

**DISEÑO METODOLÓGICO PARA LA GENERACIÓN DE ONTOLOGÍAS A
PARTIR DE LA UTILIZACIÓN DE TESAUROS EN LÍNEA**

YENNY MARCELA ROJAS GUZMÁN

**PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE COMUNICACIÓN Y LENGUAJE
CARRERA DE CIENCIA DE LA INFORMACIÓN – BIBLIOTECOLOGÍA
BOGOTÁ
2012**

**DISEÑO METODOLÓGICO PARA LA GENERACIÓN DE ONTOLOGÍAS A
PARTIR DE LA UTILIZACIÓN DE TESAuros EN LÍNEA**

YENNY MARCELA ROJAS GUZMÁN

**Trabajo de grado presentado para optar por el título de
Profesional en Ciencia de la Información - Bibliotecología**

**DIRECTORA: LUCILA REBECA ESPINOSA RICARDO
Bibliotecóloga**

**PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE COMUNICACIÓN Y LENGUAJE
CARRERA DE CIENCIA DE LA INFORMACIÓN – BIBLIOTECOLOGÍA
BOGOTÁ
2012**

Bogotá D.C., 25 de enero de 2013

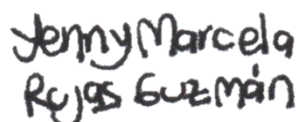
Doctor
JOSÉ VICENTE ARISMENDI
Decano académico
Facultad de Comunicación y Lenguaje
Pontificia Universidad Javeriana

Respetado decano:

Reciba un cordial saludo.

Por medio de la presente, hago entrega oficial del trabajo de grado para optar por el título de Profesional en Ciencia de la Información – Bibliotecología, titulado **“Diseño metodológico para la generación de ontologías a partir de la utilización de tesauros en línea”**.

Atentamente,



YENNY MARCELA ROJAS GUZMÁN
Ciencia de la Información – Bibliotecología

Reglamento de la Pontificia Universidad Javeriana

Artículo 23

“La Universidad no se hace responsable por los conceptos emitidos por los alumnos en sus trabajos de grado, solo velará porque no se publique nada contrario al dogma y la moral católicos y porque el trabajo no contenga ataques y polémicas puramente personales, antes bien, se vean en ellas el anhelo de buscar la verdad y la justicia”.

*A mi papi por llevarme por el camino de esta profesión
y apoyarme cada vez que lo necesité.*

*A mi mami, por su cariño, comprensión y compañía en
todo el camino recorrido.*

A mi hermana, por estar a mi lado y apoyarme siempre.

¡Gracias por todo!

Los quiero mucho

AGRADECIMIENTOS

Este ha sido un camino largo que he recorrido en compañía de todas las personas a quienes quiero, que estuvieron a mi lado siempre y me apoyaron en los momentos en los que más lo necesité...

Papi, te agradezco por haberme iniciado en esta profesión, eres mi ejemplo de vida por tu transparencia y bondad. Gracias a ti, estoy aquí, concluyendo otra etapa de mi vida.

Mami, te doy gracias por todo el apoyo que me brindaste, por tu paciencia y compañía en los momentos difíciles, por estar a mi lado siempre y compartir conmigo la alegría que siento al estar finalizando mi carrera.

Vivi, mil gracias por todo, por tu compañía, por creer en mí y por alegrarme cada vez que fue necesario.

George y Peri, gracias por hacer que cada momento fuera increíble, por las risas compartidas, por el apoyo recibido y todo lo vivido.

A mis profesores, gracias por compartir todo su conocimiento y experiencia, aportando granitos de arena que me permitieron llegar a este punto de mi formación profesional.

Profe Lucy, gracias por su guía y orientación en el desarrollo de esta investigación, por las conversaciones amenas de cada tarde y por su confianza en mí.

¡Mil gracias a todos ustedes!

CONTENIDO

	Pág.
INTRODUCCIÓN.....	10
1. DEFINICIÓN DEL PROBLEMA	12
1.1. EL ESTADO DE LA CUESTIÓN	12
1.2. EL TESAURO DE AGRICULTURA DE LA <i>NATIONAL AGRICULTURAL LIBRARY</i> (NAL)	15
2. JUSTIFICACIÓN	19
3. OBJETIVOS	21
3.1. OBJETIVO GENERAL.....	21
3.2. OBJETIVOS ESPECÍFICOS	21
4. LA INDIZACIÓN Y LOS SISTEMAS DE ORGANIZACIÓN DEL CONOCIMIENTO.....	22
4.1. Los procesos de descripción de contenidos: La indización	22
4.2. Los lenguajes documentales	24
4.3. Los tesauros	32
4.4. Los sistemas de organización del conocimiento y el entorno digital.....	37
4.5. Las ontologías como sistemas de organización del conocimiento en el entorno web	38
4.6. Las ontologías y los lenguajes documentales.....	50
5. COMPARACIÓN DE TESAUROS Y ONTOLOGÍAS.....	57
5.1. Aspectos terminológicos de los tesauros y las ontologías.....	57
5.2. Relaciones terminológicas presentes en los tesauros y las ontologías	59
5.3. Semejanzas y diferencias entre tesauros y ontologías	61
6. INDICADORES PARA LA EVALUACIÓN DE TESAUROS	64
7. METODOLOGÍA PARA LA CONVERSIÓN DE TESAUROS EN ONTOLOGÍAS	69
7.1. Diseño de la metodología.....	69
7.2. Implementación de la metodología.....	70
8. CONCLUSIONES.....	78
BIBLIOGRAFÍA.....	80

LISTADO DE TABLAS

	Pág.
Tabla 1. Términos asociados con la subcategoría Ambiente, dentro del tesoro de la National Agricultural Library.....	71
Tabla 2. Términos asociados con la subcategoría Ecología, dentro del tesoro de la National Agricultural Library.....	72
Tabla 3. Términos asociados con la subcategoría Recursos naturales, dentro del tesoro de la National Agricultural Library.....	72
Tabla 4. Identificación de las clases y subclases que agruparon los términos incluidos en la ontología desarrollada en este trabajo de grado.....	74

LISTADO DE FIGURAS

	Pág.
Figura 1. Categorías del Tesouro de Agricultura.....	17
Figura 2. Términos específicos asociados a Tierra, ambiente y recursos naturales.....	17
Figura 3. Clasificación de lenguajes documentales.....	27
Figura 4. Lenguajes documentales de acuerdo con la coordinación entre sus términos.....	28
Figura 5. Clasificación de los lenguajes documentales de acuerdo con su estructura.....	29
Figura 6. Comparación entre la terminología asociada a la Categoría Ecología, en el tesouro de la NAL y una pequeña muestra de la estructura ontológica que puede crearse si se trabajan las relaciones presentes en el tesouro.....	65
Figura 7. Relaciones jerárquicas definidas en el tesouro de la National Agricultural Library.....	66
Figura 8. Relaciones presentes en la subcategoría Recursos Naturales, de la categoría Tierra, Ambiente y Recursos naturales.....	67
Figura 9. Relaciones de la subcategoría Recursos naturales con otros términos que no se vinculan directamente en ella, y que se pueden asociar haciendo uso de verbos que enriquecen su vínculo.....	67
Figura 10. Relaciones asociativas entre las clases y subclases identificadas en la etapa anterior.....	74
Figura 11. Estructura ontológica asociada con la clase Ecología.....	75
Figura 12. Estructura ontológica asociada con la clase Problemática ambiental.....	76

INTRODUCCIÓN

Con el desarrollo de la práctica profesional en la actualización y mejora de un tesoro, se despertó el interés por encontrar la forma a través de la cual un tesoro en línea pudiera converger hacia una ontología, con el objetivo principal de favorecer el proceso de búsqueda y recuperación de información por parte de los usuarios de los sistemas de información actuales.

Como consecuencia de dicho interés, se empezó a indagar con respecto al tema y se decidió realizar como trabajo de grado para optar por el título de Profesional en Ciencia de la información - Bibliotecología, el desarrollo de una metodología que permitiera hacer la convergencia de tesoros hacia ontologías con un enfoque estructural que en caso de continuar evolucionando, pudiera ser adaptada a alguna herramienta tecnológica que posibilitara su digitalización.

En el presente trabajo de grado se hace una revisión teórica de los tesoros y las ontologías en el contexto de los lenguajes documentales y los sistemas de organización del conocimiento, con el objeto de conocer sus estructuras para hacer una posterior comparación que permitiera identificar los aspectos comunes y diferentes que tuvieran entre sí.

Para cumplir con el objetivo principal que persigue este trabajo de grado, relacionado con el desarrollo de una metodología que permita que un tesoro converja hacia una ontología, se estableció un grupo de indicadores que posibilitara la evaluación de un tesoro como el de la National Agricultural Library para identificar si tenía o no potencial de converger hacia una ontología, teniendo en cuenta especialmente las diferentes relaciones que presentaba y la representación conceptual que hacía, aspectos clave identificados en la comparación realizada entre tesoros y ontologías.

Posterior a la proposición y aplicación de estos indicadores, se inició el diseño de la metodología que se generó como producto de este trabajo de grado teniendo en cuenta los elementos clave identificados durante la comparación y algunos aspectos rescatados de las metodologías de desarrollo de ontologías consultadas durante el desarrollo del marco teórico de este proyecto.

Este trabajo de grado se realiza con la intención de aportar una posible mejora al proceso de búsqueda de información en la red, desde una perspectiva semántica, con el objeto de mejorar la efectividad en el momento en que el usuario realice la recuperación de documentos, ofreciéndole la información que necesita para dar respuesta a sus inquietudes.

Adicionalmente, como pieza clave dentro de la finalización del proceso de formación en el área de la Ciencia de la Información, el presente trabajo de grado da cuenta de la interdisciplinariedad que se puede desarrollar en esta Carrera al generar un vínculo entre algunas categorías de las ciencias ambientales y el quehacer de la profesión en materia de generación de recursos que faciliten la búsqueda y recuperación de información en las diferentes áreas del conocimiento, en las que estas herramientas se especialicen.

1. DEFINICIÓN DEL PROBLEMA

1.1. EL ESTADO DE LA CUESTIÓN

Al realizar una búsqueda de información que relacionara a los tesauros y a las ontologías se recuperaron y revisaron diferentes documentos que dan cuenta de la forma en la que se han vinculado en el contexto de la organización del conocimiento a partir de su estructuración en el entorno académico, cultural y web.

Una de las tendencias identificadas, planteada por Codina y Pedraza¹ tiene que ver con la implementación de ontologías en sistemas de información ya establecidos, como complemento o remplazo de los tesauros que antes estructuraban los contenidos y hacían posible su recuperación por parte de los usuarios de estos sistemas.

