

# Ontologías bajo *Creative Commons*. El futuro del conocimiento en la Web

Gonzalo A. Aranda-Corral  
Joaquín Borrego-Díaz

## **Resumen**

En este trabajo analizamos como se podría extender el sistema de licencias creative commons para hacer frente a algunos retos de la Ingeniería Ontológica, desde dos puntos de vista: la confianza lógica y la confianza social. Estos dos aspectos serán esenciales en la implantación de la Web Semántica. Proponemos un refinamiento de los permisos de derivación de obras para que se pueda controlar, de algún modo, dichos tipos de confianza en aras de un correcto tratamiento del conocimiento representado por la ontología.

# 1 Introducción

La Ingeniería Ontológica es una poderosa aproximación a la interoperabilidad semántica. Los productos construidos, las ontologías, se convertirán, por un lado, en la pieza clave para la implementación de dicha interoperabilidad en las empresas [Pollock y Hodgson 2004], y por otro, en el pilar necesario para la implantación de la Web Semántica. El proyecto de la Web Semántica pretende que el conocimiento almacenado en la Web sea entendido por las máquinas [Berners-Lee et al. 2001]. Es decir, los sistemas puedan razonar con los datos atendiendo al significado de los conceptos y relaciones utilizados. Ese significado se obtiene con la ontología asociada (la formalización de la conceptualización asociado al dominio). Por tanto, sería necesario que las ontologías, como cualquier otra pieza de software, fueran de acceso y uso públicos. Sin embargo, es necesario analizar ciertos aspectos de las licencias existentes, debido al carácter especial de la ontología como representación explícita del conocimiento.

Una definición comúnmente aceptada de ontología es la representación explícita de una conceptualización, donde por conceptualización entendemos una visión del mundo; una forma de pensar sobre un dominio concreto, incluyendo un conjunto de reglas que restringen la estructura de un trozo del mundo real. La ontología se expresa como un conjunto de conceptos, definiciones y relaciones.

El uso de las ontologías se ha convertido en clave para la nueva generación de servicios web [Cardoso y Sheth 2006], y su construcción, mantenimiento y manejo en una rama de la Ingeniería del Conocimiento [Gómez-Pérez et al. 2004]. Actualmente, las ontologías se consideran, en algunos aspectos, una pieza del software, aunque posee características singulares. Existen multitud de ontologías que se pueden clasificar mediante distintos factores, como el dominio representado, su finalidad de uso, el grado de formalización, etc. Ejemplos de ontologías son, por ejemplo, CyC (<http://www.cyc.com>), SUMO (<http://ontology.teknowledge.com>) y ConceptNet (<http://www.conceptnet.org>) que son de propósito general, TOVE (<http://www.eil.utoronto.ca/tove/toveont.html>) o Enterprise (<http://www.aiai.ed.ac.uk/project/enterprise/enterprise/ontology.html>) para el ámbito empresarial, NCI (<http://www.nci.org>) de medicina, GO ([www.geneontology.org/](http://www.geneontology.org/)) para genética, entre otras. Para su representación se usan formalismos variados, como KIF (<http://logic.stanford.edu/kif/kif.html>), OWL (<http://www.w3.org/2004/OWL/>) o de tipo red semántica, como en el caso de ConcepNet.

