

Reutilización de tesauros: el documentalista frente al reto de la web semántica

Por Alberto García-Torres, Antonio Pareja-Lora y Daniel Pradana-López

Resumen: Se presenta una aproximación práctica a la construcción de ontologías a partir de tesauros inspirada en la metodología Methontology, que permite hacer uso de los lenguajes propios de la definición de ontologías y anotación semántica basada en las mismas. Asimismo es una alternativa a la publicación de tesauros en la web semántica mediante SKOS-Core, que es la propuesta del World Wide Web Consortium (W3C), y ofrece más semántica que el tesoro de partida y que su implementación mediante SKOS-Core. También se detalla una arquitectura genérica para la representación de dominios con dos niveles de descripción: una ontología de representación del conocimiento, basada en las facetas de Ranganathan (1967), para representar las relaciones de alto nivel del dominio, más una ontología de dominio, que proporciona la descripción de un campo particular del dominio.



Alberto García Torres es licenciado en biblioteconomía y documentación por la Universidad Complutense de Madrid. En la actualidad compagina los estudios de doctorado en ciencias y técnicas historiográficas con un proyecto de investigación en documentación histórica en el Tribunal Supremo. Sus líneas de investigación actuales van dirigidas hacia la archivística, la informatización de los archivos y la diplomática moderna.



Antonio Pareja Lora es licenciado en informática por la Univ. Politécnica de Madrid (UPM), donde cursa el doctorado en CC de la computación e inteligencia artificial. Es especialista en procesamiento del lenguaje natural y participa en el Ontological engineering group (OEG) de la UPM. Es profesor colaborador en el Depto. de Sist. Informáticos y Computación de la UPM. Sus líneas de investigación se encaminan a la representación del conocimiento lingüístico mediante ontologías y su aplicación a la notación semántica de textos.



Daniel Pradana López es licenciado en documentación por la Universidad Complutense de Madrid y diplomado en óptica y optometría por la misma universidad. En la actualidad está realizando estudios de doctorado en documentación mientras trabaja como especialista en anomalías visuales en su propia óptica. Sus líneas de investigación van dirigidas en estos momentos a la creación de tesauros y ontologías relacionadas con la optometría.

Palabras clave: Ontologías, Tesauros, Metodologías de desarrollo, Methontology, Arquitecturas de software, Web semántica.

Title: Reusing thesauri: documentalists face the semantic web challenge

Abstract: A practical approach to the development of ontologies from thesauri is presented, using the Methontology methodology, which permits the use of any language developed so far for the implementation of ontologies, as well as semantic web annotations based on these ontologies. This approach is an alternative to SKOS-Core, the World Wide Web Consortium (W3C) approach to publication of thesauri within the semantic web. The advantage of our approach is that it provides significantly more semantic information than a thesaurus (and, hence, than does SKOS-Core). A two-level ontological architecture for domain representation is also described, consisting of a knowledge representation ontology [based on Ranganathan's (1967) facets] for the representation of high-level relationships in the domain and a domain ontology for the description of a particular field within the domain.

Keywords: Ontologies, Thesauri, Software methodologies, Methontology, Software architectures, Semantic web.

García Torres, Alberto; Pareja Lora, Antonio; Pradana López, Daniel. "Reutilización de tesauros: el documentalista frente al reto de la web semántica". En: *El profesional de la información*, 2008, enero-febrero, v. 17, n. 1, pp. 8-21.

DOI: 10.3145/epi.2008.ene.02

Introducción

La *www* ha sido y es una gran fuente de información accesible mediante nuestros ordenadores pero hasta ahora sólo comprensible para los seres humanos. En un principio los documentos de la Red nacieron orientados al intercambio informativo entre personas. Estos contenidos incluyen grandes cantidades de información (multimedia, en muchos casos) incomprensible para una computadora, siendo el usuario el principal encargado de encontrar, extraer e interpretar la información relevante.

Actualmente, dado el asombroso crecimiento de la web, resulta imposible que un único usuario realice esas tareas en un tiempo aceptable. Por otra parte, aunque se está facilitando el procesamiento automático de la información, los ordenadores todavía no pueden llevar a cabo con una precisión aceptable las tareas ya mencionadas de acceso, extracción e interpretación de la información relevante sin ayuda, pues no alcanzan el nivel de comprensión del lenguaje que poseen los seres humanos. Precisamente por ello, conseguir que los ordenadores entiendan el significado (la semántica) de los textos escritos y de las páginas web es el motor principal que impulsa el desarrollo de la web semántica (Berners-Lee, et al., 1999).

Existen dos elementos clave para lograr que un ordenador entienda el significado de un documento de la web. Por un lado, las ontologías (especificaciones formales y explícitas de una conceptualización compartida –Gruber, 1993; Borst, 1997; Studer, et al., 1998–), que establecen los términos relevantes del dominio (conceptos, atributos, relaciones, instancias, etc.); por otro lado, la anotación semántica de textos (Benjamins, et al., 1999; Motta, et al., 1999; Luke, et al., 2000; Staab, et al., 2000; Aguado, et al., 2002 (a; b; c); Benjamins, et al., 2003; Wilcock, et al., 2004), que hace explícito para un ordenador el significado de estos términos cuando aparecen en un texto, enlazán-

dolos con sus asociados dentro de las ontologías consideradas en la anotación, mediante los mecanismos de referencia apropiados de cualquiera de los lenguajes propios de la web semántica: *xml*, *rdfs*, *OWL*, etc. (Gómez Pérez; Corcho, 2002; Dean; Schreiber, 2003; Peis, et al., 2003).

“La reutilización de tesauros es un interesante punto de partida para la construcción de una ontología”

Por lo tanto, cada vez más son necesarias ontologías que representen el conocimiento de los distintos dominios que componen la realidad, y que sirvan de base para la anotación de documentos y recursos de la web semántica. Puesto que partir de cero en la construcción de semejante red ontológica es a priori inviable, dado el alto coste que supone el desarrollo de una ontología, se hace necesario reutilizar todos los recursos disponibles. La reutilización de tesauros es pues un interesante punto de partida para la construcción de una ontología, que reduciría al menos los costes del proceso de educación de conocimientos, así como gran parte del esfuerzo de consenso que requiere una ontología. De hecho, según la clasificación de Lassila y McGuinness (2001), un tesoro se puede considerar ya como una ontología ligera (*lightweight ontology*, en inglés). Esta clasificación se recoge en la figura 1, tomada y traducida de Gómez Pérez, Fernández López y Corcho (2004), donde se ven claramente las características que deben añadirse a un tesoro para convertirlo en una ontología de peso (*heavyweight ontology*).

