidades más simples es relativamente fácil. Por ejemplo, existe una extensión (*Taxonomy cut+paste plugin*) que nos permite crear una jerarquía de clases a partir de una lista tabulada. A partir de ahí podremos visualizar nuestra jerarquía usando Protégé o el visualizador OWLViz, enriquecer nuestra estructuración conceptual y, algo muy importante guardar nuestra ontología en OWL/RDF, lista para ser usada por aplicaciones de la Web Semántica.

7. Conclusión

La práctica terminográfica no debe ignorar todo lo que las ontologías ofrecen en la actualidad. Por un lado, la construcción de una ontología, especifica y formaliza un aspecto crucial del trabajo terminográfico: la definición del sistema conceptual que se asume, algo cuya importancia damos por sentada, pero que sólo en contadas ocasiones se lleva a cabo con todo el rigor necesario. Por otra parte, la materialización del sistema conceptual

subyacente a un dominio en un recurso computacional formalizado, como es una ontología, ofrece un indudable valor añadido al valor puramente lingüístico-descriptivo o al eventual carácter normativo del recurso terminológico en cuestión.

Obviamente, este valor añadido tiene su precio, pues no se puede ignorar que la creación de una ontología conlleva también un trabajo extra. Sin embargo hemos de considerar los beneficios de este trabajo: un conocimiento más profundo del dominio y una más exacta especificación de su conceptualización, además por supuesto de la disponibilidad de un recurso valioso con unas posibilidades de reutilización muy grandes en campos muy diversos. En última instancia, deberíamos ser capaces de determinar si la creación de una ontología para un dominio o aplicación determinadas merece la pena según criterios prácticos como la finalidad de nuestro trabajo o el presupuesto disponible.

La disponibilidad de herramientas para el terminólogo es suficiente como para tener un arranque con una curva de aprendizaje relativamente suave. Además, muchas de estas aplicaciones están disponibles de forma gratuita. Quizás sea el apartado relativo al intercambio de datos y la reutilización de recursos terminológicos existentes el que más reticencias haya podido generar en el pasado, pero la existencia en la actualidad de estándares, tanto a nivel de recursos ontológicos, como terminológicos, hace que tales "excusas" deban ser por fin, abandonadas.

Referencias

Borgo, S., Guarino, N. & Vieu, L. (2005). "Formal Ontology for Semanticists", Research Institute for Computer Science of Toulouse – CNRS, Laboratory for applied ontology.
<http://www.loa-cnr.it>

Denny, M. (2004) "Ontology Tools Survey, Revisited"
<http://www.xml.com/pub/a/2004/07/14/onto.html>. Fecha de última actualización: 04-07-2004.
Fecha de consulta: 19-05-2008.

Cabré, T. (1993). *La terminología: Teoría, metodología, aplicaciones*. Barcelona: Antártida/Empúries.

Gillam, L, Tariq, M. & Ahmad, K. (2005). "Terminology and the Construction of Ontology. *Terminology* 11(1): 55-81. John Benjamins Publishing Company.

Gruber, T. R. (1993) "A translation approach to portable ontology specifications". *Knowledge Acquisition vol. 5*. London, UK: Academic Press.

López-Rodríguez, C., Tercedor-Sánchez, M. & Faber, P. (2006) "Gestión terminológica basada en el conocimiento y generación de recursos de información sobre el cáncer: el proyecto Oncoterm". En: *RevistaSalud.com*, v. 2, n. 8, 2006.

<http://www.revistaesalud.com/index.php/revistaesalud/article/view/127/322>

Fecha de consulta: 12-05-2008.

ISO 1087-1 (2000). *Terminology work -- Vocabulary -- Part 1: Theory and application*. ISO, Suiza.

Lacy, L. W. (2004). *OWL: Representing information using the Web Ontology Language*. Ann Arbor: Trafford, 2004.

McGuinness, D. L. (2002) "Ontologies Come of Age" en D. Fensel, J. Hendler, H. Lieberman, & W. Wahlster (eds.). *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. Cambridge, Mass: MIT Press.

Mahesh, K. (1996). *Ontology Development for Machine Translation: Ideology and Methodology*. NMSU. Computing Research Laboratory. Technical Report MCCS-96-292. New Mexico.

Moreno-Ortiz, A. (2000a). "Managing conceptual and terminological information in a user friendly environment". Proceedings of OntoLex 2000. Workshop on Ontologies and Lexical Knowledge Bases. Septiembre 2000, Sofía, Bulgaria.

Moreno-Ortiz, A. (2000b). "OntoTerm: un sistema abierto de representación conceptual". Actas del XVI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN). Octubre 2000, Vigo, España.

Moreno-Ortiz, A. & Pérez-Hernández, C. (2000). "Reusing the Mikrokosmos ontology for concept-based multilingual terminology databases". Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC 2000). Junio 2000, Atenas, Grecia, 1061-1067.

Rees, R. van (2003). "Clarity in the Usage of the Terms Ontology, Taxonomy, and Classification". Paper for the 2003 CIB Workgroup 78 Conference in Auckland.
http://vanrees.org/research/papers/2003_cib.pdf/view

Sager, J. C. (1990). *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: Benjamins.

Soergel, D. (2004) Report on an FAO Ontology and Terminology System. Series title: Agricultural Information and Knowledge Management Papers. Food and Agriculture Organization of the United Nations.

Shirky, C. (2007) "Shirky: Ontology is Overrated -- Categories, Links, and Tags"
<http://www.shirky.com/writings/ontology/overrated.html>. Fecha de última actualización: 15-06-2007. Fecha de consulta: 12-05-2008.

Wright, S. E. & Budin, G. (1997). eds. *Handbook of Terminology Management*.
Amsterdam/Philadelphia: John Benjamins.

Recursos on-line

Actividad del W3C en cuanto a la Web Semántica: <http://www.w3.org/2001/sw/>

Especificación de OWL (Web Ontology Language): <http://www.w3.org/TR/owl-ref/>

Entrevista de Business Week a Tim Berners Lee:
http://www.businessweek.com/bwdaily/dnflash/oct2004/nf20041022_6972_db083.htm

OntoTerm: <http://www.ontoterm.com>

OpenGalen: <http://www.open-galen.com>

Protégé en el Stanford Center for Biomedical Informatics Research:
<http://protege.stanford.edu/>

World Wide Web Consortium (W3C): <http://www.w3.org/>