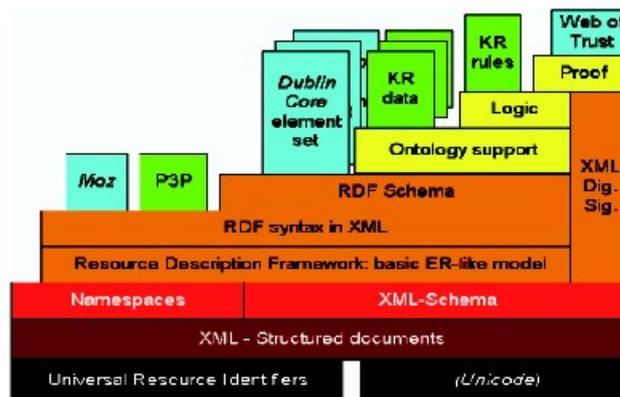


Figura 1: Esquema de la Web Semántica

### 1.1 Ontologías y confianza de la Web Semántica

Una idea de la Web Semántica nos la muestra la figura 1. En este esquema se observa el papel clave de las ontologías como soporte para la demostración (automática); es decir, para el razonamiento de los agentes que trabajarán en la WWW. Una característica importante de esta construcción es que los resultados que se siguen de una demostración son fiables lógicamente, lo que proporciona una confianza en dichos resultados que, en la actualidad, no es posible conseguir. Es, por tanto, imprescindible que la libertad de uso, derivación, tratamiento y aplicación de una ontología sea monitorizado para conseguir que la confianza en la Web Semántica no sea dañada.

Sin embargo, es también muy importante que las ontologías sean de uso universal, por lo que el conocimiento que representan debe ser de uso libre.

## 2 Ontologías de código abierto y conocimiento propietario

El tradicional dilema código libre/código propietario posee un carácter diferente en el caso de las ontologías. Partiendo del principio esencial de la Web Semántica que afirma que el código de una ontología representa el conocimiento asociado al dominio de trabajo, el dilema anterior se transforma en conocimiento libre/conocimiento propietario. Adicionalmente, existen otros aspectos como el grado de libertad para construir ontologías derivadas, para la integración ontológica, etc. que deben ser discutidos. La tesis de este trabajo es que, posiblemente, los tipos de licencias existentes sean insuficientes para abarcar diversos casos de uso atendiendo a diversos factores asociados a la confianza en el conocimiento representado.

## 2.1 Algunos problemas

Fundamentalmente, existen dos tipos de problemas en el uso libre de ontologías: la confianza lógica y la confianza en el conocimiento.

La confianza lógica en una ontología se basa en la adecuación, la consistencia y en cierta medida, en su completitud (como por ejemplo la variante de completitud descrita en [Borrego-Díaz et al. 2005]) como teoría lógica que es. Tales aspectos serían certificados, posiblemente, con sistemas de razonamiento automático. Por tanto, la modificación libre de una ontología puede dañar estas características, añadiendo inseguridad lógica a su uso.

	Licencia Creat.Comm.
	Reconocimiento
	No Comercial
	Sin Obra Derivada
	Compartir Igual

Figura 2: Características de las licencias creative commons

	Licencia Creat.Comm.	Conf. lógica	Conf. social
	Reconocimiento	NO	NO
	Reconocimiento -Sin Obra Derivada	SI	NO
	Reconocimiento -Sin obra derivada -No comercial	-	-
	Reconocimiento -No comercial	-	-
	Reconocimiento -No comercial -Compartir Igual	-	-
	Reconocimiento -Compartir Igual	NO	NO

Figura 3: Situación de las licencias con respecto a los dos tipos de confianza

Por el contrario, la confianza en el conocimiento representado es una noción de carácter social. La revisión puede producir una ontología consistente pero dañina o inadecuada para el procesamiento del conocimiento [Alonso et al. 2006]. De ahí la necesidad de combinar esa revisión con las redes de confianza asociadas a los diseñadores y revisores. En la figura 3 resumimos la situación, bajo las actuales licencias creative commons (a partir de las características descritas en la figura 2), de los dos tipos de confianza (lógica y de

reputación/social) suponiendo la aprobación/certificación del autor de la obra original:

Dos posibles soluciones a los citados problemas de confianza sería la creación del permiso para la modificación certificada y de la revisión limitada a redes de confianza, pero no serían las únicas.

## **2.2 Propuestas de extensión de las licencias para obras derivadas**

En Ingeniería Ontológica, no es suficiente la mera construcción de una ontología. Es evidente que la información no será estática. Las ontologías deben mantenerse como cualquier otra componente de los sistemas. La reutilización del conocimiento requiere que las ontologías sean extendidas, refinadas o integradas [Staab y Studer 2004] con seguridad. La evolución de las ontologías se convierte así en uno de los temas críticos en la Web Semántica, implicando tanto problemas de Representación del Conocimiento como de Procesamiento Inteligente de la Información.

