

Lenguajes documentales y ontologías

Por Rodrigo Sánchez-Jiménez y Blanca Gil-Urdiciain

Resumen: Este artículo analiza los principales puntos de convergencia y divergencia entre los lenguajes documentales y las ontologías en tanto que herramientas para la organización del conocimiento y para la recuperación de información en el ámbito de la web semántica. Se describen los aspectos fundamentales de una ontología, así como las principales características semánticas y estructurales de los lenguajes documentales para establecer una comparación entre ellos.

Palabras clave: Sistemas de clasificación, Tesoros, Ontologías, Web semántica, Relaciones semánticas, Análisis comparativo.

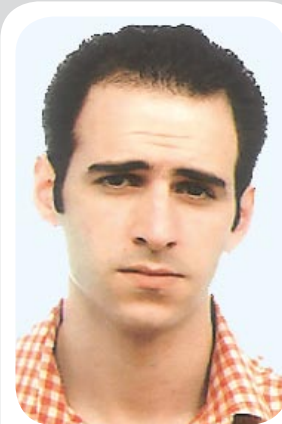
Title: Controlled indexing languages and ontologies

Abstract: The main points of convergence and divergence between controlled indexing languages and ontologies, tools for knowledge organization and information retrieval in the semantic web, are described. Fundamental aspects of ontologies are presented, as well as the basic semantic and structural characteristics of traditional controlled languages.

Keywords: Subject schemes, Thesauri, Ontologies, Semantic web, Semantic relationships, Comparative analysis.

Sánchez Jiménez, Rodrigo; Gil Urdiciain, Blanca. "Lenguajes documentales y ontologías". En: *El profesional de la información*, 2007, noviembre-diciembre, v. 16, n. 6, pp. 551-560.

DOI: 10.3145/epi.2007.nov.02



Rodrigo Sánchez Jiménez es profesor en la Facultad de Ciencias de la Información de la Universidad Complutense de Madrid. Ha participado en proyectos de investigación relacionados con lenguajes documentales, gestión de documentación audiovisual y recuperación de información.



Blanca Gil Urdiciain es profesora en la Facultad de Ciencias de la Documentación de la Universidad Complutense de Madrid. Ha dirigido proyectos de investigación en el campo del análisis documental, indización y lenguajes documentales.

1. Introducción

Tim Berners-Lee (1999) define la web semántica como "una extensión de la web actual en la que se proporciona a la información un significado bien definido, lo que permite a la gente y las computadoras trabajar en cooperación"¹. Por tanto, se trata de una nueva concepción de la www en la que el significado de las cosas y la capacidad de hacerlo inteligible a las máquinas juegan un papel esencial. Para hacerlo posible se cuenta con la capacidad de rdf para establecer descripciones de recursos mediante un modelo sencillo pero potente. Rdf ofrece la posibilidad de que distintas aplicaciones intercambien esta información en condiciones óptimas que permitan su reutilización sin que se pierda significado en el proceso, o sea, que la inicialmente dispuesta

para una aplicación o un dominio en concreto pueda ser aplicada en un ámbito distinto al original².

El modelo de descripción de recursos de rdf se basa en grafos que se pueden expresar como sentencias (oraciones), en las que el sujeto es el recurso a describir, el predicado es una propiedad o característica propia de dicho recurso y el objeto es el valor concreto que tiene dicha característica. Una sentencia rdf cualquiera tiene por tanto una forma básica como la mostrada en la figura 1. Ambos nodos del grafo (sujeto y objeto) pueden formar parte de otras descripciones, de modo que varias descripciones se pueden unir en un solo grafo descriptivo más denso y con información más rica (figura 2).

El grafo anterior describe las relaciones entre una persona y el sitio web que ha creado, su contribución



Figura 1. Grafo representando una sentencia rdf básica

en la revista *Cuadernos de documentación multimedia*, en la que se publicó el artículo identificado por el URI:

<http://multidoc.rediris.es/cdm/viewarticle.php?id=43>

Esto establece una red de relaciones utilizables para diferentes propósitos de forma sencilla.

Sin embargo, el significado de cada uno de los atributos (los arcos o “flechas” del grafo), así como de los tipos de recursos (por ejemplo el tipo “persona”), se debe definir en otro lugar. Es decir, cada grafo rdf establece tan sólo descripciones de recursos de forma similar a una oración más o menos compleja. No obstante, el significado preciso de las palabras que forman dicha oración se debe definir fuera de él (o nos veríamos obligados a concretar cada elemento en juego cada vez que lo utilizamos).

La definición de las “palabras” que se pueden utilizar en una descripción rdf se hace en “vocabularios”, mediante *schemata* rdf. La utilización de un *schema* nos permite definir el vocabulario propio de un dominio para las descripciones rdf, o sea, hace posible delimitar cualidades útiles para describir recursos, así como tipos de recursos de forma explícita para que puedan ser empleados en razonamientos automáticos.

Por otra parte, la utilización de esta codificación de las entidades y propiedades típicas de un dominio debería ser reutilizable fuera del mismo, con el objeto de crear una red de significados interrelacionada, legible y utilizable por máquinas. Es para este propósito para el que las ontologías tienen una gran importancia. No sólo nos permitirán definir vocabularios de forma más precisa, sino que admitirán la posibilidad de relacionar entre sí varios de ellos pertenecientes a diferentes dominios.

Recientemente se ha planteado como un tema de investigación de primera línea la relación entre las ontologías y los lenguajes documentales y se han llevado a cabo numerosos estudios comparativos (Miles, 2001; Guzmán Luna, et. al., 2006; Soergel, et. al., 2004). Además se ha discutido mucho sobre el papel que las ontologías podrían jugar como un nivel superior de desarrollo de los lenguajes documentales.