En el presente artículo presentamos una aproximación práctica a la construcción de ontologías a partir de tesauros inspirada en la metodología *Methontology* (Gómez Pérez; Fernández López; Corcho, 2004), que permite hacer uso de los lenguajes propios de la definición de ontologías y anotación semántica basada en las mismas, como alternativa a la publicación de tesauros en la web semántica mediante la propuesta del W3C, que es *SKOS-Core*. Como se muestra más adelante, nuestra propuesta proporciona más semántica que el tesoro de partida y que su implementación mediante *SKOS-Core*. Al final del artículo se proporciona el resultado de la aplicación de dicha

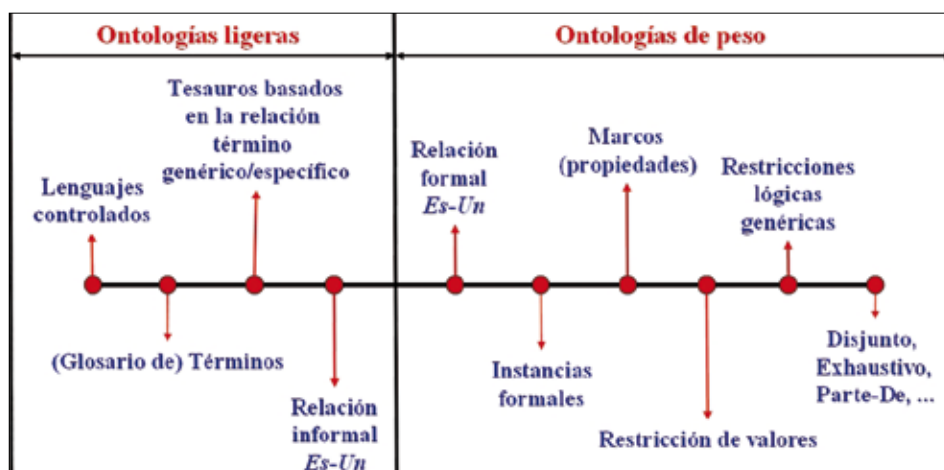


Figura 1. Clasificación de Lassila y McGuinness

metodología a un subdominio de la óptica. Asimismo, partiendo de un tesoro sencillo de dicho dominio, se detalla una arquitectura genérica en dos niveles de descripción: una ontología de representación del conocimiento para identificar dominios, basada en las facetas de **Ranganathan** (1967) –véase también **Gil Urdiciain** (1994)-, instanciada a modo de ejemplo para el dominio de la óptica, más una ontología que proporciona la descripción del tesoro en el campo de las anomalías oculares.

El resto del artículo está organizado como sigue: primeramente se presenta un breve estudio del estado de la cuestión en metodologías, herramientas y lenguajes de desarrollo de ontologías, así como de las propuestas formuladas hasta la fecha para la reutilización de tesoros en la web semántica; a continuación, se lleva a cabo un estudio comparativo de tesoros y ontologías, con sus diferencias y similitudes, así como las ventajas de las unas frente a los otros; seguidamente, se describe cómo aplicar la metodología *Methontology* al desarrollo de ontologías a partir de tesoros; posteriormente, se propone la arquitectura prototipo para la transformación de tesoros de un dominio en ontologías. Finalmente, se aportan unas conclusiones junto con las referencias bibliográficas empleadas.

Estado de la cuestión

En el desarrollo de ontologías es de gran importancia la elección de un lenguaje apropiado con el que implementarlas (**Gómez Pérez; Corcho**, 2002; **Dean; Schreiber**, 2003; **Peis**, et al., 2003), pero éste es tan sólo uno de los pasos de un proceso complejo, que requiere de la definición y aplicación de una metodología propia de desarrollo, para evitar malgastar tiempo y esfuerzos.

Una metodología de desarrollo de ontologías es un sistema detallado de pasos que deben seguirse durante el proceso de construcción de una ontología, en un determinado orden, desde el momento de su concepción, a partir de las fuentes de las que se dispone para su construcción, hasta que queda finalmente implementada; idealmente debe contemplar también cómo se va a actualizar o evolucionar con el tiempo. Existen distintos tipos de metodologías de este estilo, en función de las fuentes de partida (**Gómez Pérez; Fernández López; Corcho**, 2004): las que describen cómo construir una ontología desde cero, como la ideada por **Grüniger y Fox**, *On-To-Knowledge*, *Methontology*, etc.; otras basadas en procesos de re-ingeniería de otras ontologías implementadas previamente, para adaptarlas a un nuevo entorno de aplicación (*Methontology*, por ejemplo, ha sido extendida para incluir esta posibilidad); teniendo en cuenta técnicas de aprendizaje automático (*ontology learning*); y también asentadas en la fusión de otras

ya existentes para el dominio (similares, pero distintas) en una unificada, que implica la determinación de la correspondencia de los términos de todas las consideradas (*ontology mapping*). Estos últimos puntos mencionados (el aprendizaje automático y la comparación de ontologías) centran gran parte de la atención hoy en día en la investigación en ingeniería ontológica, junto con la evaluación de ontologías, que también es parte del proceso de desarrollo de las mismas.

“Una metodología de desarrollo de ontologías es un sistema detallado de pasos que deben seguirse en un determinado orden, desde el momento de su concepción hasta que queda finalmente implementada”

Otra de las decisiones importantes que deben tomarse en el desarrollo de una ontología es si se desea utilizar una herramienta de las existentes para este fin o no. En caso negativo, esto requeriría generar a mano por completo el código xml, *Rdfs*, *OWL*, etc. de la ontología. En caso afirmativo, la herramienta se encargará de generar el código de implementación a partir de una representación gráfica y abstracta de la ontología, lo cual es bastante más cómodo y robusto. Entre las aplicaciones existentes, algunas restringen el lenguaje de implementación en el que quedará escrita la ontología (es decir, son dependientes del lenguaje), como *OntoSaurus*, *WebOnto* y *OilEd*, y otras permiten crear una ontología independientemente del lenguaje final de implementación, como *Protégé*, *OntoEdit*, *KAON* y *WebODE*. Esta última (*WebODE*) implementa la metodología *Methontology*, lo cual es una de sus grandes ventajas (**Gómez Pérez; Fernández López; Corcho**, 2004). En la figura 2 puede contemplarse un resumen gráfico de lo hasta aquí expuesto.

<http://webode.dia.fi.upm.es>

La principal alternativa a la creación de ontologías para la estructuración del conocimiento en la web semántica mediante anotaciones, es la iniciativa *SKOS* (*Simple Knowledge Organisation System*) del *W3C*. Es una concreción en lenguaje *Rdfs* de las etiquetas y relaciones que caracterizan un tesoro, útil para la publicación de los ya desarrollados, tras su implementación en el lenguaje *SKOS-Core* y/o haciendo uso de su extensión, *SKOS-Extended*. Para realizar esta codificación automáticamente existen al menos dos herramientas disponibles: el importador *SKOS* (**Pérez Agüera**, 2004), que permite convertir tesoros codificados en texto plano mediante marcados estándar (-, >, =>, etc.) en *SKOS-Core*; y la aplicación *Tematres* (**Ferreira**,