Debido a la dificultad del problema de la confianza, parece necesario la creación de permisos con las siguientes características (la lista no pretende ser exhaustiva) que especifiquen la derivación de la ontología:

### **1. Libertad de obras derivadas certificadas (LDC).**

Es decir, se permite la obra derivada pero se debe adjuntar una certificación de su confianza, bien certificando su consistencia lógica, bien argumentando los cambios. Ambos aspectos ya han sido tratados en Ingeniería Ontológica, y afectan a la documentación adjunta a la ontología.

### **2. Libertad de obras derivadas con confianza social (LDCS).**

Esta licencia permitiría la creación de obras derivadas con cierta estimación de la confianza en el autor de la nueva obra, a partir del autor de la obra original. En casos excepcionales, es este último el que debe tener la potestad de considerar y aceptar la obra como derivada (por ejemplo, en el caso de ontologías para sistemas críticos). Nótese que este último aspecto no es un tipo de licencia. Sería más bien un tipo de certificación que podría ser remunerada. Es evidente que esta certificación depende fuertemente de la reputación social del autor de la obra original.

### **3. Libertad de integración (LI).**

La integración de ontologías es una tarea esencial en el campo de la interoperabilidad semántica. Básicamente, existen tres tipos de reconciliación del conocimiento:

- (a) la fusión (*ontology merging*), que produce una nueva ontología a partir de la mezcla de las ontologías iniciales,
- (b) la alineación, que establece relaciones entre los elementos de las dos ontologías, y
- (c) la integración, que no las une en una sola ontología, sólo establece mecanismos para utilizarlas conjuntamente (completas o en parte).

En cualquiera de los tres casos, parece necesario que la integración sea permitida sin producción de obra derivada previa. Si fuese necesaria la derivación, los aspectos comentados en el anterior punto deberían ser considerados con respecto a la certificación. Por tanto, LI permitiría estas acciones.

#### 4. Libertad de contextualización (LCon).

Este tipo de licencias es, quizás, la más difícil de precisar. Se refiere a la posibilidad de extraer o reducir la ontología para trabajar en dominios más específicos, o para hacer factible el razonamiento automático. El razonamiento contextual es un problema importante en Inteligencia Artificial, y difícil de realizar cuando se desean combinar resultados de distintos contextos. En principio, no existe una clasificación clara de los tipos de razonamiento contextual que podemos utilizar con las ontologías, y por tanto, no parece fácil hacer distinciones sobre la certificación de la ontología derivada.

#### 5. Libertad de Instanciación (LINS).

Este tipo de cesión permitiría la instanciación de ontologías generales en otras más específicas. La cuestión de la certificación de tal instanciación es similar a las ya tratadas en los casos anteriores. En principio, las ontologías generales se utilizan para instanciar los conceptos de otras más específicas (véase fig. 4), lo que induce a sugerir que las generales deberían permitir la instanciación.

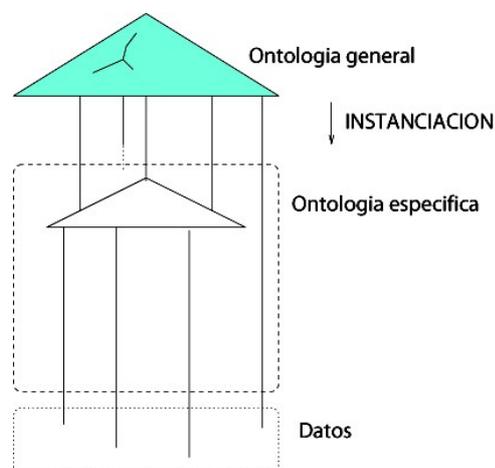


Figura 4. Libertad de instanciación

En la figura 5 presentamos la confianza de cada uno de los anteriores esbozos de refinamiento de la licencia de derivación.