2. Ontologías

Existen varios paradigmas y muchas formas concretas de representar del conocimiento, pero las ontologías parecen ser la mejor forma en el ámbito de la web semántica. El W3C está haciendo un esfuerzo bastante importante en la difusión del lenguaje para la creación de ontologías *OWL (Ontology web language)*, tal y como se expresa en el *Semantic web activity statement* (Miller, 2004) y en la necesidad de crear ontologías con las que expresar formalmente el significado de las cosas. Su utilización es clave desde el punto de vista de la reutilización del conocimiento en contextos diferentes al original, ya que por su estructura y capacidad de formalización permiten la puesta en relación de diferentes *schemata* rdf.

Una ontología, tal y como se entiende el término en filosofía, es un registro sistemático de las cosas que existen. Esta idea fue tomada del campo de la inteligencia artificial con algunas modificaciones, de forma que para un sistema lo que existe es lo que puede ser representado (Noy; McGuinness, 2000), lo que conlleva que una ontología será un registro de lo que puede ser representado.

“Existen varios paradigmas y formas concretas de representación del conocimiento, pero las ontologías parecen ser lo mejor para la web semántica”

Si buscamos una definición formal de ontología nos encontraremos con que existen varias posibilidades, por lo que optaremos por la clásica ofrecida por Noy y McGuinness (2000), según la cual “(...) es una descripción formal y explícita de los dominios del discurso”³. Gruber (1993) nos ofrece otra más abstracta: “especificación explícita de una conceptualización” siendo ésta “una visión abstracta y simplificada del mundo que queremos representar con algún propósito”⁴. Una conceptualización se compondría a su vez de objetos, conceptos y otras entidades que existen en un

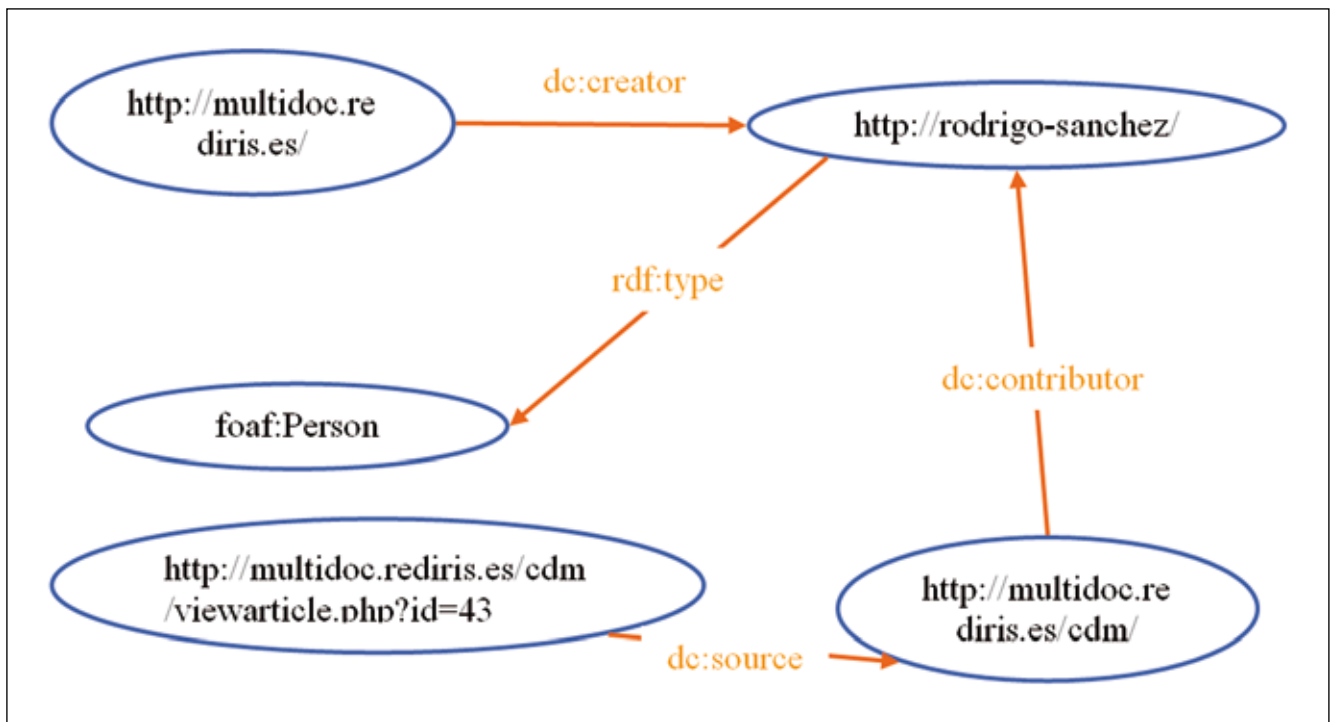


Figura 2. Grafo rdf representando múltiples relaciones sujeto/objeto

área de interés (o dominio) y las relaciones entre ambos. Existen muchas otras definiciones que se recogen en el trabajo de **García Jiménez** (2004).

Lo fundamental de la combinación de los elementos antes mencionados es la posibilidad de formular descripciones utilizables por ordenadores acerca de conceptos básicos en un dominio concreto, así como las relaciones entre ellos, lo que permite reutilizar el conocimiento entre varias áreas de conocimiento. Más concretamente **Noy** y **McGuinness** (2000) destacan cinco aspectos de utilidad:

- Compartir una comprensión común de la estructura de la información entre personas y agentes de software.
- Permitir la reutilización del conocimiento propio de un dominio.
- Hacer las asunciones propias de un dominio explícitas para terceras partes ajenas a él.
- Separar el conocimiento del dominio del conocimiento operacional.
- Analizar el conocimiento del dominio.