En el desarrollo de ese artículo, los autores hacen énfasis en las ventajas que cada uno de estos sistemas genera para la descripción y organización de la información disponible en los entornos web, planteando una comparación a partir de la combinación de tesauros y ontologías con esquemas de metadatos en función de la recuperación de contenidos e información realizada por los usuarios de estos sistemas.

Otra de las tendencias que puede identificarse en los documentos consultados se relaciona con la definición de tesauros y ontologías desde la identificación de sus principales componentes, sus características y sus aplicaciones, tema desarrollado por Gilchrist² y por de Sales y Café³, así como su diferenciación a

¹ CODINA, Lluís y PEDRAZA-JIMÉNEZ, Rafael. Tesauros y ontologías en sistemas de información documental. En: *El profesional de la información*, v. 20, no. 5 (sept.-oct, 2011) p. 555 - 563

² GILCHRIST, Alan. Thesauri, taxonomies and ontologies - an etymological note. En: *Journal of Documentation*, v. 59, no. 1 (2003) p. 7 - 18

³ DE SALES, Rodrigo y CAFÉ, Ligia. Diferencias entre tesauros e ontologías. En: *Perspectivas em Ciencia da Informacao*, v.14, no. 1 (ene.-abril, 2009) p. 99 – 116

partir de su correspondencia con taxonomías que organizan conocimiento y por las estructuras mismas que dan soporte tanto a los tesauros como a las ontologías.

Teniendo en cuenta que el objetivo de esta investigación se vincula con la convergencia de tesauros hacia ontologías, se pudo identificar una tendencia que se encuentra directamente relacionada con lo que se quiere plantear como producto de la misma. Autores como García-Torres, Pareja-Lora y Pradana-López⁴ proponen la reutilización de las estructuras planteadas en los tesauros existentes para la generación de ontologías y otros recursos semánticos que contribuyan con el desarrollo y avance de la web semántica, no solo desde los sistemas informáticos sino también desde la perspectiva de la Ciencia de la Información.

Por otra parte, Hilera, Pages, Martínez, Gutiérrez, y De-Marcos⁵ plantean que existe un proceso evolutivo visible en los sistemas de organización del conocimiento, que ha permitido que de los glosarios y otros KOS (Knowledge Organization System) más primitivos se haya podido llegar a desarrollar taxonomías y tesauros que han facilitado la descripción y recuperación de la información hasta el día de hoy, y que de seguir avanzando puede incluso generar la convergencia de los tesauros y demás sistemas de organización del conocimiento que conocemos, hacia ontologías y otras herramientas semánticas que permitan trasladar las bases bibliotecológicas de descripción en beneficio de la recuperación de información por parte de los usuarios, a un entorno digital regido por lo semántico.

⁴ GARCÍA TORRES, Alberto, et al. Reutilización de tesauros: El documentalista frente al reto de la web semántica. En: El profesional de la información, v.17, n. 1, (ene.-feb., 2008) p. 8 - 21

⁵HILERA R., José, et al. An evolutive process to convert glossaries into ontologies. En: Information Technology and Libraries, v. 29 no. 4 (dic, 2010) p. 195-204

Siguiendo esta tendencia, en la que los tesauros y las ontologías son vistos como sistemas que convergen y evolucionan en un entorno digital semántico, se recuperaron documentos como el de Ibekwe-SanJuan⁶ o el de Ospina Rodríguez⁷, en los que se plantean diferentes metodologías para hacer posible la construcción de un sistema semántico de organización del conocimiento como una ontología o un tesoro, en entornos particulares, como se presenta a continuación:

- En el artículo de Ibekwe-SanJuan⁸, la metodología que se propone consiste en la ejecución de una aplicación que extrae las expresiones clave o frecuentes que se encuentran en un texto definido, con el fin de proponer un listado de términos que componen un tesoro cuyas estructuras semánticas luego son desarrolladas a partir de la identificación del sentido y propósito del texto analizado.
- En el trabajo de grado de Ospina Rodríguez⁹, la metodología se relaciona con la identificación del proceso de construcción de un vocabulario controlado y con la reutilización de los términos incluidos en el Tesoro de Literatura infantil, para la construcción de una ontología basada en las estructuras semánticas identificadas en éste último.

Adicionalmente, Moura¹⁰ y Sánchez-Alonso y García-Barriocanal¹¹ proponen que los tesauros y las ontologías son sistemas de organización del conocimiento que cuentan con un alto nivel de interoperabilidad debido a las similitudes que pueden llegar a encontrarse en los esquemas conceptuales de los que se derivan sus

⁶IBEKWE-SANJUAN, Fidelia. Constructing and maintaining knowledge organization tools: a symbolic approach. En: Journal of Documentation, v. 62, no. 2, p.229 - 250

⁷OSPINA RODRÍGUEZ, Martha Carolina. Guía metodológica para la elaboración de una ontología a partir de un tesoro: estudio analítico comparativo (Fase 1). En: Códices, v. 7, no. 1 (ene.-jun., 2011) p. 113 - 130

⁸IBEKWE-SANJUAN, Fidelia. Óp. Cit.

⁹Ospina Rodríguez, Martha Carolina. Óp. Cit.

¹⁰MOURA, Maria Aparecida. Information, ontological tools and ad hoc social nets: the interoperability in the construction of thesaurus and ontologies. En: Informacao & Sociedade-Estudos, v. 19, no.1, (2009) p. 59-73.

¹¹SANCHEZ-ALONSO, Salvador y GARCIA-BARRIOCANAL, Elena. Making use of upper ontologies to foster interoperability between SKOS concept schemes. En: Online InformationReview, v. 30, no. 3 (2006) p. 263-277

estructuras, razón por la cual sería válido afirmar que la construcción de ontologías a partir de tesauros existentes es viable y sencilla, en la medida en que los tesauros que se vayan a usar estén bien contruidos.

Siguiendo los planteamientos que pudieron identificarse en esta última tendencia descrita, relacionados con la evolución de los sistemas de organización del conocimiento hacia ontologías, en el desarrollo de este trabajo de grado se propondrá una metodología que posibilite la conversión de un tesauo que se encuentre construido de manera adecuada, en una ontología que cuente con los elementos necesarios para facilitar la búsqueda y recuperación de información por parte de los usuarios, en un entorno más semántico.

1.2. EL TESAURO DE AGRICULTURA DE LA *NATIONAL AGRICULTURAL LIBRARY* (NAL)

Para plantear la metodología que se presentará más adelante, se partió del tesauo de la Biblioteca Nacional de Agricultura (NAL – *National Agricultural Library*) del Departamento de Agricultura de Estados Unidos¹², la unidad de información más importantes especializada en el tema, que apoya las actividades de los investigadores, agricultores y demás personas interesadas en el manejo de los cultivos y los animales en las áreas rurales de dicho país.

El Tesauo de Agricultura fue creado en 2001 por el personal de la NAL para la normalización terminológica de los descriptores con los que realizaban la indización de todos los documentos y recursos con los que contaban en su base

¹²NATIONAL AGRICULTURAL LIBRARY. Biblioteca Nacional de Agricultura [En línea] Disponible en: <http://www.nal.usda.gov/about/spanfly.pdf> (Consultado el 8 de septiembre de 2012)

de datos, para que sus usuarios los recuperaran de manera directa a través de sus buscadores¹³.

Posterior a su desarrollo en inglés y con el ánimo de apoyar las actividades investigativas y de desarrollo de la comunidad latinoamericana en el área de la agricultura, en 2007 la NAL planteó la versión del Tesoro de Agricultura en español que cuenta con las mismas características y opciones que la versión en inglés, a pesar de contar con menos cancelaciones terminológicas que ésta última.

Como resultado de la iniciativa de la NAL, se generó un vocabulario controlado que se actualiza anualmente desde entonces, y que en la actualidad cuenta con alrededor de 88.000 mil términos en su versión en inglés, y 76.000 en la versión en español¹⁴, que además de cumplir con su función inicial vinculada con la indización y recuperación de la información asociada a esta unidad de información, es una fuente de consulta para los investigadores en la temática que además hacen uso de repositorios, bases de datos, catálogos y demás fuentes de información existentes.

Este tesoro puede consultarse en un índice alfabético que reúne a todos los términos que lo conforman o a través de diecisiete listados por materias que clasifican los términos mencionados, de la siguiente forma:

¹³MOCHON BEZARES, Gonzalo y SORLI ROJO, Ángela. Tesoros de Ciencias Naturales en Internet. En: Revista Española de Documentación Científica, v. 31. no. 4 (oct.-dic., 2008) p. 650 [En línea] Disponible en: <http://redc.revistas.csic.es/index.php/redc/article/view/448/460> (Consultado el 8 de septiembre de 2012)

¹⁴Idem.

- **Alimentación y Nutrición Humana**
- **Biología**
- **Ciencia de Animales y Productos de Animales**
- **Ciencia de Plantas y Productos de Plantas**
- **Ciencia Forestal y Productos Forestales**
- **Ciencias Física y Química**
- **Clasificación Taxonómica de Organismos**
 - Algae
 - Animalia
 - Archaea, Cyanobacteria y Bacteria
 - Hongos, Levaduras, Mohos y Mildius
 - Plantae
 - Virus y Viroides
- **Crianza y Mejora Genética**
- **Economía, Negocio e Industria**
- **Gobierno, Ley y Regulaciones**
- **Granjas y sus Sistemas**
- **Insectos y Entomología**
- **Investigación, Tecnología e Ingeniería**
- **Salud y Patología**
- **Sociología Rural y Agropecuaria**
- **Tierra, Ambiente y Recursos Naturales**
- **Ubicaciones Geográficas**

Figura 1. Categorías del Tesoro de Agricultura¹⁵

Posterior a la revisión de estas categorías, para el diseño e implementación de la metodología que se construyó como producto de este trabajo de grado, se decidió que se trabajaría con la categoría Tierra, ambiente y recursos naturales, estructurada de la siguiente forma:

Tierra, Ambiente y Recursos Naturales

Términos Específicos

ambiente
 ciencia ambiental
 ciencia del sistema terrestre
 ciencia del suelo
 ciencia y manejo de incendios
 ciencias atmosféricas
 ciencias del mar
 ecología
 geografía
 geología
 hidrología
 manejo de recursos naturales
 programas ambientales
 recursos naturales

English

Natural Resources, Earth and Environmental Sciences

Categorías de Asuntos

P Tierra, Ambiente y Recursos Naturales

Term Number

127305

Figura 2. Términos específicos asociados a Tierra, ambiente y recursos naturales¹⁶

¹⁵ Tomado del Tesoro de Agricultura: Gobierno, Ley y Regulaciones [En línea] Disponible en: <http://agclass.nal.usda.gov/mtwdk.exe?k=2007es&l=115&s=1&n=1&y=0&w=Gobierno%2C%20Ley%20y%20Regulaciones&t=3> (Consultado el 8 de septiembre de 2012)

¹⁶ Tomado del Tesoro de Agricultura: Tierra, ambiente y recursos naturales [En línea] Disponible en: <http://agclass.nal.usda.gov/mtwdk.exe?k=2007es&l=115&s=1&n=1&y=0&w=Tierra%2C%20Ambiente%20y%20Recursos%20Naturales&t=2> (Consultado el 8 de septiembre de 2012)

Con el ánimo de construir la metodología propuesta dentro de los objetivos de este trabajo, se realizó una comparación entre las estructuras básicas de los tesauros y las ontologías, de manera que a partir de la identificación de los elementos comunes y diferentes entre estos dos sistemas de organización del conocimiento, se pudo determinar cuáles son los aspectos que podrían vincularse con la conversión de los tesauros en ontologías, y que se tuvieron en cuenta en el planteamiento de la metodología que se presenta más adelante.