	Refinamiento de la derivación	Conf. lógica	Conf. social
LDC	Obra Der. Certificada	SI	NO
LDCS	Obra Der. con conf. social	SI	SI
LI	Libertad de Integración	NO	SI
LCON	Libertad de Contextualización	NO	NO
LINS	Libertad de Instanciación	SI	SI

Figura 5: Situación de confianza con respecto a las licencias propuestas

Estas propuestas plantean una serie de cuestiones que deben ser solventadas:

- ¿Cuál es el mecanismo de herencia de la confianza social?
- La confianza lógica es, en términos generales, un lujo. Es imposible determinar la consistencia de ontologías de gran tamaño, y por tanto, utilizar LDC puede ser imposible a efectos prácticos.
- ¿Se puede añadir algún tipo de calificación de confianza? Con este término nos referimos a los parámetros utilizados en redes sociales para estimar la confianza en el autor, agente, etc (como en ebay, o el karma en meneame.net). La ventaja de este factor, si existiese uno estándar, es que se podría precisar mejor LDCS.

### 3 Problemas desde la Semántica Emergente

Con los esbozos de licencia anteriormente citados sólo se contemplan aspectos de seguridad lógica y confianza en el autor. Sin embargo, no se contemplan un tipo especial de derechos que, debido a su reciente aparición, no se contemplan en la actual WWW: los derechos de autor sociales. Existen multitud de proyectos realizados por redes sociales que están haciendo emerger distintos tipos de ontologías, que por el momento, no tienen la complejidad suficiente para ser consideradas como útiles fuera de las propias redes. Nos referimos, por ejemplo, a las que se pueden extraer de etiquetados, como la de del.icio.us o flickr, entre otras. Sin embargo, la inminente aparición de sistemas de etiquetados más complejos (como ya prevee la citada flickr), propiciarán la creación de ontologías muy útiles. De hecho, existe ya una comunidad científica dedicada al análisis de importantes fenómenos de semántica emergente en la Web [Cudré et al. 2006] (por ejemplo, en la Web 2.0). Esta situación convierte en algo imprescindible la discusión de dónde están los límites en la propiedad y transformación (individual o colectiva) de dichas ontologías. Un caso especialmente interesante es la emergencia de ontologías a

partir del conocimiento de los miembros de una compañía (por ejemplo, a partir de blogging semántico<sup>1</sup> o blogs corporativos [Dans 2005]).

En la figura 6 se representa la situación en este aspecto.

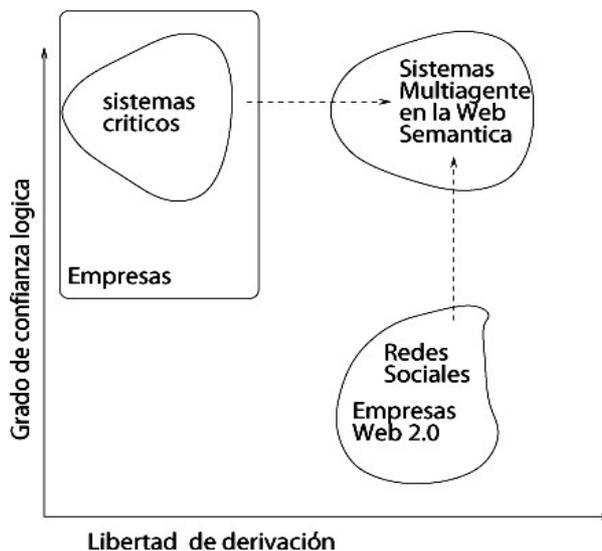


Figura 6: Confianza versus libertad de transformación de ontologías

## 4 Ventajas y desafíos de las ontologías libres para la futura Web

La Web Semántica estará poblada por agentes (software y humanos). Es previsible que, a distinto nivel, muchas aplicaciones sean de uso libre como ahora. De hecho, se desea que la presencia de agentes racionales deliberativos (es decir, aquellos agentes que deducen su comportamiento a partir del conocimiento que poseen) sea significativa. El conocimiento del agente se combinará con el conocimiento ontológico y los datos asociados a la ontología para producir los resultados (véase 7). Por tanto, es indiscutible la necesidad de uso público de las ontologías para proporcionar una ventaja competitiva de unas sobre otras, lo que redundará en un beneficio para la empresa propietaria o usuaria.