Las ontologías se han incorporado con fuerza al ámbito de la web (**Heflin**, 2004) y son utilizadas con cierta frecuencia en la esfera comercial⁵. En general se puede decir que se usan por personas, bases de datos o aplicaciones que necesitan compartir información acerca de un dominio concreto. Algunos ejemplos son los servicios avanzados de búsqueda conceptual, ayuda en la toma de decisiones, entendimiento del habla y el

lenguaje natural, gestión del conocimiento, comercio electrónico, etc.

Es muy interesante, para el caso que nos ocupa, recoger la distinción entre ontologías ligeras y pesadas expresada por **Corcho**, **Fernández** y **Gómez** (2003). Las primeras incluyen conceptos, taxonomías de conceptos, relaciones entre conceptos y propiedades que los describen, mientras que las pesadas tienen en cuenta además axiomas y restricciones de las propiedades. Algunos autores, como **García Jiménez** (2004), vinculan las ontologías ligeras a los tesauros; en nuestra opinión no se puede hablar de una asimilación total debido a la inexistencia de una formalización de las clases en los lenguajes documentales, además de algunos otros aspectos, como se verá más adelante.

Esta distinción entre ontologías ligeras y pesadas parece recogerse en el ámbito de la web semántica mediante la especificación de varios sublenguajes de *OWL* (**McGuinness**; **Van Harmelen**, 2004). La web semántica requiere ontologías con un alto grado de estructuración (**Heflin**, 2004) que especifiquen la descripción de, al menos:

- Clases, tipos de cosas propias de un dominio de interés.
- Relaciones entre esas cosas.
- Sus propiedades o atributos.

Las clases en una ontología se definen a través del conjunto de atributos que comparten los miembros (o instancias) de dicha clase; es decir, existe una formalización explícita de cada una de ellas. Cada uno de los

atributos de una clase expresa una característica, cuyo valor puede ser restringido o matizado, acotando dominios (las clases sobre las que se puede aplicar una propiedad) y rangos (las clases a las que debe pertenecer el valor de una propiedad), cardinalidad, etc.

Podemos establecer vínculos entre clases, al margen de las relaciones de herencia que las vinculan con sus sub-clases, mediante la utilización de propiedades con rangos o dominios relativos a otras clases, de forma que un atributo deba tener como valor un miembro de una determinada clase.

El aspecto externo de una ontología es bastante similar al de un sistema de clasificación, como se puede apreciar en la figura 3. En realidad, los lenguajes documentales permiten establecer relaciones entre conceptos, que pueden ser en principio similares a las que ofrece una ontología, aunque cubriendo eso sí un espectro menos amplio. Sin embargo, un análisis más detallado de las características de los lenguajes documentales y las ontologías muestra que existen diferencias importantes entre ambos.

3. Diferencias entre ontologías y lenguajes documentales

Son apreciables las divergencias tanto a nivel semántico como estructural. En el primer caso, ya sean los sistemas de clasificación o las listas de encabezamientos de materias, se limitan a representar mediante términos la información contenida en los documentos. Los tesauros, en cambio, han tratado de representar mediante conceptos dicha información, con la ayuda de un sistema de relaciones que, aunque complejo, no alcanza la capacidad descriptiva que se desarrolla en las ontologías.

En este sentido, podemos decir que los tesauros intentan llegar al nivel conceptual a través de la utilización de relaciones que están fuertemente ancladas a nivel léxico. Es decir, que se interpreta que el nivel conceptual representado por los términos que componen un tesoro está directamente asociado a dichos términos. Las redes de asociaciones entre términos nos llevan a un término que actúa como representante único de un determinado concepto, pero su asociación con él está ahí desde el principio, anclada en la relación entre el léxico y el significado. En el caso de las ontologías el significado de las cosas se explicita a través de atributos, de las características que son propias de dichas cosas, y no de una representación léxica de las mismas.

Creemos que una de las mejores formas de observar las diferencias entre los lenguajes documentales y las ontologías en el nivel semántico es analizar con detenimiento el concepto de clase, el cual es primordial para entender el funcionamiento de las ontologías, y

también un elemento crucial desde la perspectiva de la representación del conocimiento.

3.1. En torno al concepto de clase

Maniez (1992) lo define, en el caso de los lenguajes documentales como: “un conjunto de objetos (en el sentido amplio de la palabra) que tienen al menos un carácter en común”. Por nuestra parte, según definíamos en un trabajo anterior (**Gil**, 2004), se puede entender como: “grupo de objetos o asuntos que comparten una o más características, identificadas normalmente por una notación específica”. En el ámbito de las ontologías (**Bechhofer**, et. al., 2004) se define como: “un mecanismo de abstracción para agrupar recursos con similares características”.

Sin embargo, al margen de las definiciones del concepto, existen algunas diferencias prácticas; una de ellas hace referencia al hecho de que en el ámbito de las ontologías una clase se define por un conjunto de propiedades, o atributos, que tendrán valores distintos en diferentes instancias de dicha clase y se declararán de forma explícita.

En el ámbito de los lenguajes documentales una clase efectivamente hace referencia a un conjunto de elementos con características similares, pero que no se explicitan en lugar alguno, salvo quizá en forma de notas explicativas, lo que en cualquier caso no es suficiente para su tratamiento automatizado. El problema de esta definición es que no se formaliza, es decir, no se asigna un conjunto de atributos que deben estar presentes en una entidad cualquiera para que pueda ser considerada miembro de una clase. Esto reduce mucho la capacidad de llevar a cabo razonamientos sobre la estructura del lenguaje documental, ya que las clases no se codifican de forma que resulten explícitas para una máquina, por lo que el conocimiento inherente en cada una de las clases queda implícito y no es utilizable.