Teniendo en cuenta lo dicho anteriormente, la pregunta que se responde con el desarrollo de esta investigación es la siguiente:

¿Cuál sería la metodología más adecuada para posibilitar la convergencia de un sistema de organización del conocimiento, como el Tesoro de la NAL, hacia una ontología que contribuya con la búsqueda y recuperación de información por parte de los usuarios de fuentes relacionadas con agricultura y todas las temáticas asociadas?

2. JUSTIFICACIÓN

El desarrollo de metodologías que orienten la convergencia de tesauros hacia ontologías es un tema que ha empezado a ser desarrollado por teóricos en diferentes lugares del mundo, que se han enfocado en la creación de herramientas semánticas que faciliten y simplifiquen la búsqueda y recuperación de información por parte de los usuarios de diferentes sistemas de información.

Muestra de ello son las iniciativas que pudieron recuperarse cuando se revisó qué se había escrito sobre el tema. Entre los artículos que se revisaron, se encuentra el escrito por Popovici, Bogdan, Polceanu y Querrec¹⁷, en el que proponen que el desarrollo de ontologías en un entorno web representa una gran ventaja para los sistemas de información especializados en áreas particulares de conocimiento por dos razones: la primera, relacionada con la disminución de los costos de mantenimiento de los sistemas de organización del conocimiento que los respaldan, llámense tesauros, taxonomías y/o glosarios; y la segunda, asociada con la automatización que podría llegar a alcanzarse en materia de actualización de los términos con los que se recupera información en esos sistemas de información.

Teniendo en cuenta lo anterior, el desarrollo de este trabajo de grado se justifica a partir de los siguientes argumentos:

- Con el desarrollo de ontologías a partir de tesauros existentes, se está reutilizando y actualizando el conocimiento relacionado con un área específica del conocimiento, lo que beneficia directamente al usuario de la información asociada a estos sistemas de organización del conocimiento, porque no debe

¹⁷POPOVICI, Dorin-Mircea, et al. Applying of an ontology based modeling approach to cultural heritage systems. En: Advances in Electrical and Computer Engineering v. 11, no. 3 (2011) p. 105 - 110

reconocer nuevos términos sino identificar la forma como se hacen explícitas las relaciones entre estos.

- En el momento en que se generan ontologías y estas se integran con sistemas de información, como el que maneja la *National Agricultural Library*, se puede conseguir la unificación completa de todos los recursos ofrecidos por las unidades de información y con ello, favorecer el aprovechamiento de los mismos y la generación integral de conocimiento por parte de los usuarios de estas unidades.
- El desarrollo de una metodología que permita la conversión de un tesoro como el de la *National Agricultural Library* hacia una ontología, favorece el proceso de búsqueda y recuperación de información por parte de los investigadores y demás usuarios de estos sistemas de organización del conocimiento, al flexibilizar y hacer más visibles las relaciones existentes entre los conceptos que alojan.
- Al poseer una estructura jerárquica establecida y contar con un esquema de relaciones asociativas y de equivalencia bien establecidas, el tesoro de la *National Agricultural Library* puede relacionarse directamente con una ontología, en la medida en que ya tiene definidas las relaciones entre los términos que lo componen y estas puedan vincularse más fácilmente dentro de una ontología.

3. OBJETIVOS

3.1. OBJETIVO GENERAL

Desarrollar una metodología que posibilite la construcción de ontologías a partir de tesauros existentes, desde la identificación de las estructuras básicas de estos sistemas de organización del conocimiento, hasta los aspectos comunes que se presentan entre ellos, de manera que se contribuya con la recuperación de la información por parte de los usuarios en el entorno web.

3.2. OBJETIVOS ESPECÍFICOS

- Identificar y evaluar las estructuras y componentes básicos de los tesauros y las ontologías, para establecer sus semejanzas y diferencias.
- Diseñar la metodología a partir de la comparación de las estructuras básicas de los tesauros y las ontologías, teniendo en cuenta los indicadores que permitirán evaluar si un tesoro posee potencial para converger hacia una ontología.
- Evaluar un tesoro a la luz de los indicadores establecidos para determinar si la metodología que se plantea le permita converger hacia una ontología.
- Implementar la metodología propuesta para validarla.

4. LA INDIZACIÓN Y LOS SISTEMAS DE ORGANIZACIÓN DEL CONOCIMIENTO

Para establecer una metodología que permita la convergencia de un tesoro hacia una ontología, es necesario identificar primero las definiciones, las características y los componentes de cada uno de estos sistemas de organización de conocimiento, así como su relación con la indización, la búsqueda y la recuperación de información en los sistemas de información en los que son empleados.

4.1. Los procesos de descripción de contenidos: La indización

Se inicia esta revisión teórica discutiendo las relaciones que estos sistemas de organización del conocimiento tienen con la Ciencia de la Información, principalmente con la *descripción de contenidos*, proceso que puede definirse como el conjunto de actividades que hacen posible la representación condensada y normalizada de la información contenida en un documento¹⁸, sea cual sea su formato, de manera que esta pueda identificarse, ser clasificada, recuperada y usada por quien pueda necesitarla.

Las actividades que se vinculan con la descripción de contenidos pueden clasificarse de acuerdo con el tipo de análisis. Estos se centran en dos grupos, la descripción y catalogación, y el proceso de indización¹⁹, este último vinculado directamente con el uso de los lenguajes normalizados, en los que se enmarcan los tesauros y las ontologías, sistemas de organización del conocimiento en los que se concentra este trabajo de grado.

¹⁸ GIMENO PERELLÓ, Javier. Sistemas de indización aplicados en bibliotecas: Clasificaciones, tesauros y encabezamientos de materias. En: Tratado básico de Biblioteconomía. Madrid: Editorial Complutense, 2004. p. 200. ISBN: 84-7491-750

¹⁹ PINTO MOLINA, María. Análisis y síntesis del Proceso Documental (Consultado el 8 de septiembre de 2012) Disponible en: http://www.mariapinto.es/alfintra/contenidos/ANALISIS_SINTESIS.pdf

La indización, de acuerdo con el enfoque de Slype, puede definirse como el conjunto de actividades que se relacionan con la representación del contenido de los diferentes documentos que se analicen, a través del uso de lenguajes documentales normalizados que faciliten la recuperación de estas fuentes, mediante los catálogos de las unidades de información en los que se desarrollen²⁰.

Una de las actividades que se realizan en la indización se relaciona con el análisis del contenido de los documentos que se describieron y catalogaron previamente, en busca de los conceptos con los que puede representarse la información contenida en ellos²¹, de manera que se pueda indicar qué temáticas abordan y con qué áreas del conocimiento se encuentran relacionados.

Posterior a esta identificación, se realiza la traducción de estos conceptos a términos normalizados que conforman los lenguajes documentales que se encuentran relacionados con las temáticas identificadas en los documentos que se indizan²² de manera que, de acuerdo con los parámetros establecidos dentro del sistema de información que los aloja, estos términos puedan complementar los registros descriptivos de los documentos analizados previamente y garantizar la recuperación y uso de los mismos.

Para que los resultados de la indización sean los esperados y se favorezca la búsqueda y recuperación de información por parte de los usuarios que accedan a los catálogos de las unidades de información, este proceso debe contar con las siguientes características:

²⁰ SLYPE, Georges Van. Utilización de los lenguajes de indización. En: Los lenguajes de indización: Concepción, construcción y utilización en los sistemas documentales. Madrid: Fundación Germán Sánchez Ruipérez, 1991. p. 116. ISBN: 84-86168-60-0

²¹ PINTO MOLINA, María. Óp. Cit.

²² Idem.

- **Precisión:** Se puede medir la precisión a la hora de seleccionar los conceptos más importantes dentro de los documentos, de manera que se resalten los datos que resulten pertinentes para la recuperación de estas fuentes de información por parte de los usuarios²³.
- **Especificidad:** Se relaciona con el nivel de precisión en el momento en que se seleccionan los descriptores que representan los conceptos identificados en el documento que se analiza, teniendo en cuenta la jerarquía que forma parte dentro del lenguaje documental al que pertenece y la coherencia que se guarde con esta, a la hora de hacer uso de los términos dentro del sistema de información que da soporte a la unidad de información²⁴.
- **Concreción:** Durante el desarrollo del proceso de indización, se debe evitar el uso de términos genéricos que propicien la aparición de ambigüedades que pongan en riesgo la recuperación efectiva de información por parte de los usuarios, cuando intenten hacerlo a través de los catálogos de las diferentes unidades de información²⁵.

Para que la descripción de contenidos a partir de la realización de las actividades que se describieron anteriormente sea posible, se requiere de diferentes herramientas que les ofrezcan soporte como aquellas que plantean lineamientos, como las Reglas de Catalogación Angloamericanas, que ofrecen formatos que orientan su desarrollo, como el formato MARC, o que sean fuentes de información que las validen y normalicen, como los lenguajes documentales.

4.2. Los lenguajes documentales

²³ SLYPE, Georges Van. Óp.Cit. p. 123

²⁴ *Ibíd*em, p. 123 - 124

²⁵ PINTO MOLINA, María. Óp. Cit.

Teniendo en cuenta lo que se ha dicho hasta ahora de la descripción de contenidos y de las actividades con las que esta se encuentra vinculada, y volviendo sobre el objetivo que persigue este trabajo de grado, el siguiente paso en la construcción de una base teórica y conceptual, es la definición de los lenguajes documentales y de las características que poseen, haciendo énfasis en los tesauros y las ontologías.