### 4.1 Nuevas cuestiones, nuevas propuestas

El uso de las ontologías en la Web Semántica depende de la respuesta adecuada a las siguientes cuestiones (entre otras):

---

<sup>1</sup><http://www.semanticblogging.org/semblog/blog/default/>

- Si se permite la transformación ¿Hasta dónde está dispuesta la empresa propietaria a aceptar como *similar o compatible* la nueva ontología? ¿Cómo puede el autor certificar y/o supervisar las nuevas versiones? Y en el caso extremo de que sea un agente software el que transforma ¿Se permite?

Es evidente que el concepto de ontología de uso libre es la respuesta. Incluso con la cesión de autoría a redes sociales

- El interés en migrar agentes o sistemas que reparen/transformen ontologías [Alonso et al. 2004, Alonso et al. 2003] a la futura Web dependerá de la libertad de transformación. Sin embargo, la transformación no restringida provocaría un efecto totalmente contrario al espíritu de la Web semántica, pues dificultaría el uso compartido del conocimiento de la Web al proliferar ontologías sobre el mismo dominio.

¿Cómo controlar este tipo de permisos? Parece que esta es una cuestión que se puede resolver con licencias del tipo LDC. Una transformación realizada por un agente software es fácilmente documentable (existe algún tipo de traza lógico-computacional). Sin embargo, las transformaciones humanas podrían afectar a aspectos esenciales, y no poseer documentación. Esto puede ocurrir, por ejemplo, cuando el usuario pretende ajustar la ontología a su modelo de conocimiento. De ahí que puedan existir situaciones en las que el propietario de la ontología cierre el derecho a transformación.

Para solucionar estas cuestiones, y para impulsar la interoperabilidad, parece necesario la persistencia de ciertas ontologías propietarias, incluso con código cerrado en el caso de necesitar certificación lógica. Esta solución debe combinarse con la existencia de suficientes herramientas para la integración con otras ontologías asociadas. Por ejemplo, *ontology mapping* parciales, o el uso de articulaciones en base a ontologías más generales establecidas (por ejemplo, TOVE [Fox 1992] para empresas).