En otras palabras, es fundamental que una ontología cumpla bien su papel de base de conocimiento para un ámbito en concreto. Para que sea así, debe ser capaz de responder una serie de preguntas a través del conocimiento codificado en ella, denominadas “preguntas de competencia” (**Noy; McGuinness**, 2000), que en el caso de una ontología que modelara una universidad, podrían ser las siguientes:

- ¿Es el decano de una facultad responsable de las infraestructura de redes?
- Las clases de la licenciatura de documentación ¿se imparten en la facultad de documentación o en la de ciencias de la información?
- ¿Dónde debo presentar mis impresos de matriculación?



newspaper Class Hierarchy

- Author
 - Columnist (2 instances)
 - Editor (4 instances)
 - News_Service (2 instances)
 - Reporter (3 instances)
- Content
 - Advertisement
 - Personals_Ad (4 instances)
 - Standard_Ad (1 instance)
 - Article (9 instances)
- Layout_info
 - Billing_Chart (3 instances)
 - Content_Layout
 - Prototype_Newspaper (7 instances)
 - Rectangle
 - Section (8 instances)
- Library (1 instance)
- Newspaper (6 instances)
- Organization (1 instance)
- Person
 - Employee
 - Columnist (2 instances)
 - Editor (4 instances)
 - Manager (3 instances)
 - Director (1 instance)
 - Reporter (3 instances)
 - Salesperson (1 instance)

[^ back to top](#)

Generated: 10/01/2005, 10:57:29 AM, Hora de verano de Europa Central

Protégé is a trademark of Stanford University, Copyright (c) 1998-2005 Stanford University.

Figura 3. Presentación de la jerarquía de clases de una ontología. Fuente: Protégé (Universidad de Stanford)

Sin embargo, la mera existencia de una clase no proporciona la información necesaria para responder ese tipo de preguntas y por lo tanto no podría cumplir adecuadamente su papel. Para que esto sea posible es indispensable definir propiedades para cada una de las clases.

Las clases en un lenguaje documental, por ejemplo, en un sistema de clasificación, sí poseen atributos, ex-

presados en forma de relaciones con otras clases y propiedades con valores determinados. El problema es que dichos atributos se aplican de forma indistinta a todas las clases del sistema de clasificación, o dicho de otra forma, sólo existe una clase formalmente diferenciada en un sistema de clasificación, la “clase concepto”. Miles (2005) defiende la existencia de una “capa de indirectación” entre los lenguajes documentales y las rea-

lidades que modelan. Hace referencia al hecho de que una entrada de un sistema de clasificación o de un tesoro representa un concepto útil para la descripción de documentos, pero no se constituye en el equivalente en la realidad de dicho concepto. Esto se encuentra directamente relacionado con el hecho de que los lenguajes documentales surgen para normalizar e internacionalizar conceptos que a su vez se utilizan para representar entidades del mundo real. Podemos utilizar los ejemplos del propio **Miles** para esclarecer este aspecto.

El grafo de la figura 4 nos muestra cómo el concepto “Enrique VIII”, identificado por el URI: <http://www.example.org/concepts#henry8> (del tipo *skos:Concept*) y los atributos “fecha de creación” y “creador” se pueden aplicar únicamente al concepto generado para representar a la persona de Enrique VIII y no evidentemente a la propia persona de Enrique VIII. Podemos observar cómo se genera de forma natural una clase que describe un tipo de entidades con características comunes (los conceptos de un lenguaje documental). Sin embargo, no se genera una clase para cada una de las entidades del mundo real que vamos a representar, de manera que los atributos de Enrique VIII como instancia de la clase persona no se formalizan y no podemos llevar a cabo ninguna tarea que implique la utilización del concepto Enrique VIII en tanto que persona, sino que tendremos que manejar dicha entidad a través de un mecanismo indirecto, mediante los conceptos que el lenguaje documental utiliza para representarla.

En resumen, podemos decir que la única clase existente en un lenguaje documental, al menos desde el punto de vista de la metodología para crear ontologías, es la clase “concepto”, que tendría a su vez como subclases materia, encabezamiento de materia o descriptor, cada una de las cuáles comparte una serie de características básicas con su superclase.

En el caso de un sistema de clasificación o de una lista de encabezamientos de materias, cada una de las entradas existentes puede ser entendida a su vez como una instancia de una única clase, la de materia, y los únicos atributos serían los de “notación”, nota de alcance, materias genéricas o específicas. Esto supone que no es posible definir atributos específicos para cada una de las instancias de la clase materia que forman un sistema de clasificación ni, por tanto, asimilarlas a las clases de una ontología.

Otra diferencia importante reside en el hecho de que las ontologías modelan entidades, cosas, mientras que los lenguajes documentales trabajan con el nivel léxico en algunos casos y con el nivel conceptual en otros. Esto se materializa en la capacidad de las ontologías para modelar clases de entidades (con sus subcla-

“Una ontología sobre una universidad debería responder preguntas como: ¿Es el decano de una facultad responsable de las infraestructura de redes? o ¿Dónde debo presentar mis impresos de matriculación?”

ses correspondientes) e instancias (ejemplos concretos de uno de esos tipos de entidad) de forma diferenciada, mientras que en el ámbito de los lenguajes documentales esto no es posible.