Para empezar a hablar de los lenguajes documentales es necesario aclarar que por sus características y por los cambios que se han generado en el campo de la documentación a raíz de la evolución de las tecnologías de la información y de los cambios en las necesidades de los usuarios, existen diferentes perspectivas desde las que se pueden definir los lenguajes documentales.

De acuerdo con Slype, los *lenguajes documentales* son aquellos sistemas de signos que permiten la representación de los contenidos de cualquier fuente de información, de manera que se posibilite la recuperación y uso de cada una de estas fuentes²⁶, a través de un catálogo o un sistema de información, definición que coincide con la propuesta por Gil, quien los considera un “sistema artificial de signos normalizados”²⁷ que favorecen la recuperación de la información solicitada por los usuarios.

Desde otras perspectivas como la de García Gutiérrez, los *lenguajes documentales* son listas de términos que se vuelven vocabularios dinámicos que generan relaciones entre los términos que los componen, que son producto del análisis de los documentos y cuyo objetivo principal es controlar un conjunto de términos para dar respuesta a las consultas de los usuarios²⁸.

²⁶ SLYPE, Georges Van. Óp. Cit. p. 21

²⁷ GIL URDICIÁIN, Blanca. Listas de encabezamientos de materia. En: _____. Manual de lenguajes documentales. Gijón (Asturias): Ediciones Trea, 2004. p. 18. ISBN: 84-9704-138-0

²⁸ GARCÍA GUTIÉRREZ, Antonio citado por MAGÁN WALS, José Antonio. Óp. Cit. p. 221

Mientras que para C. Guinchat y M. Menou los *lenguajes documentales* son aquellos “lenguajes convencionales” que se usan en las unidades de información para describir documentos y que posibilitan su recuperación por parte de los usuarios²⁹.

Teniendo en cuenta las anteriores definiciones, los *lenguajes documentales* pueden definirse como sistemas dinámicos de signos normalizados que son producto del análisis y la comprensión de las diferentes áreas del conocimiento, que favorecen la descripción de los contenidos de los documentos y que facilitan la recuperación de estos últimos por parte de los usuarios de las unidades de información existentes.

Con el ánimo de beneficiar a los usuarios en sus procesos de búsqueda y recuperación de información, los objetivos que persiguen los lenguajes documentales son los siguientes:

- En la etapa de descripción documental, su objetivo es mediar entre los contenidos de los documentos y la indización, al contribuir con el control de la ambigüedad terminológica que pueda presentarse, así como con el análisis de los contenidos de estos documentos, al ofrecer términos unívocos que favorecen la identificación del sentido en que deben entenderse, para así poder representarlos³⁰.
- En las etapas de búsqueda y recuperación de información, su principal y más importante objetivo es convertirse en el canal de comunicación que permita al

²⁹ GUINCHAT, Claire y MENOUE, Michel citado por MARTÍN GAVILÁN, César. Lenguajes documentales. Principales tipos de clasificación. Encabezamientos de materia. Descriptores y tesauros. (Consultado el 8 de septiembre de 2012) Disponible en: <http://eprints.rclis.org/bitstream/10760/14817/1/lendoc.pdf>

³⁰Idem.

usuario interactuar con el sistema de información, para recuperar los documentos que pueden dar respuesta a sus necesidades³¹.

Teniendo claros los objetivos y los principios que orientan y guían los lenguajes documentales, es viable empezar a hablar de la forma como se clasifican. Para ello, es importante tener en cuenta que estos lenguajes pueden tipificarse de acuerdo con tres características, que son:

1- Control ejercido sobre el lenguaje

De acuerdo con esta característica, los lenguajes documentales pueden dividirse en dos grupos, los lenguajes que son libres y los que son controlados³²:

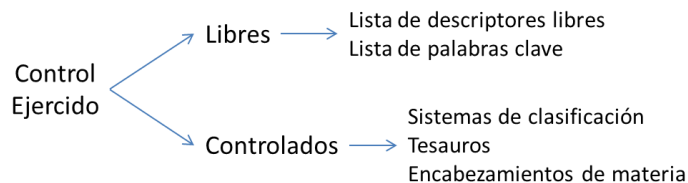


Figura 3. Clasificación de lenguajes documentales

Los *lenguajes libres* son aquellos como las listas de palabras clave o de términos libres, que se alimentan y desarrollan continuamente, a medida que se realiza el proceso de indización, y que se caracterizan por contener términos ambiguos y presentar problemas de sinonimia y homonimia³³.

Este tipo de lenguajes documentales contribuyen con la recuperación de un alto número de documentos a través de la combinación de varios de los

³¹ SLYPE, Georges Van. Óp.Cit. p. 21

³² GIL URDICIAIN, Blanca. Óp. Cit. p. 22 - 25

³³ Idem.

términos que contiene, facilitando la construcción de estados de arte y marcos teóricos, durante el desarrollo de las primeras etapas de la investigación.

Por otro lado, *lenguajes controlados* son aquellos que cumplen con los principios de los lenguajes documentales mencionados anteriormente, relacionados con la reducción de la redundancia terminológica y con la univocidad de cada uno de los términos que contienen, para favorecer la recuperación efectiva de la información por parte del usuario³⁴.

2- Coordinación de los términos

Si se tiene en cuenta el momento en que se combinan los términos que componen los lenguajes documentales, se identifica otra forma de clasificar estos lenguajes:

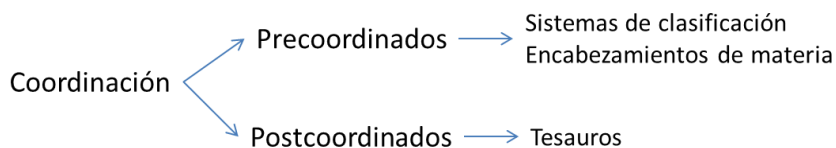


Figura 4. Lenguajes documentales de acuerdo con la coordinación entre sus términos

Los lenguajes precoordinados son aquellos cuyos términos presentan relaciones sintácticas y se combinan gramaticalmente en el momento en que se realiza la descripción del contenido de los documentos que se analizan³⁵, favoreciendo el aumento de la precisión en la búsqueda hecha por los usuarios, al disminuir la aparición de documentos irrelevantes en la recuperación de información.

³⁴Idem.

³⁵ SLYPE, Georges Van. Óp. Cit. p. 23

Los lenguajes coordinados o poscoordinados carecen de sintaxis y se caracterizan por contener términos simples o unitérminos que se combinan en el momento de la recuperación de la información a través de relaciones lógicas booleanas propuestas por el usuario. Pueden propiciar la recuperación de documentos irrelevantes como resultado de una combinación errada de términos de búsqueda³⁶.

3- Estructura del lenguaje

De acuerdo con la estructura que presentan los lenguajes documentales, estos pueden clasificarse en tres grupos:

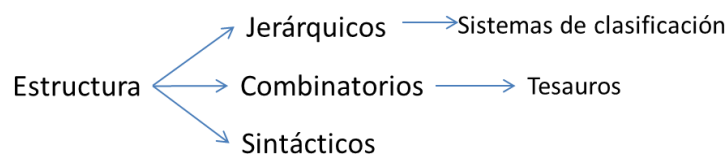


Figura 5. Clasificación de los lenguajes documentales de acuerdo con su estructura

Para empezar, los lenguajes documentales jerárquicos son los que organizan el conocimiento en categorías correspondientes con las áreas en las que puede ser dividido³⁷, que representan las relaciones entre los términos con estructuras arborescentes cuyas partes pueden ser identificadas por medio de códigos numéricos o alfanuméricos.

Por otro lado, los lenguajes con estructura combinatoria son aquellos que facilitan la combinación de los términos de acuerdo con las necesidades específicas de las unidades de información en las que es aplicado durante la

³⁶ GIL URDICIÁIN, Blanca. Óp. Cit. p. 22 - 25

³⁷ Ídem.

indización del conocimiento contenido en los diferentes recursos bibliográficos³⁸.

Finalmente, los lenguajes documentales con estructura sintáctica son aquellos que representan los contenidos de los documentos a través de la sintaxis terminológica³⁹, que contribuye con la visualización de las relaciones que se presentan entre los contenidos de distintos documentos, en el momento en que se realizan búsquedas de información asociadas a dichos recursos.

Teniendo claridad con respecto a la clasificación de los lenguajes documentales de acuerdo con las características de forma y fondo que se mencionaron anteriormente, resulta más sencillo explicar algunos de los lenguajes documentales que se conocen en la actualidad. Entre ellos están:

- Listas de encabezamientos de materia: Son lenguajes precoordinados de estructura combinatoria que reúnen en un listado alfabético, un conjunto de términos que representan los contenidos de los documentos que se analizan.

Las listas de encabezamientos de materia se presentan en un solo idioma y plantean relaciones jerárquicas, *Términos generales* y *Términos específicos*, relaciones asociativas, *Términos relacionados*, y relaciones de equivalencia entre los términos que las componen, a partir del uso de las referencias de *Use o Véase* y *Usado por*⁴⁰.

- Sistemas de clasificación decimales: Pueden definirse como lenguajes documentales que dividen el conocimiento en grandes áreas representadas por

³⁸ Ídem.

³⁹ Ídem.

⁴⁰ Ibídem, p. 31 - 32

números decimales que, divididas en diferentes sub-áreas, enriquecen la descripción de cada una de estas áreas, facilitando su comprensión⁴¹.

Estos lenguajes documentales cuentan con características que los diferencian claramente de los demás. Entre ella se encuentran su naturaleza generalizadora porque pretenden abarcar todas las áreas del conocimiento existentes, y su flexibilidad, porque permiten representar conceptos desde diferentes perspectivas.

Además, estos lenguajes documentales son explícitos en la medida en que desglosan las áreas de conocimiento hasta donde les es posible, siendo muy puntuales en las categorías y los términos que emplean, representados todos por conjuntos de símbolos numéricos que además de identificar al documento clasificado, permite ubicarlo dentro de la unidad de información⁴².

Finalmente, estos lenguajes se caracterizan por ser expandibles y modificables en la medida en que pueden actualizarse y transformarse de acuerdo con los avances que se presenten en las diferentes áreas del conocimiento, sin que esto comprometa su integridad o su funcionalidad⁴³.

- Listas de palabras clave: Son lenguajes documentales que reúnen un conjunto de términos no normalizados y escritos en lenguaje natural, organizados en orden alfabético, que representan los conceptos e ideas claves desarrolladas en los documentos que se analizan en una unidad de información cualquiera⁴⁴.