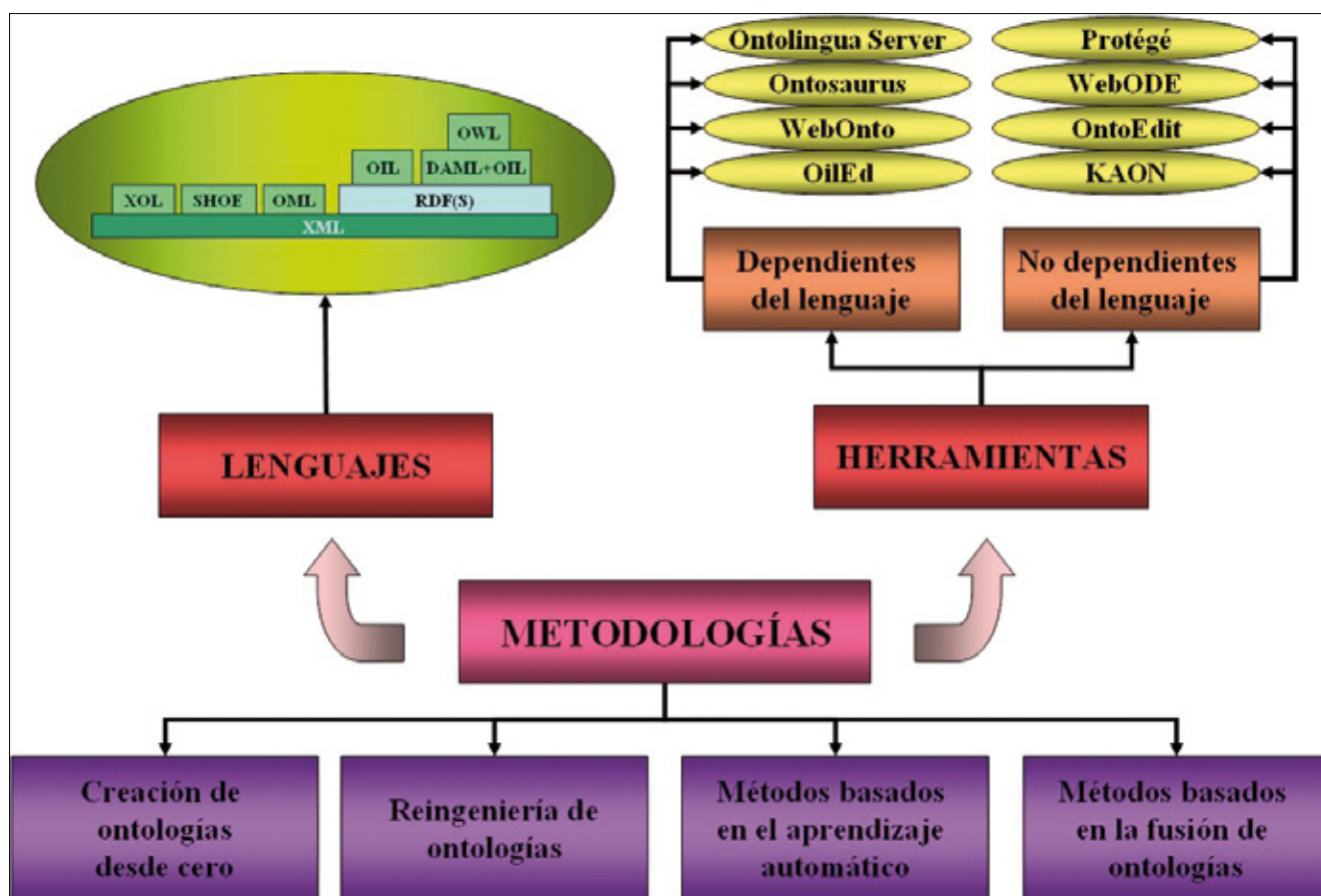


Figura 2. Desarrollo de ontologías. Estado de la cuestión

2007), que proporciona una implementación en *SKOS-Core* del tesoro al completo. La tarea de generar *SKOS* se puede llevar a cabo de forma automática si desarrollamos el tesoro con la propia aplicación, aunque también se puede importar en html, xml u otros formatos a su base de datos y, a partir de ella, generar el *SKOS-Core*. Sin embargo, el nivel de descripción semántica de un tesoro (aun implementado en *SKOS*), comparado con el de una ontología, es muy limitado, por lo que la formalización del conocimiento como ontología y no como tesoro aporta una serie de beneficios adicionales reseñados en el estudio comparativo de la sección siguiente.

<http://www.w3.org/TR/swbp-skos-core-guide>

<http://www.w3.org/2004/02/skos/extensions.rdf>

Es por eso por lo que en este artículo se presenta una concreción de *Methontology* para la construcción de ontologías a partir de tesauros, de forma que puedan ser reutilizados en el contexto de la web semántica.

“El nivel de descripción semántica de un tesoro, comparado con el de una ontología, es muy limitado”

Las ventajas son evidentes: se aprovecha el esfuerzo y el tiempo empleados en el desarrollo del tesoro; la mayor expresividad semántica de las ontologías deja abiertas las puertas para un uso más extensivo; hay ya disponibles herramientas de desarrollo (incluyendo su implementación) de ontologías, lo que facilita esta engorrosa tarea; y sólo hay que adaptar alguna metodología de elaboración de ontologías existente cuando el punto de partida sea un tesoro, para convertirla en una tarea sistemática y robusta.

Tesauros y ontologías: un estudio comparativo

Las similitudes existentes entre ambos sistemas hacen que la conversión de unos a otros sea sencilla, siempre y cuando se tengan en cuenta las diferencias existentes entre ellos (véase la tabla 1).

En primer lugar, las instancias, que son cada uno de los objetos particulares que se incluyen dentro del dominio que estamos modelando, son elementos básicos en ontologías que no poseen correspondencia en los tesauros. El resto de elementos de una ontología pueden hallarse especificados unas veces, e implícitos (camuflados) otras, entre los distintos componentes de un tesoro.

Las siguientes entradas de la tabla 1 son los descriptores y los no descriptores. En *Methontology* (y, por ende, en *WebODE*) la información de cada concepto en una ontología incluye los campos: nombre del concepto, sus sinónimos, acrónimos, instancias, atributos de

clase y atributos de instancia, más las relaciones en las que participa. Los descriptores (o términos preferentes en *SKOS-Core*) son implementados dentro de una ontología, en general, como el nombre del concepto o de la clase. En el caso excepcional de que el descriptor fuera

Tesauros	Definición	Ejemplos	Ontologías	Definición	
-	-	-	Instancia	Objeto particular del dominio que se está formalizando	
Descriptor (SKOS: término preferente)	Término representativo de un concepto	Globo ocular	Nombre del concepto (clase)	Identificador del concepto o clase dentro de la ontología	
			Acrónimo (en caso de que se decida elegir como descriptor el desarrollo de unas siglas)	Siglas del identificador del concepto en caso de existir	
No descriptor (SKOS: término no preferente)	Término no representativo de un concepto (sinónimos, cuasi-sinónimos, etc.)	Ojo	Sinónimo	Término alternativo para referirse al concepto	
			Nombre del concepto (si el descriptor es un acrónimo)	Véase más arriba	
Nota aclaratoria (NA)	Nota utilizada para la contextualización de cada término	Anomalía ocular NA Úsese en caso de referirse a aspectos relacionados con anomalías de refracción o estructurales	Descripción	Nota utilizada para la contextualización de cada término (definición en lenguaje natural, etc.)	
			Nombre del concepto (desarrollo de un acrónimo, en el caso de que se elija éste como descriptor)	Véase más arriba	
			Atributo	de clase	Propiedad que adopta un mismo valor para todas las instancias de una misma clase
				de instancia	Propiedad particular para cada una de las instancias incluidas en una determinada clase
			Definición del tipo de datos de atributos	El tipo de datos de atributos es el conjunto de valores que puede adoptar una propiedad (atributo) de una clase o concepto (entero, real, booleano, cadena de caracteres, etc.)	
			Restricción del rango de valores de atributos	El rango de valores que puede adoptar una propiedad puede limitarse en función del dominio	
			Regla	Afirmación de la lógica de primer orden del tipo $P \rightarrow Q$ que debe entenderse como: si se verifica P, entonces se verifica Q; o bien, regla de producción de la forma $P_1 \dots P_n \Rightarrow A_1, \dots, A_n$ donde las premisas P_i son condiciones que deben verificarse simultáneamente y las A_i son acciones que se realizan si se verifican dichas premisas	
Axioma formal	Afirmación siempre cierta dentro del dominio, expresada como una fórmula de la lógica de primer orden				