3.2. Relaciones connotativas

Estas relaciones se sitúan en realidad entre la esfera semántica y la estructural en los lenguajes documentales. Se pueden dar entre los términos que componen los diferentes útiles de representación del conocimiento, que aquí estamos analizando. En el caso de los sistemas de clasificación, la connotación se da de forma particular en aquellos que organizan el conocimiento en base a disciplinas, es decir, por campos del saber, no por materias; con lo que una materia puede figurar en varios lugares de una misma clase y, a su vez, en diferentes clases del esquema clasificatorio. Son ejemplos de esta forma de organización la *Dewey Decimal Classification (DDC)* y la *Clasificación Decimal Universal (CDU)*, siendo típica la organización por materias en la *Library of Congress Classification (LCC)*.

Las relaciones connotativas derivadas de esta característica de algunos sistemas de clasificación, es decir, la posible existencia de un mismo término referido, no obstante, a distintas disciplinas o aspectos, se resuelve en los sistemas de clasificación mediante su representación con una numeración específica y unívoca para cada uno de sus potenciales significados. En los tesauros hablaríamos en este caso de polijerarquía, que se puede resolver mediante notas aclaratorias. Una ontología, por su parte, le puede indicar a una aplicación que dos términos polisémicos no tienen ninguna relación y pertenecen a distintas disciplinas.

Por lo que se refiere a la estructura, la que caracteriza a los tesauros es tradicionalmente arborescente, representada mediante árboles jerárquicos. Las ontologías son flexibles y multidimensionales, en contraste con la estructura jerárquica. Por su parte, los sistemas de clasificación por facetas no tienen la misma estructura arborescente de los modelos tradicionales, y además tienen en cuenta aspectos (facetas) que no son consideradas por aquellos. Por ejemplo, el sistema de clasificación de **Bliss** contempla las disciplinas desde los puntos de vista filosófico, teórico, histórico y prác-

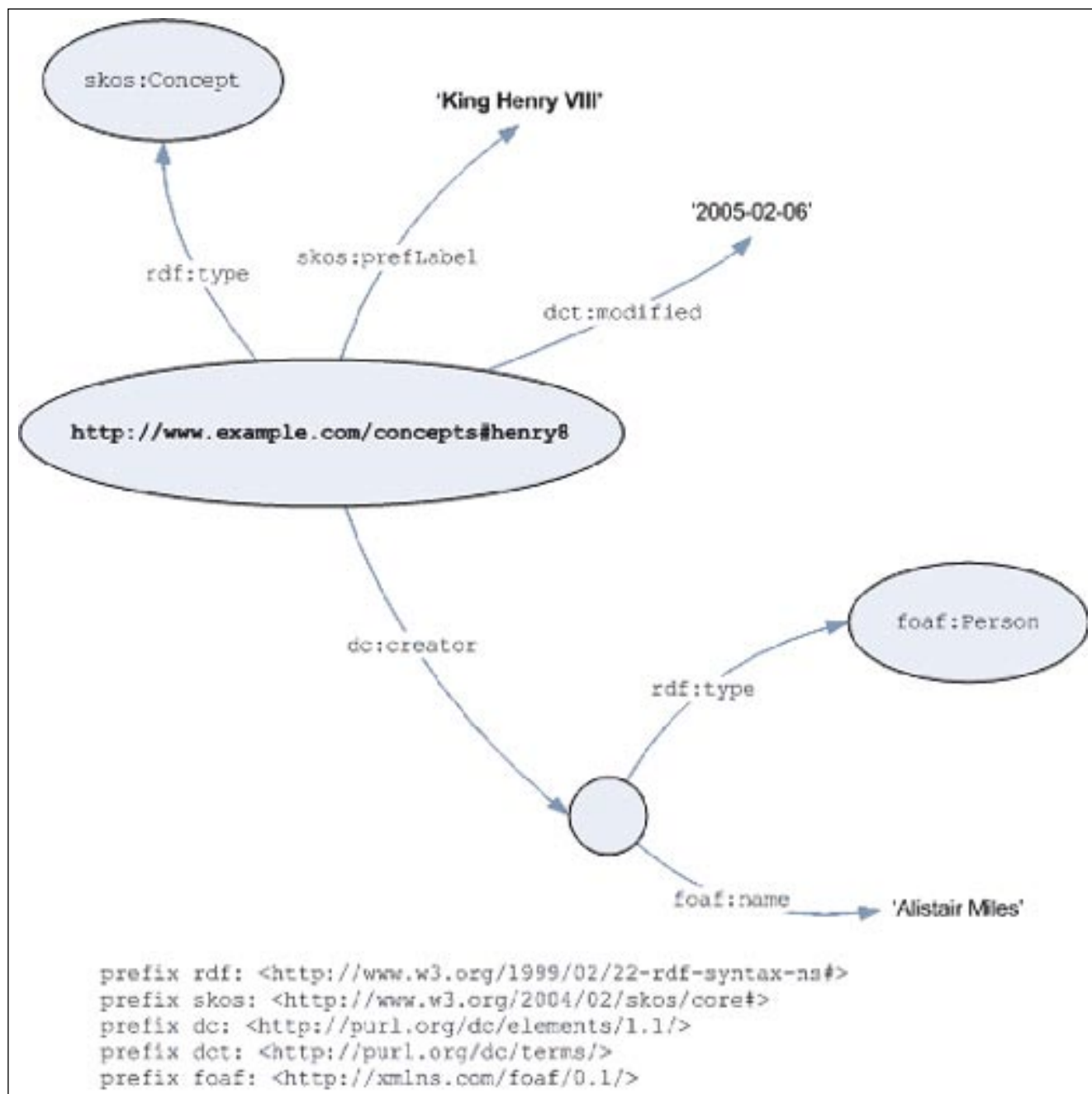


Figura 4. Capa de indirección en los lenguajes documentales. Fuente: SKOS Core Guide

tico; es decir, base teórica de la disciplina, la disciplina en sí misma, su historia y sus aplicaciones, y sustituye la ordenación jerárquica por una gradación de lo simple a lo complejo en las disciplinas y de lo general a lo particular entre los distintos conceptos de una materia.