En estos listados se encuentran unitérminos compuestos por una sola palabra sin ningún tipo de control, que fueron extraídos de documentos analizados y

⁴¹ GIMENO PERELLÓ, Javier. Óp. Cit. p. 211 - 213

⁴² Idem.

⁴³ Idem.

⁴⁴ SLYPE, Georges Van. Óp. Cit. p. 23

que presentan problemas serios de ambigüedad y redundancia⁴⁵, al extraerse del contexto en el cual fueron identificados.

Este tipo de lenguaje documental representa un gran desafío para el usuario que quiere realizar una búsqueda de información puesto que puede llegar a incluir términos creados por los autores de los documentos de los que se alimenta. Los términos se refieren a temas muy puntuales y que resultan muy difíciles de reconocer por parte de los usuarios, o pueden incluir un término que representa muchas ideas a la vez y resulta imposible recuperar información más precisa, vinculada con las necesidades de los usuarios⁴⁶.

4.3. Los tesauros

Son lenguajes documentales especializados en áreas particulares del conocimiento que agrupan términos relacionados semánticamente⁴⁷ y que facilitan los procesos de descripción, búsqueda y recuperación de la información dentro de los catálogos de las diferentes unidades de información.

Se caracterizan por contar con una estructura combinatoria jerárquica y por aplicar un amplio control terminológico que además de traducir conceptos expresados en lenguaje natural a un vocabulario técnico, contribuye con la recuperación de información por parte de los usuarios.

Adicionalmente, los tesauros son considerados vocabularios controlados dinámicos que permiten que los usuarios que hagan uso de ellos puedan realizar las combinaciones necesarias para representar un concepto particular dentro de un área de conocimiento⁴⁸, que si bien no es desarrollada con la misma

⁴⁵Idem.

⁴⁶Ibidem, p. 27 - 28

⁴⁷GIL URDICIAIN, Blanca. Óp. Cit. p. 183 - 184

⁴⁸Idem.

profundidad que los sistemas de clasificación, si reúne los aspectos principales que son de interés para el usuario que lo consulte.

Como consecuencia de sus características, los vocabularios controlados como los tesauros deben cumplir con las siguientes funciones⁴⁹:

- *Normalización del vocabulario*, relacionada con la eliminación de la sinonimia y polisemia en pro del cumplimiento de los principios de univocidad y coherencia explicados anteriormente.
- *Inducción*, vinculada con el ofrecimiento de una perspectiva más amplia de un término buscado, que incluye términos genéricos, específicos y asociados.
- *Representación*, relacionada con la coherencia que existe entre los términos y los conceptos con los que se vinculan, en beneficio del usuario.

En cumplimiento de las funciones mencionadas en los puntos mencionados anteriormente, los tesauros cuentan con dos componentes principales que pueden definirse de la siguiente forma:

- Unidades léxicas: Representan los conceptos asociados al área de conocimiento en la que se especializa el tesoro. Pueden ser de dos tipos:
 - o Descriptores: Términos autorizados dentro del tesoro para representar un concepto particular dentro de un campo de conocimiento, que se usan para indizar y recuperar documentos⁵⁰.

⁴⁹Ibidem, p. 184 - 185

⁵⁰ SLYPE, Georges Van. Óp.Cit. p. 37

- No-descriptores: Términos equivalentes a los descriptores que se cancelan, porque propician problemas relacionados con la sinonimia, cuando se combinan con los descriptores. Estos términos presentan referencias a las formas autorizadas de sí mismos, para orientar al usuario en su búsqueda⁵¹.
- Relaciones semánticas: Son los vínculos que conectan a los términos y los estructuran de manera que puedan ser entendidos en un contexto y en un sentido particular, de manera que se simplifique su uso.

Estas relaciones pueden ser de cuatro o cinco tipos, de acuerdo con la característica monolingüe o multilingüe del tesaurus, estas son:

- Pertenencia a un grupo: Es una relación asimétrica que se presenta entre los descriptores y la temática a la que representan, dentro de un área particular del conocimiento. Puede ser mono-jerárquico, en la que un término solo puede pertenecer a una temática, o poli-jerárquico, en la que un descriptor puede pertenecer a dos o más temáticas al tiempo⁵².
- Equivalencia interlingüística: Es el tipo de relación que se presenta en los tesaurus multilingües cuando se crea una equivalencia entre términos de diferentes lenguas que representan una misma idea o concepto, y que favorece la búsqueda de información en diferentes idiomas⁵³.
- Equivalencia semántica: Es la relación que se presenta entre descriptores y no-descriptores, que busca la univocidad en el momento de representar

⁵¹ Ídem.

⁵² Ibídem, p. 44 - 45

⁵³ Ibídem, p. 45 - 46

conceptos o ideas particulares, dentro del área de conocimiento en la que se especializa el tesoro⁵⁴.

- Jerarquía: Es la relación que se observa entre términos relacionados con una temática particular dentro de un área de conocimiento. En este tipo de relación se puede evidenciar el nivel de generalidad y especialización que alcanza un descriptor frente a otros descriptores, y el nivel de profundidad con el que se aborda un tema dentro del tesoro⁵⁵.
- Asociación: Es la relación que se presenta entre descriptores que si bien no se encuentran vinculados semántica ni jerárquicamente, pueden relacionarse por la cercanía de los conceptos que representan o por la incidencia que puede tener un descriptor en otro, al ser analizado en un contexto particular⁵⁶.

A pesar de que los dos componentes de los tesauros mencionados anteriormente son los más reconocidos y aceptados por los autores, y los que mayor desarrollo teórico presentan, existe un tercer componente que vincula al usuario de una manera más directa con este lenguaje documental, al indicarle algunos aspectos que facilitarán la interacción que tenga con los tesauros.

El tercer componente al que se hace referencia son las *notas de alcance*, ya sean bibliográficas, de definición o históricas, cuyas funciones principales se relacionan con el ofrecimiento de toda la información posible que permita la identificación plena de los términos del tesoro, en el contexto del lenguaje como tal y del sistema o unidad de información con el que se encuentran vinculados.

⁵⁴ Ibídem, p. 47 - 48

⁵⁵ Ibídem, p. 52 - 53

⁵⁶ Ibídem, p. 57 - 58

Las notas de alcance bibliográficas son aquellas que dan cuenta de las fuentes de información que se emplearon para la proposición y normalización de los términos del tesoro, así como aquellas que se consultaron para construir las notas de alcance que hacen parte del vocabulario controlado.

En el caso de las notas de alcance de definición, se puede decir que son notas que contienen definiciones asociadas a los términos contenidos por el tesoro, la identificación de los descriptores que son incluidos o excluidos en este vocabulario y de las relaciones que existen entre ellos, así como la indicación de instrucciones adicionales de uso de un término dentro del tesoro⁵⁷.

Teniendo claridad con respecto a las formas en que deben usarse las definiciones de descriptores dentro de un tesoro desde el desarrollo de notas de alcance, vale la pena resaltar un último aspecto que desde la perspectiva de Del Valle Gastaminza resulta importante a la hora de desarrollar dichas notas:

“La información incluida en una nota de alcance debe ser provechosa a los usuarios del tesoro, en este sentido, debe agregar información adicional a la que el término dice ya por sí mismo. Repetir el término o dar una definición obvia de un término inequívoco no es provechoso. Recuérdese que un tesoro no es un diccionario, una enciclopedia, o aún un índice.”⁵⁸

Otro tipo de nota que se asocia con los tesauros es la *nota histórica*, que se encarga de dar cuenta de la evolución terminológica que precede a los términos del tesoro, incluyendo las expresiones que se han remplazado⁵⁹, así como las

⁵⁷ DEL VALLE GASTAMINZA, Félix. Óp. Cit.

⁵⁸ Ídem.

⁵⁹ TEMATRES. Gestión de términos. Notas. (Consultado el 8 de septiembre de 2012) Disponible en: <http://r020.com.ar/tematres/manual/notas.html>

fechas en las que se fusionó con otro término del tesoro o en las que se determinó que entró en desuso.

4.4. Los sistemas de organización del conocimiento y el entorno digital

Con el desarrollo y evolución de los sistemas de información tradicionales de la mano de las tecnologías de la información, se hizo necesario propiciar el desarrollo de sistemas de organización de conocimiento que respondieran a las necesidades de los usuarios de información en el entorno digital, razón por la que los tesauros, las listas de encabezamientos de materia y otros sistemas de este tipo, fueron transferidos al medio digital, con el objeto de apoyar los procesos de búsqueda y recuperación de información por parte de dichos usuarios.

Sin embargo, con el avance de la web surgió la necesidad de crear sistemas de organización del conocimiento, que trabajando en conjunto con los sistemas existentes, faciliten la búsqueda de información dentro de ese gran cúmulo de documentos y recursos que se encuentran disponibles en la red y a los que debe enfrentarse un usuario, siempre que quiera recuperar información de calidad que responda a sus necesidades e inquietudes particulares.

En respuesta a esta nueva necesidad, comunidades de científicos de la información y de la computación, crearon sistemas expertos que facilitarían la búsqueda de los recursos que satisficieran las necesidades de los usuarios de la información en el entorno digital a partir del aprovechamiento de nuevos sistemas de organización del conocimiento como las ontologías.

Con el desarrollo de estos nuevos sistemas de organización del conocimiento y su implementación en los entornos digitales, la información ha podido ser estructurada de manera tal, que diferentes fuentes de información relacionadas con un dominio particular se encuentran vinculadas y contribuyen con la

recuperación de los recursos para dar respuesta a los interrogantes de los usuarios.

4.5. Las ontologías como sistemas de organización del conocimiento en el entorno web

Los sistemas de organización del conocimiento actuales se preocupan por contribuir con el desarrollo de la web semántica, aportando estructuras que fortalezcan la recuperación de información por parte de los usuarios, e incluyendo características que hacen que este proceso de búsqueda sea más efectivo, partiendo de las expresiones cotidianas de los usuarios, en lenguaje natural.

Uno de los sistemas de organización de conocimiento que ha tenido desarrollo en el contexto de la web semántica ha sido la ontología, un instrumento que, al igual que los sistemas de organización tradicionales, contribuye con la indización de los contenidos digitales que a diario se publican en dicho contexto.

Para hablar de ontologías, se debe empezar por identificar las diferentes definiciones inherentes a ellas, ya que de acuerdo con la literatura revisada, las ontologías han sido trabajadas desde dos enfoques: el ofrecido por los planteamientos filosóficos y otro, relacionado con la inteligencia artificial y la informática.