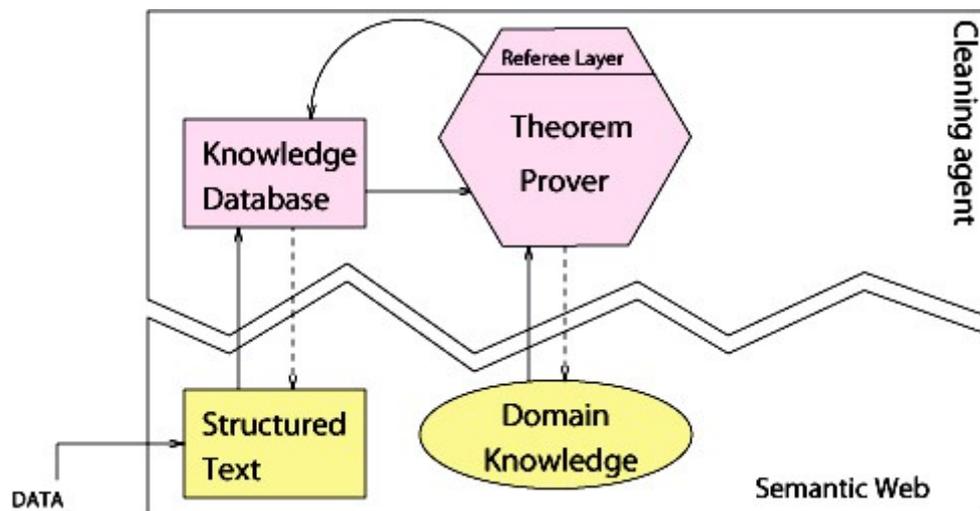


Figura 7: Interacción un agente con datos referenciados con ontologías

## Conclusiones

Los diferentes aspectos comentados en este artículo conducen a pensar en que, para establecer una clasificación adecuada de las nuevas licencias emergentes, es necesario la creación de una ontología de los derechos asociadas a las ontologías. De esta forma el ingeniero ontológico, una vez estudiados los diversos aspectos anteriormente comentados, podría decidir qué aspectos de su obra deben ser protegidos, teniendo en cuenta no sólo su beneficio, sino también para asegurar el correcto uso de las tecnologías semánticas asociadas.

## Bibliografía

[Alonso et al. 2003] J. A. Alonso-Jiménez, J. Borrego-Díaz, A. M. Chávez-González and J. D. Navarro-Marín, A Methodology for the Computerb #Aided Cleaning of Complex Knowledge Databases. 28th Conf. of IEEE Industrial Electronics Society IECON 2002, pp. 1806-1812, 2003.

[Alonso et al. 2004] J.A. Alonso-Jiménez, Joaquín Borrego-Díaz, Antonia M. Chávez-González, Ontology Cleaning by Mereotopological Reasoning. DEXA Workshop on Web Semantics WEBS-2004, pp. 132-137 (2004).

[Alonso et al. 2006] J. A. Alonso-Jiménez, J. Borrego-Díaz, A. M. Chávez-González and F. J. Martín-Mateos, Foundational Challenges in Automated Data and Ontology Cleaning in the Semantic Web, IEEE Intelligent Systems 21(1):42-52 (2006)

[Berners-Lee et al. 2001] T. Berners-Lee, J. Hendler and O. Lassila, The Semantic Web Scientific American 284(5):34-43 (2001).

[Borrego-Díaz et al. 2005] J. Borrego-Díaz and A. M. Chávez-González (2005a), Extension of Ontologies Assisted by Automated Reasoning Systems, Proc. 10th Int. Conf. on Computer Aided Systems Theory (EUROCAST 2005), LNCS 3643, 247-253, Springer.

[Cardoso y Sheth 2006] J. Cardoso, A.P. Sheth (eds.), Semantic Web Services, Processes and Applications, Springer, 2006.

[Cudré et al. 2006] Ph. Cudré-Mauroux y otros, Viewpoints on Emergent Semantics, J. Data Semantics VI:1-27, Lecture Notes in Computer Science vol. 4090 (2006).

[Dans 2005] E. Dans, Blogs y empresa: de la oscuridad a la luz, pasando por las tinieblas, en J.M. Cerezo (ed.) La blogosfera hispana. Pioneros de la cultura digital, pp. 158b #169 Fundación France Telecom, 2005.

[Fox 1992] Fox, MS., The TOVE project: A Common-sense Model of the Enterprise, Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Lecture Notes in Computer Science 604, Springer, pp. 25-34 (1992).

[Gómez-Pérez et al. 2004] A. Gómez-Pérez, M. Fernández-López and O. Corcho, Ontological Engineering: With Examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web, Springer-Verlag, 2003.

[Pollock y Hodgson 2004] J. T. Pollock, R. Hodgson, Adaptive Information, Willey Interscience, 2004.

[Staab y Studer 2004] S. Staaby R. Studer (eds.), Handbook of Ontologies in Information Systems, Springer-Verlag, 2004.

## **Agradecimientos**

Esté trabajo está parcialmente financiado por el proyecto TIN2004-03884 *Sistemas verificados para el razonamiento en la Web Semántica*, Ministerio de Educación y Ciencia, cofinanciado con fondos FEDER.

Gonzalo A. Aranda Corral  
Depto. Ingeniería Electrónica, Sistemas Informáticos y Automática - Univ. Huelva,  
Gonzalo.Aranda@cs.us.es

Joaquín Borrego-Díaz  
Depto. Ciencias de la Computación e inteligencia Artificial - Univ. Sevilla,  
jborrego@us.es