La *Colon Classification* tampoco es un sistema de clasificación jerárquico como los tradicionales, ya que no contempla un esquema de conocimientos que se va organizando jerárquicamente, sino una “realidad básica”, como afirmaba **Ranganathan**, que se descompone en facetas. Esto es, el sistema consiste en clasificar las materias no sólo por la relación jerárquica de género a especie, sino también por los vínculos existentes entre una cosa y sus partes, las materias que la componen,

sus propiedades, los tratamientos a los que pueden ser sometidos, el espacio y el tiempo. De esta forma, todos los conceptos son susceptibles de caracterizarse por una serie de aspectos o características comunes.

Estos mismos principios son los que inspiran a los tesauros facetados. Algunas de las facetas en las que suelen organizar el conocimiento son: fenómeno, procesos, materiales, propiedades, disciplina, etc. Desde el punto de vista estructural, su utilización proporciona seguramente el punto de conexión más viable entre un lenguaje documental y una ontología que, como veremos en otro punto, se basa en una concepción muy flexible de los tipos de relaciones posibles entre los diversos elementos del sistema.

3.3. Relaciones jerárquicas entre clases

Tienen un papel fundamental a la hora de definir la estructura de los lenguajes documentales. Utilizamos el término relaciones jerárquicas aunque quizá no describe completamente el caso, ya que las que se establecen entre una clase y una subclase en un sistema de clasificación, o entre los términos específicos y genéricos en un tesoro, no son exactamente del mismo tipo que las existentes entre las clases y las subclases de una ontología.

Las clases y las subclases de una ontología se relacionan entre sí mediante un mecanismo de *subsunción*, lo que implica que dada una clase C con una subclase C_i si m es un miembro de C_i también lo es de C . En otras palabras, si la clase “bebida” tiene una sub-clase “vino”, podemos decir que “vino” es una “bebida”. Esto implica que todos los atributos que son propios de C lo son también de C_i , aunque no se cumpla a la inversa, lo que se define comúnmente como un mecanismo de herencia (Corcho; Fernández; Gómez, 2003).

Podemos analizar el ámbito de los lenguajes documentales en torno a dos enfoques, el que es propio de los tesauros y el de las clasificaciones. En el primer caso, las relaciones jerárquicas pueden ser *grosso modo* de tipo partitivo o de generalidad/especificidad. En el primero de los casos se establece que un “término específico” lo es de otro porque constituye una parte de aquél, como en el caso de: *Italia TE Roma*. En el segundo caso “el término genérico se define como aquel descriptor que designa una noción que engloba a otras nociones más específicas representadas por los términos específicos” (Gil, 2004).

La norma *ISO 2788:1986* para la construcción de tesauros monolingües es muy ambigua al respecto de las relaciones jerárquicas, de forma que el significado exacto que ofrece depende del criterio de los expertos que crean cada tesoro en concreto. Esto ha llevado a la definición de relaciones mejor perfiladas por parte de los profesionales pero que habitan fuera del estándar. De esta manera se definen relaciones “*broader term generic*”, que recogen los términos genéricos en general, relaciones “*broader term instantive*”, que tienen en cuenta las relaciones de instancia entre una clase y sus miembros, o “*broader term partitive*”, que recogen las de una entidad y sus partes.

La tabla 1 muestra un extracto del tesoro *Agrovoc* y la reformulación del mismo en forma de ontología que hace Soergel (2004). De esta tabla se deduce fácilmente que la relación existente entre “milk” y “cow milk” no es en absoluto igual a la que existe entre “milk” y “milk fat”. Esto se debe a que el tipo exacto de relación cubierto por los términos específicos se halla implícito. Esto es así porque su explicitación no es ne-

cesaria para un indizador humano. Cuando trabajamos en el ámbito de la indización y recuperación manual, un experto es perfectamente capaz de apreciar de forma directa los matices del tipo de relación y juzgar su idoneidad para cada situación. Sin embargo, no es en modo alguno irrelevante para una máquina que quisiera utilizar provechosamente la relación existente entre ambos conceptos. Y éste es precisamente el enfoque que se intenta lograr en la web semántica.

Agrovoc		Hypothetical ontology	
milk		milk	
NT	cow milk	<i>includesSpecific-</i>	cow milk
NT	milk fat	<i>containsSubstance-</i>	milk fat
cow		cow	
NT	cow milk	<i>hasComponent-</i>	cow milk
cheddar cheese		cheddar cheese	
BT	cow milk	<i>madeFrom-</i>	cow milk

Tabla 1. Relaciones jerárquicas. Comparación entre el tesoro *Agrovoc* y una hipotética ontología. Fuente: Soergel 2004

En el ámbito de los sistemas de clasificación ocurre algo similar. Tomemos por ejemplo la *CDU*; podemos encontrar cómo se produce con cierta frecuencia la relación entre materias y sus divisiones en torno al criterio de pertenencia a una clase, como en “616.91 *Enfermedades infecciosas febriles agudas*” que se divide en “616.912 *Variola*”, “616.913 *Varicela*”, “616.914 *Sarampión*”, etc. Como se puede observar, “*Varicela*” es una instancia de la clase “*Enfermedades infecciosas febriles agudas*” y por tanto un miembro concreto de un grupo de enfermedades con una serie de características en común. No es un tipo de enfermedad, sino una enfermedad específica, por lo que en ningún caso estaríamos hablando de la relación de subsunción existente en el ámbito de las ontologías entre una clase y su subclase, sino de una relación clase/instancia bien distinta.