Desde la filosofía, la ontología puede ser comprendida como el estudio del ser que se concentra en el análisis de cada uno de sus componentes, explícitos e implícitos, con el ánimo de poder describir su esencia y explicarla a quienes estén interesados en conocer al ser estudiado en profundidad, teniendo en cuenta los aspectos que pueden identificarse a simple vista y aquellos que se encuentran ocultos a ella⁶⁰.

⁶⁰ FERNÁNDEZ BREIS, Jesualdo Tomás. Estado del arte. En: _____. Un entorno de integración de ontologías para el desarrollo de sistemas de gestión de conocimiento. 2003. p. 29 – 32. Tesis doctoral

Teniendo como base este planteamiento filosófico relacionado con las ontologías, inteligencia artificial y en general la informática y las ciencias de la computación las definen de las siguientes formas:

De acuerdo con Neches, *et al*, las ontologías son herramientas que se encargan de definir la terminología asociada a un dominio de conocimiento específico, así como las relaciones que se desarrollan en su interior, de manera que se pueda construir un vocabulario que describa dicho dominio en su totalidad, con un tiempo de construcción menor al de otros sistemas de organización del conocimiento y que favorece directamente la estructuración de sistemas de información capaces de soportar grandes cantidades de información⁶¹.

Para Tom Gruber, una ontología debe entenderse como la representación formal de conceptos que describen todo aquello que existe en una determinada área de conocimiento, a partir de la identificación de términos que los reflejan, las relaciones existentes entre ellos y las limitaciones que restringen la interpretación que se le dé, en el entorno de los sistemas de información en los que se utilice⁶².

Mientras que para Nicola Guarino, reconocido por criticar las posturas tradicionales de la inteligencia artificial representadas en su mayoría por Gruber, las ontologías van más allá de la simple conceptualización de una base de conocimiento definida. Para él, las ontologías deben ser estudiadas desde la interdisciplinariedad teniendo en cuenta la propuesta filosófica, la de la inteligencia artificial y la lingüística.

(Ciencia de la computación e inteligencia artificial) Universidad de Murcia. Departamento de Ingeniería de la Información y las Comunicaciones. (Consultado el 20 de octubre de 2012) Disponible en: <http://hdl.handle.net/10803/10921>

⁶¹ NECHES ROBERT, *et al*. Enabling technology for knowledge sharing. En: A.I. Magazine, v. 12, no. 3 (jul.-sept., 1991) p. 37. ISSN: 0738-4602 (Consultado el 20 de octubre de 2012) Disponible en: <http://www.aaai.org/ojs/index.php/aimagazine/article/view/902/820>

⁶² GRUBER, Thomas. A translation approach to portable ontology specification. En: Knowledge acquisition, v. 5, no. 2 (abril, 1993) p. 199. ISSN: 1042-8143 (Consultado el 20 de octubre de 2012) Disponible en: <http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>

En este sentido, para Guarino las ontologías son conceptos organizados jerárquicamente y relacionados entre sí, que intentan describir a las entidades que se vinculan con áreas particulares de conocimiento haciendo uso de expresiones controladas del lenguaje natural, que contienen reglas que definen la forma como se relacionan y el sentido en que deben ser comprendidas, por los sistemas basados en la gestión de conocimiento y por las personas que los usen⁶³.

Teniendo en cuenta lo que se ha dicho hasta el momento, las ontologías pueden definirse como la descripción formal de conceptos asociados a una determinada área de conocimiento, que se encargan de describirla y detallarla a través de términos que los representan, de manera tal que contribuya con la recuperación de información por parte de los usuarios.

Además de apoyar los procesos de búsqueda y recuperación de información en el entorno digital, las ontologías cumplen otras funciones vinculadas con la estructuración de la información de manera que esta pueda ser usada por personas y sistemas expertos para la generación de conocimiento, en respuesta a necesidades puntuales de información⁶⁴.

En consecuencia, las ontologías se encargan de analizar toda la información relacionada con un dominio particular de conocimiento y de sintetizarlo a partir de la identificación de los conceptos más importantes y de las relaciones que existen entre ellos⁶⁵, para crear estructuras que favorecen la recuperación de información digital disponible en los sistemas de información actuales.

⁶³ GUARINO, Nicola. Formal Ontology, Conceptual Analysis and Knowledge Representation. En: International Journal of Human-Computer Studies, v. 43, no. 5/6 (nov., 1995) p. 625 – 627. ISSN: 1071-5819 (Consultado el 20 de octubre de 2012) Disponible en: <http://www.loa.istc.cnr.it/Papers/FormOntKR.pdf>

⁶⁴ LUCAS DA SILVA, Daniela. ROCHA SOUZA, Renato y BARCELLOS ALMEIDA, Mauricio. Ontologias e vocabularios controlados: comparação de metodologias para construção. En: Ciência da Informação, v. 37, no. 3(sept.-dic., 2008) p. 60. ISSN: 0100-1965

⁶⁵ Ídem.

Adicionalmente, en caso de ser un producto derivado de sistemas de organización del conocimiento preexistentes, una ontología cumple con la función de reutilizar todas las estructuras que estuvieron asociadas a esos sistemas⁶⁶ y que requieren ser actualizadas para dar respuesta a los nuevos interrogantes que tienen los usuarios que los utilizan.

Para concluir con los objetivos y las funciones que cumplen las ontologías en el entorno actual, es necesario resaltar la importancia que estos sistemas de organización del conocimiento dan al lenguaje natural⁶⁷, en la medida en que tienen en cuenta expresiones cotidianas para representar conceptos del dominio definido, dando una mayor flexibilidad al proceso de búsqueda y recuperación de información en la web, por parte de los usuarios.

Siguiendo con la explicación de ontologías que se ha hecho hasta ahora, se debe prestar atención a los componentes particulares de estos sistemas de organización del conocimiento que posibilitan su construcción y aseguran que desde su estructura, la información esté bien representada en términos coherentes con los conceptos que quieren comunicarse en las ontologías y que explicitan el área de conocimiento específico en la que se especializan.

Los componentes de los que se habla en el párrafo anterior son los siguientes:

- Clases o conceptos

Las clases o conceptos son consideradas el centro de las ontologías⁶⁸ en la medida en que constituyen la base de conocimiento de la que se habla y sobre la que se trabaja en estos sistemas de organización del conocimiento.

⁶⁶ NOY, Natalia F. y McGUINNESS, Deborah L. Desarrollo de Ontologías-101: Guía para crear tu primera ontología. Stanford University: Stanford, 2005. p. 3(Consultado el 29 de octubre de 2012) Disponible en: http://protege.stanford.edu/publications/ontology_development/ontology101-es.pdf

⁶⁷ LUCAS DA SILVA, Daniela. ROCHA SOUZA, Renato y BARCELLOS ALMEIDA, Mauricio. Óp. Cit. p. 60

⁶⁸ NOY, Natalia F. y McGUINNESS, Deborah L. Óp. Cit. p. 5

Las clases son las descripciones que los autores de la ontología hacen de los conceptos del dominio o universo en el que se especializa este KOS, y que son objeto de la clasificación y la estructuración terminológica que puede evidenciarse al interior de una ontología⁶⁹.

Estas clases a su vez cuentan con subclases que las detallan y que permiten alcanzar un nivel mayor de especificidad en relación con el universo al que pertenecen, ofreciendo más posibilidades al usuario de recuperar la información particular cuando hacen uso de una ontología⁷⁰.

- Propiedades o atributos (*slots*)

Pueden definirse como aquellos elementos que permiten la descripción de las clases y subclases que se encuentran en una ontología⁷¹, en este sentido, son los aspectos que contribuyen con la definición de cada concepto que se representan en las clases y subclases que constituyen las ontologías.

- Instancias

Cuando se habla de los atributos que definen a las clases y subclases presentes en las ontologías, se hace necesario hablar de las instancias debido al papel que cumplen dentro de estos sistemas de organización del conocimiento.

Las instancias son los ejemplos o los objetos⁷² que dan cuenta de las características o propiedades que definen a las diferentes clases y subclases

⁶⁹ LUCAS DA SILVA, Daniela. ROCHA SOUZA, Renato y BARCELLOS ALMEIDA, Mauricio. Óp. Cit. p. 61

⁷⁰ Ídem.

⁷¹ GARCÍA JIMÉNEZ, Antonio. Instrumentos de representación del conocimiento: tesauros versus ontologías. En: Anales de documentación, v. 7 (2004) p. 83 (Consultado el 30 de octubre de 2012) Disponible en: <http://revistas.um.es/analesdoc/article/view/1691/1741>

⁷² Ídem.

que se encuentran presentes en las ontologías y que ubican al usuario en el universo del que se habla en estos sistemas de organización del conocimiento.

- Facetas

Luego de revisar los componentes de los tesauros que se expusieron en la primera parte de este marco teórico, es necesario decir que cuando se habla de facetas dentro de una ontología, se habla de las limitaciones que tienen los conceptos representados en las ontologías.

Las facetas o restricciones de rol identifican, de acuerdo con García Jiménez, “los tipos de valores, los valores permitidos, el número de valores”⁷³ y todas las características adicionales que puedan tener los atributos o slots de las clases contenidas en las ontologías.

- Las relaciones entre las clases

Son aquellas interacciones que pueden identificarse entre los conceptos que se encuentran en las ontologías y que se asocian con la clasificación taxonómica que estructura a estos sistemas de organización del conocimiento, al tiempo que dan cohesión a las clases que describen un dominio particular del conocimiento⁷⁴.

Teniendo claridad con respecto a los principales componentes estructurales de una ontología, se puede empezar a hablar de las metodologías que existen para su construcción y de los diferentes aspectos que deben tenerse en cuenta a la hora de plantear la generación de un sistema de organización de conocimiento de este tipo.

⁷³ Ídem.

⁷⁴ LUCAS DA SILVA, Daniela. ROCHA SOUZA, Renato y BARCELLOS ALMEIDA, Mauricio. Óp. Cit. p. 61

Para empezar, es necesario aclarar que no existe un patrón metodológico para la creación de ontologías. De acuerdo con Noy y McGuinness, no existe un modelo correcto para generarlas debido a que su estructuración depende de la visión y los objetivos que tengan quienes quieren construir un sistema de organización del conocimiento de este tipo, y de la amplitud y profundidad que quieran darle a la ontología⁷⁵.

En concordancia con esto, se pueden identificar diferentes metodologías que se usan para generar ontologías, entre las que se encuentran las siguientes:

- Método CYC:

El método CYC fue propuesto por Douglas Lenat y Ramanathan V. Guha en 1990 para la creación de una base de conocimiento que ofrece una representación formal de diferentes áreas del conocimiento, a través del establecimiento de reglas que posibilitan la descripción de diferentes elementos que se ven involucrados en las actividades cotidianas humanas⁷⁶.