Tabla 1. Comparativa entre los elementos de un tesauro y de una ontología

Tesauros		Definición	Ejemplos	Ontologías	Definición
Relación de equivalencia	Usado por (UP)	La relación usado por se establece desde el descriptor al no descriptor mientras que use se establece desde el no descriptor al descriptor	Globo ocular	Methontology: representaciones intermedias (tareas 1, 6, 7, 9 y 10)	
	Use (USE)		UP ojo Ojo USE Globo ocular		
Relación jerárquica	Término genérico (TG)	Hiperónimo de otros conceptos	CÓRNEA TG Globo ocular	Relación IsA	El concepto C1 es una subclase de otro concepto C2 si y sólo si todas las instancias de C1 lo son también de C2
			Miopía magna TG Miopía	Descomposición disjunta	División de un concepto C en subclases, de forma que éstas no tienen instancias en común; puede haber instancia de C que no pertenezcan a ninguno de los conceptos de la descomposición
	Término específico (TE)	Cada uno de los hipónimos de otro concepto	Globo ocular TE Córnea	Descomposición exhaustiva	División de un concepto C en subclases, de forma que todas las instancias deben pertenecer al menos a una de las subclases de C, pero puede haber instancias que pertenezcan a más de una subclase
			Miopía TE Miopía magna	Partición	División de un concepto C en subclases, de forma que todas las instancias deben pertenecer a una y sólo una subclase de C
			Relación PartOf	Relación metonímica del todo a cada una de sus partes a un todo	
Relación asociativa	Término relacionado (TR)	Relacionan descriptores que no suponen sinonimia ni jerarquía; se trata de una relación de afinidad semántica	Miopía TR Globo ocular	Relaciones ad hoc	Relaciones que se establecen entre conceptos, pero que no se ajustan a ninguno de los tipos de relaciones jerárquicas anteriores
				Methontology: representaciones intermedias (tareas 1, 6, 7, 9 y 10)	

Tabla 1 (continuación). Comparativa entre los elementos de un tesoro y de una ontología

un acrónimo será reflejado en la ontología en el campo acrónimo de un concepto o de una clase.

Los no descriptores (o términos no preferentes en *SKOS-Core*) se implementarán en las ontologías dentro de los campos de descripción citados en la tabla 1. Es decir, en el caso de que el no descriptor fuera sinónimo de un descriptor, el primero será incluido en el campo de sinónimos del concepto que representa al descriptor en la ontología. En el caso de que el no descriptor fuera el desarrollo de un acrónimo, éste se verá reflejado en la ontología como el nombre del concepto.

Las notas aclaratorias utilizadas en los tesauros pueden implementarse en las ontologías de diferentes formas, atendiendo a las funciones que cumplen dentro del tesoro:

1. Si es utilizada para contextualizar el descriptor a través de una definición en lenguaje natural, quedará

reflejada en la ontología dentro de la descripción asociada al concepto (que puede verse en el glosario de términos de la misma).

2. En el caso de que sea el desarrollo de un acrónimo, como ya vimos más arriba, se incluirá en la ontología precisamente como el nombre del concepto.

3. Si se refiere a las restricciones de aplicación o de uso de un descriptor, se plasmará en la ontología a través de:

3.a. La definición de un atributo propio de un concepto. Los atributos son las propiedades o características que identifican cada concepto o instancia. Existen dos tipos de atributos: si un valor del atributo caracteriza a un grupo de instancias (lo que conocemos como clase), se trata de un atributo de clase; por el contrario, si el valor del atributo es propio de cada instancia, se trata de un atributo de instancia.

3.b. La definición del tipo de un(os) cierto(s) atributo(s) y/o,

3.c. La restricción del rango de valores determinada por el tipo de esos u otros atributos.

4. Si la nota se refiere a relaciones entre valores de atributos de uno o varios conceptos, por ejemplo, se verán implementadas en una ontología como reglas. Por otro lado, es deseable que una nota aclaratoria que contiene una definición en lenguaje natural del término se traduzca en la medida de lo posible en un axioma, pues en ellos y en las reglas reside en gran parte la potencia razonadora de los mecanismos de inferencia basados en ontologías.

Las relaciones de equivalencia que existen en el tesoro entre los descriptores y los no descriptores no ven reflejadas directamente en las ontologías, sino que se recogen en el glosario inicial de la metodología de construcción *Methontology*, por ejemplo, donde se establecen sinónimos y acrónimos de cada uno de los términos admitidos. También pueden recogerse en el resto de las representaciones intermedias en las que se dan los atributos de clase, de instancia y en la definición de los propios axiomas y reglas especiales.

Un caso mucho más interesante son las relaciones jerárquicas del tesoro, que se establecen entre los términos más generales y los más específicos del dominio. En lo que respecta a las ontologías, existen cinco tipos de relaciones para concretar estas relaciones, en función de la semántica con la que se usen dentro del tesoro: la relación *IsA* (o *SubClaseDe*), la descomposición exhaustiva, la descomposición disjunta, la partición y la relación *PartOf* (*Parte-Todo*), todas ellas explicadas más adelante, en el apartado dedicado a la descripción de la metodología.

Por último, en los tesoros tenemos relaciones asociativas entre descriptores que expresan cierta afinidad semántica entre conceptos distintos. En el caso de las ontologías se concretarían como relaciones ad hoc entre términos afines. No obstante, algunas de las funciones de las relaciones asociativas de **Aitchison, Gilchrist** y **Bawden** (1997) como, por ejemplo, aquella que se recoge en una ontología como la relación entre un concepto y sus atributos, se verán reflejadas en el glosario inicial y en las representaciones intermedias resultado de las tareas en las que se detallan los atributos de clase y de instancia, los axiomas y las reglas (explicadas asimismo en la metodología).

A partir de todo lo expuesto anteriormente, puede deducirse fácilmente que la formalización del conocimiento como una ontología y no como tesoro aporta los seis beneficios adicionales siguientes:

– En la creación de una ontología se utilizan modelos matemáticos, lo que la dota de mayor precisión frente al tesoro.

– A diferencia de los tesoros, se pueden añadir elementos concretos denominados instancias, que aportan una información esencial para la extracción de información (por ejemplo, las personas concretas que realizan una acción, con su lugar y fecha de realización).

– Se obtiene un desarrollo semántico más profundo (**Ding; Foo**, 2002; **García-Jiménez**, 2004), pues los términos son desarrollados en toda su extensión, proveyéndolos de atributos (propiedades) y también se formalizan y detallan más aún las relaciones jerárquicas o asociativas de un tesoro, lo que provoca que un concepto esté mucho más definido dentro de la globalidad del vocabulario.