“La capacidad de precisión de las ontologías es fundamental para el procesamiento automático del conocimiento”

Otro fenómeno común es el de las relaciones “parte de” que se establecen entre muchas materias y sus divisiones, como entre “004.42 *Programación de ordenadores. Programas de ordenador*” y “004.421 *Algo-*

ritmos para la construcción del programa”, “004.422 Componentes de los programas de ordenador”. Son tan sólo algunos ejemplos pero, en cualquier caso, son representativos de las diferencias existentes entre la estructura jerárquica típica de un lenguaje documental y la de una ontología. Esta capacidad de precisión de que hacen gala las ontologías es fundamental para el procesamiento automático del conocimiento.

3.4. Otros tipos de relaciones

Si observamos un tesoro tradicional, podemos comprobar cómo las relaciones asociativas se pueden reformular mediante múltiples tipos de relaciones más expresivas y formalmente correctas.

Eric Thesaurus	Hypothetical ontology
reading instruction	reading instruction
BT instruction	<i>isa-</i> instruction
RT reading	<i>hasDomain-</i> reading
RT learning standards	<i>governedBy-</i> learning standards
reading ability	reading ability
BT ability	<i>isa-</i> ability
RT reading	<i>hasDomain-</i> reading
RT perception	<i>supportedBy-</i> perception

Tabla 2. Relaciones asociativas. Comparación entre el tesoro Eric y una hipotética ontología. Fuente: Soergel 2004

Si consideramos los términos relacionados de “*reading ability*” podemos observar cómo la relación asociativa típica de los tesoros es poco explícita y se podría concretar en otras mucho más específicas y significativas, como “*hasDomain*” o “*supportedBy*”. En realidad los dos tipos de relación propuestos en la ontología no tienen nada que ver entre sí, son de hecho tan distintos como para hacer poco justificable que se agrupen bajo un mismo tipo genérico de relación, la asociativa.

Los tipos de relaciones que se pueden establecer entre las materias de un sistema de clasificación, especialmente en los sistemas de tipo mixto como la CDU, son bastante más expresivos, lo que nos dota de cierta capacidad de combinación y adaptación a las necesidades reales de la clasificación de documentos, aunque siguen siendo limitados, ya que existe un número finito de ellos, que no afecta a la definición de las relaciones entre conceptos en una ontología.

4. Conclusiones

Los lenguajes documentales no pueden utilizarse de la misma forma que las ontologías para tareas de formalización del conocimiento, ni siquiera desde la perspectiva de una funcionalidad reducida. A día de

hoy no pueden emplearse como una base de conocimiento, porque no han sido construidos de forma que este conocimiento resulte explícito para su tratamiento automatizado.

Desde el punto de vista de la reutilización de los lenguajes documentales hemos de decir que las diferencias entre éstos y las ontologías complican la posibilidad de trasladar el conocimiento existente en un tesoro o un sistema de clasificación a una ontología. Reducir los tipos de relaciones que se establecen en los lenguajes controlados tradicionales haría de la ontología resultante un sistema inoperante para las tareas de razonamiento deductivo. Explicitar los tipos de relaciones es una tarea que, como mucho, se puede semi-automatizar, pero que en cualquier caso tendría un coste elevado en términos de tiempo y esfuerzo.

“Las diferencias entre lenguajes documentales y ontologías complican el traslado del conocimiento existente en un tesoro o un sistema de clasificación a una ontología”

Por otra parte, los lenguajes documentales pueden servir como base para el desarrollo de ontologías, como se ha hecho en varios de los artículos mencionados, pero tienen un papel limitado en este sentido. Trazando un símil con el proceso de creación de un tesoro, podemos comparar la eficacia que tendrían los lenguajes documentales para la creación de ontologías con la utilidad que tienen los glosarios o listados de términos utilizados durante la fase de extracción de terminología. Los lenguajes documentales existentes proporcionan un acercamiento a los conceptos típicos de un dominio, pero tendrían un papel mucho más limitado a la hora de construir la estructura, y sobre todo para la formalización de las clases y el establecimiento de reglas y axiomas formales.

Parece más factible que los lenguajes documentales puedan adaptarse a los requerimientos de una ontología para la recuperación de información. Se han llevado a cabo numerosos estudios sobre la utilidad de este tipo de ontologías (Bhagal, et. al., 2007; Paralic; Kostial, 2003; Vallet, et. al., 2005) aunque los resultados no son concluyentes.

En cualquier caso nosotros nos hemos centrado en una visión sobre las ontologías para su utilización en el ámbito de la web semántica, lo que implica capacidad para expresar formalmente el significado de las cosas y la posibilidad de reutilización del conocimiento en

entornos diferentes al original. En este contexto los lenguajes documentales están muy alejados de los requerimientos formales de una ontología señalados por el W3C, como se ha expuesto con anterioridad. Por estos motivos creemos que la reutilización de los lenguajes documentales dentro del ámbito de la web semántica requerirá bastantes esfuerzos por parte de la comunidad de profesionales y expertos en lenguajes documentales.

Notas

1. En el original: “*the semantic web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation*”.
2. Véase: **McBride, Brian** (2004) para una explicación en profundidad de éstos y otros aspectos básicos de rdf.
3. En el original: “*a formal explicit description of concepts in a domain of discourse*”.
4. En el original: “*an abstract, simplified view of the world that we wish to represent for some purpose*”. “*A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose*”.
5. Algunos sitios web de referencia como *Amazon* o *Yahoo* hacen uso de ontologías como forma de gestión del conocimiento, una práctica que parece estar extendiéndose paulatinamente.