De acuerdo con esta metodología los pasos que deben seguirse para construir una ontología son⁷⁷:

- 1) Extracción del conocimiento vinculado con un objeto de estudio particular, de las fuentes de información disponibles.
- 2) Procesamiento del conocimiento extraído a partir de herramientas de análisis que formalicen el lenguaje natural y lo hagan comprensible por los computadores.

⁷⁵ NOY, Natalia F. y McGUINNESS, Deborah L. Óp. Cit. p. 6

⁷⁶ FERNÁNDEZ BREIS, Jesualdo Tomás. Óp. Cit. p. 62

⁷⁷ Ídem.

3) Adquisición de nuevo conocimiento presente en el sentido común, con los computadores configurados para ello.

- Metodología de Gruninger y Fox:

Propuesta por Michael Gruninger y Mark Fox en 1995 como producto de su investigación y desarrollo vinculado con la inteligencia artificial, esta metodología pretendía propiciar el desarrollo de un modelo adaptable al entorno empresarial, que a partir del tratamiento del conocimiento compartido, pudiera dar respuesta a las preguntas generadas por este sector de la economía en relación con temas específicos⁷⁸.

La metodología Gruninger y Fox propone las siguientes etapas en el desarrollo de ontologías⁷⁹:

- 1) Identificación de los problemas y situación actual del sector empresarial en el que se quiere trabajar, para la generación de escenarios que simulen sus características.
- 2) Definición de los requisitos que debe cumplir la ontología a partir de la especificación de los aspectos formales con los que debe contar este sistema de organización del conocimiento.
- 3) Generación de la taxonomía que estructura la ontología asociada con el escenario creado, con lenguajes formales que representan expresiones del lenguaje natural, en términos de la lógica de primer orden, o la lógica de predicados.

⁷⁸ LUCAS DA SILVA, Daniela. ROCHA SOUZA, Renato y BARCELLOS ALMEIDA, Mauricio. Óp. Cit. p. 70

⁷⁹ Idem.

- 4) Representación formal de los problemas identificados en la primera etapa, de acuerdo con la taxonomía planteada en el paso inmediatamente anterior.
- 5) Establecimiento de los axiomas que limitan la interpretación de los términos empleados para la descripción de los problemas representados.
- 6) Verificación del cumplimiento de las condiciones con las que debe contar la ontología para dar respuesta a las preguntas generadas por los usuarios de este sistema de organización del conocimiento.

- Método de Uschold y King

Planteado por Mike Uschold y Martin King en 1995 como resultado de su participación en el proyecto Enterprise Ontology del Instituto de Aplicaciones en Inteligencia Artificial de la Universidad de Edimburgo⁸⁰. Este método considera importante realizar las siguientes actividades para construir una ontología⁸¹:

- 1) Identificación del propósito que persigue la ontología o de la necesidad que se presenta y a la que se quiere dar respuesta, para así definir el grado de formalidad que tendrá (del lenguaje natural al uso de las expresiones lógicas más avanzadas)
- 2) Construcción de la ontología
 - a. Captura de la ontología, a partir de la identificación con los conceptos y relaciones clave en el dominio de conocimiento en el que se esté

⁸⁰ Idem.

⁸¹ USCHOLD, Mike y KING, Martin. Towards a methodology for building ontologies. Artificial Intelligence Applications Institute. University of Edimburgh, 1995. p. 2 – 4 (Consultado el 3 de noviembre de 2012) Disponible en: http://www1.cs.unicam.it/insegnamenti/reti_2008/Readings/Uschold95.pdf

trabajando, la desambiguación de las definiciones y asociaciones seleccionadas y la coherencia entre ellos.

b. Codificación de la ontología, o representación de los conceptos y relaciones seleccionadas a partir de lenguajes formales de representación, que hagan comprensible dicha conceptualización por los computadores.

c. Integración con ontologías existentes que manejen temáticas relacionadas o similares con la desarrollada por la ontología que se está construyendo, a partir de la migración de las representación conceptuales que contienen, haciendo uso de los lenguajes y estándares a los que se acogen, en aquellos casos en los que esto sea posible.

3) Evaluación de la ontología construida a partir de la revisión de sus estructuras semánticas, el cumplimiento de los requerimientos mínimos de la ontología, las preguntas de competencia a las que debe dar respuesta, así como la calidad de la representación formal que posee.

4) Documentación de la ontología creada, en la que se redacten manuales o guías que especifiquen la forma en que debe ser entendida la ontología, su funcionamiento, forma de uso, componentes y forma de actualización.

- Método 101

Esta metodología, resultado del análisis de experiencias vinculadas con el desarrollo de ontologías y planteada por Natalia F. Noy y Deborah L.

McGuinness⁸², propone que una ontología debe ser construida a partir de las siguientes etapas:

- 1) Determinación del dominio y alcance que limita a la ontología.
- 2) Revisión de ontologías existentes, en caso que se pueda recuperar alguna, y el estudio de la posibilidad de reutilizarla.
- 3) Identificación de los términos relevantes relacionados con el dominio de conocimiento definido para la ontología.
- 4) Definición de los conceptos claves y de las clases que los representan, así como su jerarquización y clasificación.
- 5) Determinación de las propiedades y atributos de las clases (slots), a través de las cuales podrán asociarse con las demás clases que se encuentran en la ontología.
- 6) Identificación de las restricciones (facetas) que contribuyen con la definición y limitación de cada una de las clases alojadas en la ontología.
- 7) Revisión e inclusión de los términos que pueden ejemplificar las diferentes clases que hacen parte de la ontología, para darles un mayor nivel de detalle y especificidad.

Al revisar estas metodologías, seleccionadas por presentar cuatro enfoques diferentes para orientar la construcción de ontologías (desde la inteligencia artificial, los lenguajes documentales y la administración), se pueden concluir los siguientes aspectos:

⁸² NOY, Natalia F. y McGUINNESS, Deborah L. Óp. Cit. p. 6 - 14

- Si bien las cuatro metodologías explicadas anteriormente establecen criterios diferentes para el desarrollo de ontologías, se puede observar que todas culminan con la generación de este tipo de sistemas de organización del conocimiento.

- Existen varios caminos desde los que se puede iniciar la construcción de ontologías como puede comprobarse si se comparan las metodologías, ya que los autores de las mismas proponen:
 - La reutilización de ontologías existentes.
 - La construcción desde cero de la conceptualización que las estructura, empezando con la captura de las expresiones del lenguaje natural que mejor representen el dominio en el que se especializan.
 - La creación de la taxonomía que las soporten con la ayuda de lenguajes documentales ya creados.
 - La formalización de los conceptos a través de herramientas tecnológicas, como los puntos de partida para creación de ontologías.

- Al analizar los procedimientos propuestos por las cuatro metodologías planteadas anteriormente, se pueden establecer algunos aspectos comunes, vinculados con:
 - La importancia de la revisión de fuentes de información de todo tipo, del área del conocimiento en la que se centrará la ontología, ya sea para la extracción de los conceptos que resultan clave para lograr su completa representación o para reusar las estructuras taxonómicas que tengan y puedan ser la base de desarrollo de nuevas ontologías.

- El procesamiento de la información asociada con el área del conocimiento en la que se especializa la ontología ya sea con el uso de lenguajes documentales existentes para la normalización de las expresiones del lenguaje natural, la intervención de tecnologías de información para su formalización y posterior comprensión por computadores, o la reutilización de ontologías relacionadas.
- El establecimiento de estructuras taxonómicas conceptuales que moldeen la ontología y la hagan usable en los procesos de descripción, búsqueda y recuperación de la información, asociados a este tipo de sistema de organización del conocimiento.

4.6. Las ontologías y los lenguajes documentales

El constante aumento del volumen de información disponible en entornos digitales ha hecho que se promueva la construcción y el desarrollo de sistemas de organización de conocimiento que, además de contribuir con la descripción e indización de los contenidos que se publican, son fuentes de conocimiento comprensibles por los sistemas de información en los que son implementados para contribuir con la recuperación efectiva de información por parte de los usuarios.

Para que una ontología pueda cumplir con estas dos funciones, como herramienta para indizar y como fuente de conocimiento, debe contar con algunas características que garanticen la recuperación de información en determinadas áreas de conocimiento, aspecto que de acuerdo con García Jiménez⁸³, vincula a las ontologías con los lenguajes documentales tradicionales que se conocen.

⁸³ GARCÍA JIMÉNEZ, Antonio. Instrumentos de representación de conocimiento: Tesoros versus Ontologías. En: Anales de Documentación, v. 7 (2004) p. 87 – 88

De acuerdo con este autor, existen tres aspectos que evidencian que las ontologías, como sistemas de organización del conocimiento, son la evolución de los lenguajes documentales tradicionales que se mencionaron al inicio del marco teórico, al tiempo que constituyen herramientas diferentes. Estos son:

1. El uso del lenguaje natural en la representación de conceptos

Los lenguajes documentales como los tesauros y las listas de encabezamiento de materias y las ontologías tienen como base terminológica a las expresiones sustantivadas del lenguaje natural para representar los conceptos que contienen⁸⁴.

En el caso de las ontologías, y como muestra clara de la evolución que representan, además de contar con expresiones sustantivadas, incluyen otro tipo de palabras como los adverbios, los prefijos, los sufijos, los adjetivos y los verbos⁸⁵, situación que da cuenta de la flexibilidad lingüística que ha alcanzado este sistema de organización del conocimiento para dar respuesta a las nuevas necesidades de los usuarios de los entornos digitales.

2. Las relaciones terminológicas

Las relaciones terminológicas cumplen un papel fundamental en el uso de los lenguajes documentales y las ontologías, en la medida en que ofrecen opciones para refinar los resultados de la búsqueda y recuperación de información, de acuerdo con los intereses particulares de los usuarios.

Como se dijo cuando se habló de las relaciones que se presentan en un lenguaje documental como los tesauros, los vínculos que pueden crearse entre

⁸⁴ *Ibidem*, p. 85

⁸⁵ *Ídem*.

los términos que los componen son en su mayoría jerárquicos⁸⁶, debido a que responden a taxonomías estrictas que van de los aspectos más generales a los más específicos, para desarrollar una temática.

Sin embargo, si se revisan las relaciones que se establecen entre las clases de una ontología, se puede encontrar que la mayoría de ellas son de carácter asociativo para contextualizar los conceptos en el universo de conocimiento al que pertenecen, y hacer explícitas todas las formas a partir de las cuales se vinculan para darle sentido a dicho universo.