– El conocimiento formalizado puede reutilizarse dentro de un dominio, por definición, al ser previamente consensuado, y se posibilita su incorporación en distintas aplicaciones. Es decir, al describir objetos del mundo reales, y no sólo conceptos o ideas, éstos pueden ser incluidos en diversos sistemas a la vez (**Qin; Paling**, 2002; **García-Jiménez**, 2004).

– Por ser consensuada, una ontología favorece una representación normalizada de la información (**Ding; Foo**, 2002; **García-Jiménez**, 2004).

– La utilización de las ontologías permite la comunicación entre humanos y máquinas (así como entre las propias máquinas o agentes entre sí –**Ding; Foo**, 2002; **García-Jiménez**, 2004-); además, la inclusión de axiomas y reglas allana el camino para la concreción y aplicación de mecanismos de inferencia y razonamiento de la web semántica basados en ontologías, lo cual a su vez permitirá mejorar la precisión de los mecanismos de extracción y recuperación de información.

Aplicación de *Methontology* al desarrollo de ontologías a partir de tesoros

El proceso de construcción de una ontología puede verse simplificado mediante la utilización de la metodología apropiada. Existen varias aplicables, pero reco-

“La utilización de ontologías permite la comunicación entre humanos y máquinas, así como entre las propias máquinas o agentes entre sí”

mendamos tanto por su sencillez como por su probada eficacia el uso de *Methontology* (Gómez-Pérez, Fernández-López, Corcho, 2004). Consta de once pasos (o tareas) que se han de realizar de forma secuencial, y que son descritos a continuación ilustrándolos mediante ejemplos tomados de un tesauro en el dominio de la óptica y la optometría actualmente en construcción. Este tesauro hace un uso abundante de las notas aclaratorias con toda su capacidad expresiva, según Aitchison, Gilchrist y Bawden (1997), de forma que todo el conocimiento que no es capaz de captar quede en cierta manera ya recogido en él informalmente, pues el objetivo final es la construcción de una ontología de ese dominio.

Tarea 1. Construcción de un glosario de términos

Se desarrollan los términos relevantes propios del dominio, indicando su nombre, sinónimos, acrónimos, descripción (en lenguaje natural) y el tipo de término que constituye dentro de una ontología (si es un concepto, una instancia, un atributo, una relación entre conceptos, una constante, etc.).

Si disponemos de un tesauro como punto de partida, en este glosario se incluirá cada descriptor como el nombre de un nuevo concepto; los no descriptores, bien como sinónimos o bien como nombre de un concepto, en el caso de que se eligiera un acrónimo como descriptor (en cuyo caso, dentro de la ontología, el descriptor se transformaría en un acrónimo del concepto cuyo nombre es el desarrollo del acrónimo), algunas de las notas aclaratorias (que se transcribirán como una descripción del concepto en la ontología, si contienen una definición, como un atributo si identifican una pro-

Ejemplo

A partir del descriptor "Lente de contacto" y de los no descriptores "lentilla" y "LC", y del descriptor "Miopía magna" y su nota de aplicación:

Miopía magna

NA Úsese para referirse a aquella miopía con una graduación superior a 6,00 dioptrías.

podemos consignar dentro del glosario de la ontología el nombre de los conceptos "Lente de contacto" y "Miopía magna", que cada uno de ellos es un "Concepto" (en el campo de "Tipo" de término), que "Lentilla" es un "Sinónimo" de "Lente de contacto", que las siglas "LC" son un "Acrónimo" del concepto al que se refieren ("Lente de contacto"), y que la "Descripción" de una "Miopía magna" es: "Aquella miopía con una graduación superior a 6,00 dioptrías".

iedad asociada a un concepto o a las instancias del mismo) y las relaciones asociativas, a las que habrá que dar primeramente un nombre a partir de su semántica y de los términos relacionados en el tesauro.

Tarea 2. Construcción de taxonomías

Las cuatro relaciones taxonómicas que se emplean para formalizar una ontología de peso son:

- Subclase de (*IsA*): diremos que el concepto C_1 es una subclase de otro C_2 si, y sólo si, todas las instancias de C_1 lo son también de C_2 . Ejemplo: una opacidad ocular es una subclase de anomalías estructurales.

- Descomposición disjunta: es una división de un concepto C en subclases, de forma que éstas no tienen instancias en común; puede haber instancias de C que no pertenezcan a ninguno de los conceptos de la descomposición. Ejemplo: la miopía puede dividirse en miopía normal ($0 \leq \text{graduación} < 3$ dioptrías), de riesgo ($3 \leq \text{graduación} < 6$ dioptrías) y magna (graduación ≥ 6 dioptrías). Evidentemente, un ojo miope no puede pertenecer a la vez a dos subclases de miopía.

- Descomposición exhaustiva: es una división de un concepto C en subclases, de forma que todas las instancias deben pertenecer al menos a una de las subclases de C , pero puede haber instancias de C que pertenezcan a más de una subclase. Ejemplo: el caso anterior también constituye una descomposición exhaustiva pues, por su forma de definición, no puede haber más subclases de miopía que la normal, la de riesgo y la magna.

- Partición: se define como una división de un concepto C en subclases, de forma que todas las instancias deben pertenecer a una y sólo una subclase de C . Las particiones son, por lo tanto, descomposiciones exhaustivas y disjuntas a la vez. Ejemplo: puesto que los tipos de miopía constituyen una descomposición exhaustiva y disjunta, también forman una partición del concepto miopía.

Además de estas relaciones taxonómicas, conviene distinguir, asimismo, las metonímicas, o de tipo "Parte De" (*PartOf*), que se establecen entre cada una de las partes y el todo. Pueden ser de dos tipos: transitivas o intransitivas. Las primeras son las que cumplen la afirmación: "si un concepto C_1 es parte de otro concepto, C_2 , y éste lo es de otro concepto, C_3 , entonces C_1 es parte de C_3 "; e intransitivas las que no la cumplen. Ejemplo: el iris es parte del ojo; el ojo es parte de mi cabeza; como además se verifica que el iris es parte de la cabeza, se llega a la conclusión de que estas relaciones "Parte De" son transitivas.

Si ya disponemos de un tesauro, las relaciones jerárquicas existentes nos ayudarán a obtener las distintas relaciones taxonómicas dentro de la ontología.

Ejemplo

Si contamos con un tesoro con los siguientes descriptores y relaciones jerárquicas:

	Miopía
TE	Miopía magna
TE	Miopía normal
TE	Miopía de riesgo

y sabemos que sólo existen estos tipos de miopía y que son incompatibles entre sí, por las notas aclaratorias de cada tipo de miopía que restringen la graduación de cada tipo, a partir de estas relaciones jerárquicas sabemos que hay que incluir una partición para representarla dentro de la ontología.

Tarea 3. Construcción de diagramas de relaciones binarias ad hoc

Se entiende por relación ad hoc cualquiera que no pertenezca a las especificadas en el apartado anterior. Puede establecerse entre conceptos de la misma o de distinta taxonomía de conceptos. En el caso de disponer de un tesoro, vendrán desarrolladas a partir de las relaciones asociativas de las que disponemos.