Bibliografía

- Bechhofer, Sean; Van Harmelen, Frank; Hendler, Jim; Horrocks, Ian; McGuinness, Deborah L.; Patel-Schneider, Peter F.; Stein, Lynn Andrea; Olin, Franklin W.** *OWL Web ontology language. Recomendación W3C*, febrero 2004. Consultado en: 12-02-06.
<http://www.w3.org/TR/owl-ref/>
- Berners-Lee, Tim.** *Transcripción de la conferencia en la conmemoración del 35 aniversario del Instituto Tecnológico de Massachusetts el 14 de abril de 1999.*
<http://www.w3.org/1999/04/13-tbl.html>
- Bhagal, Jagdev; MacFarlane, Andrew; Smith, Peter.** “A review of ontology based query expansion”. En: *Information processing & management*, 2007, julio, v. 43, n. 4, pp. 866-886.
- Corcho, Óscar; Fernández-López, Mariano; Gómez-Pérez, Asunción.** “Methodologies, tools and languages for building ontologies. Where is their meeting point?”. En: *Data and knowledge engineering*, 2003, n. 46, pp. 41-64.
- García-Jiménez, Antonio.** “Instrumentos de representación del conocimiento: tesauros versus ontologías”. En: *Anales de documentación*, 2004, n. 7, pp. 79-95.
- Genesereth, Michael J.; Nilsson, Nils J.** *Logical foundations of artificial intelligence*. San Francisco: Morgan Kaufmann, 1987.
- Gil-Urdiciain, Blanca.** *Manual de lenguajes documentales*. Gijón: TREA, 2004.

Gruber, Thomas R. “Towards principles for the design of ontologies used for knowledge sharing”. En: **Guarino, N.; Poli, R.** (eds.). *Formal ontology in conceptual analysis and knowledge representation*. Deventer: Kluwer Academic Publishers, 1993.

Guzmán-Luna, Jaime; Torres-Pardo, Durley; López-García, Alba-Nubia. “Desarrollo de una ontología en el contexto de la web semántica a partir de un tesoro documental tradicional”. En: *Revista interamericana de bibliotecología*, 2006, julio-diciembre, v. 29, n. 2, p. 79-94.

Heflin, Jeff. *OWL Web ontology language use cases and requirements. Recomendación W3C*, febrero 2004.
<http://www.w3.org/TR/webont-req/>

Maniez, Jacques. *Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales*. Madrid: Fundación Germán Sánchez Ruipérez, 1992, p. 22.

McBride, Brian. *Rdf primer. Recomendación W3C*, febrero 2004.
<http://www.w3.org/TR/rdf-primer/>

McGuinness, Deborah L.; Van Harmelen, Frank (eds.). *OWL Web ontology language overview. Recomendación W3C*, febrero 2004.
<http://www.w3.org/TR/owl-features/>

Miles, Alistair. *Modelling thesauri for the semantic web, W3C*, 2001.
<http://www.w3c.rl.ac.uk/SWAD/thesaurus/tif/deliv81/final.html#sec-owl>

Miles, Alistair. *SKOS core guide. W3C Working Draft*, mayo 2005.
<http://www.w3.org/TR/swbp-skos-core-guide/>

Miller, Eric. *Semantic web activity statement. W3C*, noviembre 2005.
<http://www.w3.org/2001/sw/Activity>

Noy, Natalya F.; McGuinness, Deborah L. *Ontology development 101: a guide to creating your first ontology*. Informe técnico. Universidad de Stanford, 2000.
<http://protege.stanford.edu/ontology101-noy-mcguinness.html>

Paralic, Jan; Kostial, Ivan. “Ontology-based information retrieval” En: *Proceedings of the 14th International conference on information and intelligent systems, IIS 2003*, pp. 23-28.

Pérez-Agüera, José-Ramón. *Generación automática de tesauros documentales*. Trabajo para la obtención del diploma de estudios avanzados. Dirigido por **Lourdes Araujo**. Universidad Complutense de Madrid, septiembre de 2005, pp. 107-112.

Soergel, Dagobert; Liang, Anita; Lauser, Boris; Fisseha, Frehiwot; Keizer, Johannes; Katz, Stephen. “Reengineering thesauri for new applications: the Agrovoc example”. En: *Journal of digital information*, 2004, v. 4, n. 4.

Vallet, David; Fernández, Miriam; Castells, Pablo. “An ontology-based information retrieval model”. En: *Proceedings of the Second European semantic web conference, ESWC 2005*, 2005, pp. 455-470.

Rodrigo Sánchez-Jiménez; Blanca Gil-Urdiciain, *Facultad de Ciencias de la Información, Universidad Complutense de Madrid.*
rsanchezj@ccinf.ucm.es
blanca@caelo.eubd.ucm.es

Suscripción EPI sólo online

Pensando sobre todo en los posibles suscriptores latinoamericanos, ya no es obligatorio pagar la suscripción impresa de EPI para acceder a la online.

EPI se ofrece a instituciones en suscripción “sólo online” a un precio considerablemente más reducido (85 euros/año), puesto que en esta modalidad no hay que cubrir los gastos de imprenta ni de correo postal.



OvidSP

Piensa rápido Busca velozmente

Ovid es la plataforma de búsqueda profesional más usada en el mundo. Ahora la mejor plataforma es todavía mejor.

Les presentamos OvidSP: la nueva herramienta de búsqueda y descubrimiento de Ovid, que convierte a Ovid Gateway y a SilverPlatter en una experiencia de búsqueda más sencilla y precisa. Mejor integración en el flujo de trabajo del usuario. Una interfaz más rápida y más intuitiva. Respaldada por la excelente tecnología de búsqueda de precisión de Ovid. Ahora, ya es mucho más fácil para los investigadores llegar a dónde quieren ir.

Potente. Simplificada. Mucho mejor.

Para más información
spain@ovid.com, o 91 4186275

www.ovid.com