Ejemplo

Gafas

TR	Anomalía de refracción
----	------------------------

En nuestra ontología podremos incorporar la relación ad hoc "Compensa", que se establece entre el concepto "Gafa" y "Anomalía de refracción"; su relación inversa sería "Compensada por", que se establece entre el concepto "Anomalía de refracción" y "Gafa".

Tarea 4. Construcción de un diccionario de conceptos

En él especificaremos cuáles son las propiedades (los atributos) y las relaciones que describen cada concepto, así como sus instancias características. Es de vital importancia explicar el nombre del concepto, sus instancias, los atributos de clase (aquellos cuyo valor caracteriza a los individuos de la misma clase), los atributos de instancia (distintos y distintivos de cada individuo de la clase) y las relaciones (de las que son origen).

Por el proceso seguido anteriormente, en este paso se incluyen ya de una manera u otra prácticamente la

mayoría de los elementos de un tesoro; es una etapa de resumen y de clasificación de la información obtenida del tesoro en etapas anteriores.

Tarea 5. Definición detallada de las relaciones binarias ad hoc

A partir del diccionario de conceptos creado en la tarea anterior, especificaremos el nombre de la relación, el concepto de origen (denominado también fuente), la cardinalidad (el número mínimo y máximo de instancias del concepto destino con las que se puede corresponder, mediante una misma relación, una instancia del concepto origen), el concepto de destino, las propiedades matemáticas que se cumplen -reflexiva, simétrica, transitiva, etc.; véase su definición, por ejemplo, en: **Caballero Roldán**, et al., (2007)- y la relación inversa a que da lugar.

La cardinalidad de una relación muy difícilmente podrá ser extraída de un tesoro, por lo que habrá que recurrir a las fuentes de elaboración del mismo para determinarla. Las propiedades matemáticas se pueden deducir de la semántica de la relación asociativa del tesoro de la que proviene la relación ad hoc de la ontología.

Tarea 6. Definición detallada de los atributos de instancia

Asimismo, a partir del diccionario de conceptos, incluiremos los atributos de la instancia que consideremos relevantes en el dominio, identificando para cada uno de ellos su nombre, clase a la que se asocia, el tipo de valores que adopta (cadena de caracteres, número entero, número real, booleano, fecha, etc.), las unidades

Ejemplo 1

Dioptría

TR	Miopía
----	--------

Ejemplo 2

Presbicia

NA Útese en casos de anomalías en la visión próxima padecidas por personas mayores de 40 años de edad

Del ejemplo 1 podemos deducir que la "Dioptría" es la "Unidad de medida" de la "Miopía"; del ejemplo 2 se infiere: el nombre del atributo ("Edad"), el tipo de valor del atributo (entero), así como que "Años" es la unidad en la que se mide una "Edad". En este caso, existe una restricción sobre el rango de valores viene dada por la semántica propia del concepto "Edad".

de medida, la precisión que se observa con cada variación de valor y el rango de valores (en el caso de números), el valor por defecto, las cardinalidades máxima y mínima (se admite la posibilidad de que haya atributos multivaluados) y otros datos que consideremos necesarios (como por ejemplo, qué otros atributos y constantes se usan al inferir el valor de este atributo).

En el caso de partir de un tesoro, las unidades de medida suelen venir reflejadas como descriptores y a la vez como términos relacionados dentro del tesoro, o a veces aparecen dentro de una nota aclaratoria, asociadas a un valor del atributo que califican y que dará a su vez el tipo de valores del mismo.

Tarea 7. Definición detallada de los atributos de clase

Especificaremos su nombre, el concepto en el que el atributo está definido, el tipo de valor que posee, su valor (pues caracteriza a la clase, lo que lo distingue de los atributos de instancia), las unidades de medida, la precisión, el rango de valores, su cardinalidad y los atributos de instancia cuyos valores se pueden inferir con los valores de este atributo de clase. Si se parte de un tesoro para construir la ontología, las consideraciones son análogas al caso anterior.

Ejemplo

Miopía magna

NA Úsese en el caso de aquellas miopías con una graduación mayor de 6,00 dioptrías.

Del ejemplo podemos deducir el nombre del atributo ("Graduación"), el tipo de valor del atributo (real) así como que "Dioptría" es la "Unidad" con la que se mide una "Graduación" y que existe una restricción sobre el rango de valores (> 6,00) para dicho atributo.

Tarea 8. Definición detallada de las constantes

Gracias al glosario desarrollado en la tarea 1 y al diccionario de conceptos del paso 4 podremos listar las constantes que necesitamos en nuestra ontología. Para cada una de ellas es importante mencionar su nombre, el tipo de valor, su valor, las unidades de medida y los atributos que se pueden inferir mediante su uso. Difícilmente en un tesoro se ven definiciones de constantes, por lo que para la realización de esta tarea habrá que recurrir a las fuentes de la elaboración del mismo.

Tarea 9. Modelización de los axiomas

Un axioma se puede definir como una "afirmación siempre cierta en el ámbito de la ontología que se está

formalizando". Es necesario, pues, la identificación y descripción de los axiomas del dominio que se está conceptualizando, indicando para cada uno de ellos su nombre, descripción y la expresión lógica que lo representa, así como los conceptos, atributos, relaciones ad hoc, y variables utilizadas en dicha expresión. En el caso de contar con un tesoro como punto de partida, los axiomas vendrán definidos a partir de las notas de aplicación.

Ejemplo

Miopía magna

NA Úsese en el caso de aquellas miopías con una graduación mayor de 6,00 dioptrías.

De la nota de aplicación anterior podemos extraer también el siguiente axioma en lógica de primer orden, según el escritor de axiomas de WebODE:

Forall (?X, ?Y)

([Miopía Magna] (?X) and [Graduación] (?X, ?Y) -> [Mayor] (?Y, 6.00))

en el que estarían implicados el concepto "Miopía magna", el atributo "Graduación" y las variables ?X e ?Y.

Tarea 10. Definición de las reglas

Son un tipo especial de axioma por el que calculamos, por ejemplo, el valor de un atributo en función de otros ya especificados o de valores de constantes. Han de formalizarse especificando el nombre, su descripción en lenguaje natural y la expresión lógica que representa la regla, así como los conceptos, atributos, relaciones ad hoc y variables usados.

En el caso de contar con un tesoro, también las reglas vendrán definidas a partir de las notas de aplicación. La diferencia entre implementar una nota aclaratoria como una regla o como un axioma es contextual: si, por ejemplo, un atributo recibe un valor determinado según la misma, se implementará mediante una regla; en caso contrario, a través de un axioma.

Tarea 11. Definición de las instancias

La última tarea pasa por indicar las características de las instancias incluidas en el diccionario de conceptos creado en la tarea 4. Indicaremos, para cada una de ellas, su nombre, el del concepto al que pertenece, los atributos de instancia que posee y el valor de dichos atributos. Todo este conocimiento, dado que las instancias no figuran en forma alguna en un tesoro, deberá obtenerse de las fuentes de partida para la construcción del tesoro.

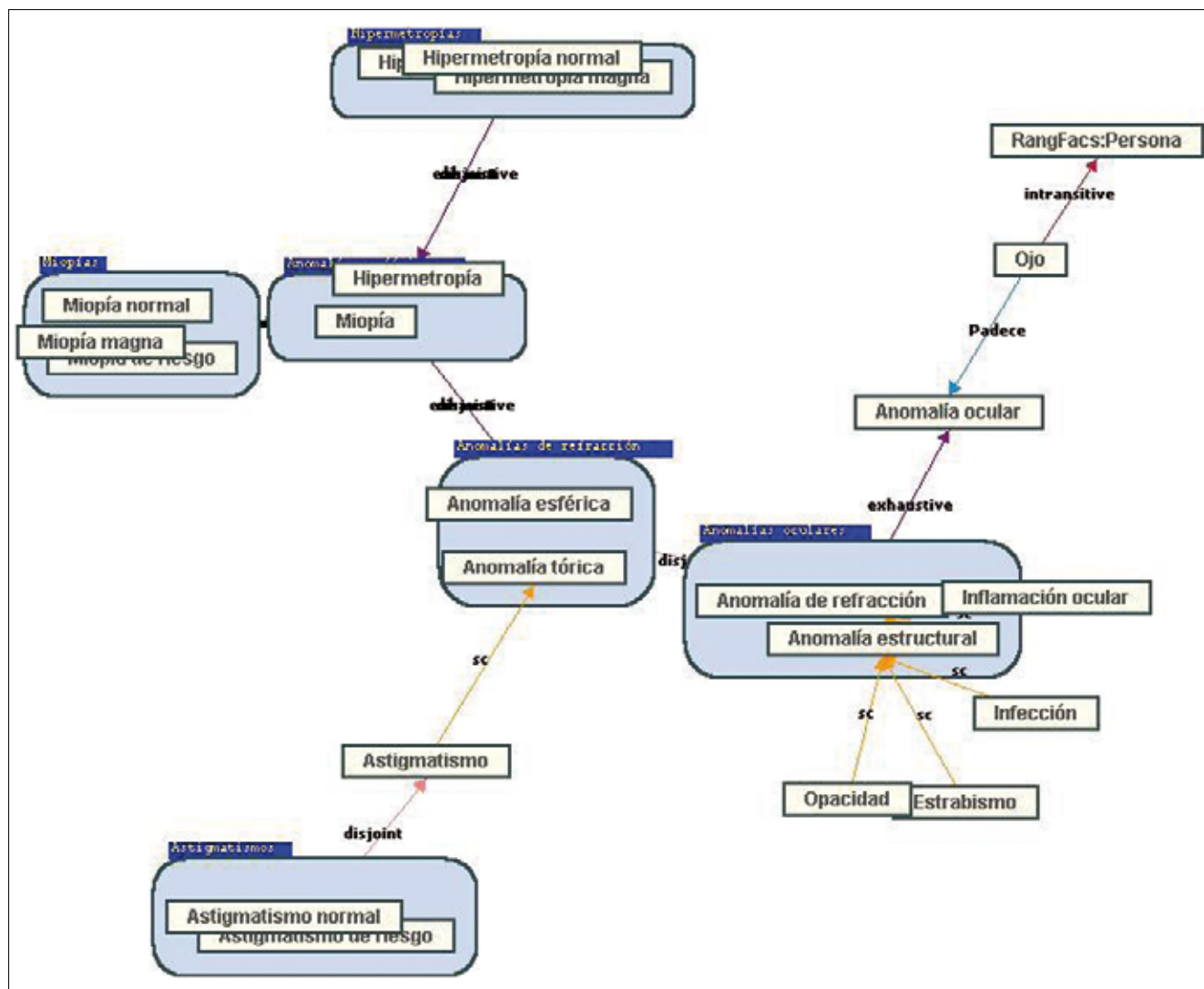


Figura 4. Representación gráfica de una ontología de dominio para anomalías oculares, derivada de un tesoro

tarea, además de sistematizarla y robustecerla. La formalización del conocimiento como una ontología y no como un tesoro aporta los beneficios adicionales mencionados al final del estudio comparativo entre ambas herramientas.

Por otro lado, y como se puede observar en el apartado anterior, se pueden articular varias ontologías, como si de un puzzle se tratara, para obtener una arquitectura ontológica que representa globalmente un dominio, lo cual es impensable en el caso de tesauros, ya que operan en un campo muy restringido del conocimiento. Éste es un aspecto en el que se espera que la investigación sea muy fructífera en el futuro.

El inconveniente que presenta nuestra propuesta para un documentalista es que deberá conocer y manejar una serie de conceptos nuevos propios de la lógica (lógica de predicados, descriptiva, etc.), la matemática discreta (propiedades de relaciones: reflexiva, simétrica, transitiva, etc.), la ingeniería del conocimiento (técnicas de educción y formalismos de representación del conocimiento, etc.), aparte de aquellos propios de su

profesión, como los relacionados con las bases de datos (formatos de representación de la información: números reales, enteros, booleanos, etc.).

“Se pueden articular varias ontologías, como si de un puzzle se tratara, para obtener una arquitectura de ontologías que representan globalmente un dominio”

En cualquier caso, el principal problema (sobre todo en la web semántica) es la adopción de ontologías concretas por parte de un grupo tan heterogéneo de usuarios como es internet para un dominio determinado. Las ontologías necesitan ser aceptadas masivamente por toda la red para conseguir sus objetivos: hacer explícitos conocimientos y servir de base para la inferencia sobre los mismos (en caso contrario, harían falta mecanismos de traducción entre ontologías similares a los existentes para la traducción entre distintas len-

Concept name	Synonyms	Acronyms	Instances	Class attributes	Instance attributes	Relations
Acción	-	-	-	Natural	-	SeRealizaEn SeRealizaSobre Posee
Animal	-	-	-	Microscópico Racional SeTraslada	-	-
Disciplina	-	-	Optometría	-	-	-
Edificio	-	-	Óptica	-	-	-
Ente	-	-	-	Individual	-	Intervienen SeCaracterizaPor
Ente Colectivo	-	-	-	Individual	-	-
Ente Individual	-	-	-	Animado Individual	-	-
Equipamiento	-	-	-	-	-	PropioDe UsadoParaRealizar Tiene
Fenómeno	-	-	Infección Inflamación ocular Opacidad del ojo Refracción de la luz	Natural	-	AjenoA
Herramienta	-	-	Queratómetro Frontofocómetro Gafa Autorrefractómetro Destornillador	-	-	OperaSobre
Material	-	-	Ojo Lente	-	-	-
Microorganismo	-	-	-	Microscópico SeTraslada	-	-
Organización	-	-	Colegio Nacional de Ópticos Optometristas	-	-	-
Persona	Ser Humano	-	Optometrista Óptico Paciente	Racional	-	Usa Provoca
Proceso	-	-	Compra de herramientas Compra de materiales Montaje de gafa Graduación	Natural	-	-
Propiedad	-	-	Apellidos Nombre Dirección Magnitud de refracción Signo de refracción	-	-	-
Ser Inanimado	-	-	-	Animado	-	-
Ser Vivo	-	-	-	Animado Microscópico SeTraslada	-	-
Vegetal	Planta	-	-	Microscópico SeTraslada	-	-
Vehículo	-	-	-	-	-	-

Tabla 2: Diccionario de conceptos de la ontología de facetas de Ranganathan

guas); pero es difícil coordinar y organizar la web para alcanzar esta aceptación masiva, pues no hay forma de imponer el uso de una ontología concreta a semejante volumen de usuarios implicados.

Bibliografía

- Aguado, G.; Álvarez-de-Mon I.; Gómez-Pérez, A.; Pareja-Lora, A.; Plaza-Arteche, R.** *A semantic web page linguistic annotation model. Semantic web meets language resources. Technical report WS-02-16.* American Association for Artificial Intelligence. Menlo Park, California, EUA: AAAI Press, 2002.
- Aguado-de-Cea, G.; Álvarez-de-Mon, I.; Pareja-Lora, A.; Plaza-Arteche, R.** "OntoTag: a semantic web page linguistic annotation model". En: *Proceedings of the ECAI 2002 workshop on semantic authoring, annotation and knowledge markup*, 2002.
- Aguado-de-Cea, G.; Álvarez-de-Mon, I.; Pareja-Lora, A.; Plaza-Arteche, R.** "Rdf(S)/xml linguistic annotation of semantic web pages". En: *Proceedings of the Coling 2002, 2nd workshop on nlp and xml (Nlpxml-2002)*, 2002.
- Aitchison, J.; Gilchrist, A.; Bawden, D.** *Thesaurus construction and use: a practical manual*. 3^a ed. Londres: Aslib, 1997.
- Benjamins, V. R.; Fensel, D.; Decker, S.; Gómez-Pérez, A.** "(KA)²: building ontologies for the internet: a mid term report". En: *International journal of human computer studies (Ijhc)*, 1999, n. 51, pp. 687-712.
- Benjamins, R.; Contreras, J.; Martín, F.; Navarrete, B.; Aguado-de-Cea, G.; Álvarez-de-Mon, I.; Pareja-Lora, A.; Plaza-Arteche, R.** *State of the art on annotation tools and services*. Esperanto Services (IST-2001-34373), deliverable D3.1. 2003. <http://www.esperanto.net>
- Berners-Lee, T.; Fischetti, M.** *Weaving the web: the original design and ultimate destiny of the world wide web by its inventor*. San Francisco: Harper, 1999.
- Borst, W. N.** *Construction of engineering ontologies for knowledge sharing and reuse. PhD Thesis*. University of Twente, The Netherlands, Enschede, 1997.
- Caballero-Roldán, R.; Hortalá-González, T.; Martí-Oliet, N.; Nieva-Soto, S.; Pareja-Lora, A.; Rodríguez-Artalejo, M.** *Matemática discreta para informáticos: ejercicios resueltos*. Madrid: Pearson Educación SA, 2007.
- Dean, M.; Schreiber, G.** *OWL web ontology language reference. W3C Recommendation*. Edición 2004. Consultado en: 28-09-07. <http://www.w3.org/TR/owl-ref/>
- Ding, Y.; Foo, S.** "Ontology research and development. Part 1—a review of ontology generation". En: *Journal of information science*, 2002, v. 28, n. 2, pp. 123-136.
- Ding, Y.; Foo, S.** "Ontology research and development. Part 2—a review of ontology mapping and evolving". En: *Journal of information science*, 2002, v. 28, n. 5, pp. 375-388.
- Ferreira, D.** *TemaTres: software libre para gestión de tesauros*. Consultado en: 12-09-07. <http://www.r020.com.ar/tematres/>
- García-Jiménez, A.** "Instrumentos de representación del conocimiento: tesauros versus ontologías". En: *Anales de documentación*, 2004, n. 7, pp. 79-95.
- Gil-Urdiciain, B.** "La 'Colon Classification'". En: *Revista general de información y documentación*, 1994, v. 4, n. 1, pp. 111-121.
- Gómez-Pérez, A.; Corcho, O.** "Ontology specification languages for the semantic web". En: *IEEE intelligent systems*, 2002, enero, v. 17, n. 1, pp. 54-60.
- Gómez-Pérez, A.; Fernández-López, M.; Corcho, O.** *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. Londres; Nueva York: Springer-Verlag, 2003.
- Gruber, R.** "A translation approach to portable ontology specification". En: *Knowledge acquisition*, 1993, n. 5, pp. 199-220.
- Lassila, O.; McGuinness, D.** *The role of frame-based representation on the semantic web. KSL tech report number KSL-01-02*, enero de 2001. <http://www.ksl.stanford.edu/people/dlm/etai/lassila-mcguinness-fbr-sw.html>
- Luke S.; Hefflin J.** *SHOE 1.01. Proposed Specification. SHOE Project*. Consultado en: 14-09-07. <http://www.cs.umd.edu/projects/plus/SHOE/spec.html>
- Motta, Enrico; Buckingham-Shum, Simon; Domingue, John.** "Case studies in ontology-driven document enrichment". En: *Proceedings of the 12th Banff knowledge acquisition workshop*, Alberta, 1999.
- OntoMat*. Consultado en: 12-09-07. <http://annotation.semanticweb.org/ontomat/index.html>
- OntoWeb*. Consultado en: 12-09-07. http://ontoweb.aifb.uni-karlsruhe.de/About/Deliverables/D13_v1-0.zip
- Peis, E.; Herrera-Viedma, E.; Hassan, Y.; Herrera, J. C.** "Análisis de la web semántica: estado actual y requisitos futuros". En: *El profesional de la información*, 2003, septiembre-octubre, v. 12, n. 15.
- Pérez-Agüera, J. R.** "L'automatització de tesauros i la seva utilització en el web semàntic". En: *BiD, textos universitaris de biblioteconomia i documentació*, 2004, diciembre, n. 13. http://www2.uib.es/bid/consulta_articulos.php?fichero=13perez1.htm
- Qin, J.; Paling, S.** "Converting a controlled vocabulary into an ontology: the case of GEM". En: *Information research*, 2001, v. 6, n. 2. <http://informationr.net/ir/6-2/paper94.html>
- Ranganathan, S. R.** *Prolegomena to library classification*, 3^a ed. New Delhi: Asia Publishing House, 1967.
- Staab, S.; Angele, J.; Decker, S.; Erdmann, M.; Hotho, A.; Mädche, A.; Schnurr, H. P.; Studer, R.** "Semantic community web portals". En: *Proceedings of the 9th international world wide web conference (www9)*, 2000, v. 33, n. 1-6 (número especial), pp. 473-491.
- Studer, R.; Benjamins, R.; Fensel, D.** "Knowledge engineering: principles and methods". En: *Data and knowledge engineering (DKE)*, 1998, v. 25, n. 1-2, pp. 161-197.
- WebODE*. Consultado en: 01-07-07. <http://webode.dia.fi.upm.es>
- Wilcock, G.; Buitelaar, P.; Pareja-Lora, A.; Bryant, B.; Lin, J.; Ide, N.** "The roles of natural language and xml in the semantic web". En: **Huang, C. R.; Lenders, W.** (eds.). *Computational linguistics and beyond: perspectives at the beginning of the 21st century*. Frontiers in linguistics series, Academia Sinica, Taiwan, 2004.

Alberto García-Torres, Tribunal Supremo, Plaza Villa de París, s/n, 28071 Madrid.
torres.al@gmail.com

Antonio Pareja-Lora, Dpto. Sistemas Informáticos y Computación, Universidad Complutense de Madrid, C/ Profesor García Santesmases, s/n, 28040 Madrid.
apareja@sip.ucm.es

Daniel Pradana-López, Óptica Gasset, C/ José Ortega y Gasset 96, 28006 Madrid.
opticagasset@yahoo.es