Adicionalmente, y en respuesta a la necesidad de inmediatez de la información que tienen los usuarios en la actualidad, el hecho de que las relaciones de tipo asociativo sean las que primen en las ontologías refleja la flexibilidad que este sistema de organización del conocimiento tiene para que, en el momento en que un usuario realice una búsqueda, se realicen diferentes asociaciones que garanticen su satisfacción en materia de resultados recuperados⁸⁷.

3. Necesidad de actualización

Al estar vinculados con áreas particulares de conocimiento, los lenguajes documentales y las ontologías requieren de actualizaciones periódicas que les permitan estar al día en relación con la temática que desarrollan, para así poder dar respuesta a los requerimientos de los usuarios⁸⁸.

En este sentido, y teniendo como contexto el entorno digital, la actualización periódica se torna más frecuente para adaptarse al creciente volumen de información disponible en la Web, de modo que las ontologías y los lenguajes documentales que se adapten al entorno digital deben ser mejoradas y puestas

⁸⁶ *Ibidem*, p. 86

⁸⁷ *Ídem*.

⁸⁸ *Ídem*.

al día continuamente para cumplir con su papel en los procesos de búsqueda y recuperación de información.

Por otra parte, si se tiene en cuenta la propuesta de Sánchez-Jiménez y Gil-Urdiciain⁸⁹, se encuentran varias razones que dan cuenta de las grandes diferencias que existen entre los lenguajes documentales tradicionales y las ontologías a partir de dos aspectos clave:

1. Las clases o conceptos representados

De acuerdo con estos autores, una de las diferencias más marcadas radica en el significado y la función que tienen las clases o conceptos que conforman los lenguajes documentales clásicos y las ontologías, esto debido a que para los primeros una clase es un grupo de términos con características comunes y para las segundas, es una agrupación de atributos que posibilitan la descripción de instancias o ejemplos que representan a cada clase⁹⁰.

Adicionalmente si se piensa en el entorno digital, en el momento en que se establecen clases que agrupan propiedades para describir las instancias asociadas a una clase que representa un concepto particular en sí misma⁹¹, se genera una estructura semántica comprensible por las máquinas que explicita el concepto como tal, facilitando la recuperación de información por parte de los usuarios.

Mientras que los lenguajes tradicionales como las listas de encabezamientos de materia o los listados de palabras clave, carecen de ese conjunto de

⁸⁹ SÁNCHEZ-JIMÉNEZ, Rodrigo y GIL-URDICIAIN, Blanca. Lenguajes documentales y ontologías. En: El profesional de la información, v. 16, no. 6, (nov.-dic., 2007) pp. 551 – 561. ISSN 1386-6710 (Consultado el 27 de octubre de 2012) Disponible en: <http://www.elprofesionalde lainformacion.com/contenidos/2007/noviembre/02.pdf>

⁹⁰ *Ibidem*, p. 554

⁹¹ *Ídem*.

propiedades que le permitan validar las razones por las que un término o entidad cualquiera pueda pertenecer a las clases que los conforman⁹², aspecto que hace que parte del conocimiento organizado en ellos quede implícito y no pueda ser comprendido por las máquinas o los sistemas de información en los que se encuentran alojados y por ende, dificultar la búsqueda y recuperación de los recursos que contengan.

Otra diferencia asociada con el manejo de las clases por los lenguajes documentales tradicionales y las ontologías se relaciona con lo que cada una de ellas representa en el contexto en el que se encuentra, es decir, mientras que los términos de una lista de encabezamiento de materias representan significados, los términos o conceptos que se encuentran en las ontologías representan entidades que pueden asociarse con las clases y subclases que estas contienen, o las instancias que relacionen a esas clases con la realidad vinculada con el universo que se maneja en una ontología⁹³.

2. Las relaciones presentes en los lenguajes documentales y las ontologías

En relación con este punto, Sánchez-Jiménez y Gil-Urdiciain afirman que las relaciones terminológicas existentes en los lenguajes documentales clásicos, como los sistemas de clasificación y los tesauros, se diferencian de las relaciones que se evidencian en las ontologías en cuanto a:

- La semántica

Mientras que en los sistemas de clasificación y en los tesauros hay pocas materias que pertenecen a más de un área de conocimiento y pocos términos que contenidos por más de una clase, aspecto que se conoce

⁹² Ídem.

⁹³ Ibídem, p. 556

como polijerarquía, en las ontologías se presentan varios términos que además de pertenecer a varias clases, tienen diferentes significados, situación conocida como polisemia.

- La jerarquía

En el momento en que se analizan las jerarquías presentes en los lenguajes documentales clásicos y las ontologías, se puede evidenciar una diferencia clara, ya que mientras las jerarquías en los sistemas de clasificación, las listas de encabezamiento de materias y los tesauros son muy estrictas y poseen pocos términos polijerárquicos por el alto contenido semántico que tienen, las jerarquías en una ontología contienen más casos de polijerarquía, debido a que la propiedad de univocidad no es requisito para la estructuración de este sistema de organización del conocimiento.

- La estructura

Si se revisa la estructura de los lenguajes tradicionales como los tesauros o como los sistemas de clasificación se puede evidenciar que tienen una estructura arborescente que responde a las jerarquías estrictas que se encuentran implícitas en ellos, aspecto que contrasta con las características estructurales de las ontologías, que son flexibles debido a que no poseen una jerarquía rigurosa y estricta que organice los términos que contiene.

Teniendo en cuenta esta última parte desarrollada, vinculada con la relación existente entre las ontologías y los lenguajes documentales tradicionales, se puede concluir que si bien las ontologías cuentan con una gran flexibilidad que permite al usuario recuperar aquellos recursos que dan respuesta a sus inquietudes de una manera más directa y en lenguaje natural, y que la aleja de los

lenguajes clásicos; tiene varios aspectos en común con los sistemas de organización del conocimiento clásicos.

Se puede afirmar esto si se tiene en cuenta que estos sistemas de organización de conocimiento tienen la necesidad de establecer estructuras semánticas que los soporten y relaciones terminológicas que compacten los dominios de conocimiento en los que se centran, y lo hagan más comprensible por los usuarios y las herramientas tecnológicas intermedias que facilitan su interacción.

En el siguiente capítulo se hará una mayor profundización en el establecimiento de los aspectos comunes y diferentes existentes entre los tesauros y las ontologías, sistemas de organización de conocimiento en los que se centra este trabajo.

5. COMPARACIÓN DE TESAuros Y ONTOLOGÍAS

Dando continuidad a la diferenciación propuesta hacia el final del capítulo anterior y teniendo en cuenta el objetivo principal que persigue este trabajo, asociado con el desarrollo de una metodología para generar ontologías a partir de tesauros existentes, a continuación se hará una comparación entre estos sistemas de organización del conocimiento a partir de dos aspectos: las características terminológicas que presentan y las relaciones existentes entre los conceptos que representan.

5.1. Aspectos terminológicos de los tesauros y las ontologías

Como se ha visto hasta este momento, los tesauros y las ontologías como sistemas de organización del conocimiento son herramientas que se involucran directamente con los procesos de descripción e indización de fuentes de información en los entornos analógicos y digitales con el objeto de favorecer su recuperación por parte de los usuarios que puedan requerirlas.

Para cumplir con su función en los procesos de búsqueda y recuperación de información, los tesauros y las ontologías cuentan con representaciones terminológicas de conceptos propios de las áreas del conocimiento en las que se especializan, que pueden ser analizadas desde:

- Su base léxica:

Los tesauros, en cumplimiento de sus funciones como lenguajes controlados, se preocupan por normalizar el lenguaje natural y las expresiones de uso cotidiano empleadas por las personas del común para referirse a temas específicos.

Dentro del proceso de normalización de las expresiones del lenguaje natural, los desarrolladores de tesauros realizan la sustantivación de dichas expresiones⁹⁴, consiguiendo que la categoría léxica que prima en los tesauros, sean los sustantivos, dejando de lado aquellos elementos que, a la luz de las teorías consultadas acerca de los tesauros, puedan distorsionar los resultados obtenidos en la recuperación de información, tales como los adjetivos, verbos, prefijos, sufijos, entre otros.

Sin embargo, si se tienen en cuenta las características terminológicas de las ontologías, es posible identificar diferencias en torno al componente léxico de las mismas, ya que las ontologías no realizan un control tan estricto como los tesauros de las expresiones del lenguaje natural, aspecto que logra que las ontologías sean más flexibles y más usables por los usuarios, en la medida en que funciona en “sus propios términos”.

Otra de las diferencias que se encuentran al revisar la parte léxica de las ontologías es que a diferencia de los tesauros, estos sistemas de organización del conocimiento utilizan diferentes formas léxicas en los términos que manejan para representar los conceptos que incluyen, enriqueciéndolos, como es el caso de los adjetivos cuando califican un término, o de los verbos, cuando representan acciones asociadas a los conceptos trabajados⁹⁵.

- La independencia o dependencia que tienen del contexto en el que se desarrollan

Otro tipo de diferencias relacionadas con la terminología que se puede establecer entre los tesauros y las ontologías, se asocia con el nivel de independencia que los términos de estos sistemas de organización del

⁹⁴ GARCÍA JIMÉNEZ, Antonio. Óp. Cit., p. 85

⁹⁵ Ídem.

conocimiento tienen respecto de las fuentes de información que los alimentaron durante su desarrollo.

Se considera esto una diferencia ya que mientras el sentido de los términos de un tesoro siempre es el mismo dentro del área de conocimiento en la que se especializa y puede ser utilizado por los analistas para describir el contenido que analizan en todo momento, los términos de una ontología deben asociarse a sistemas de información que le den contexto y la hagan útil para la recuperación de información⁹⁶.

Para afirmar lo anterior se parte de la idea de una ontología que no se encuentra asociada a ningún sistema de información, al carecer de jerarquías que diferencien los términos aprobados y de las cancelaciones para representar conceptos que permitan recuperar contenido, o de notas de alcance que indiquen el sentido en que deben ser entendidos tales términos en el contexto de la ontología, pueden hacer que se recuperen recursos que no tengan nada que ver con la necesidad de información a la que se quiere dar respuesta y que el usuario no obtenga lo que requiere.

5.2. Relaciones terminológicas presentes en los tesauros y las ontologías

Dando continuidad a los planteamientos que se han hecho hasta el momento, es necesario hablar de las relaciones terminológicas que se presentan en los tesauros y las ontologías, por la importancia que estas adquieren cuando los usuarios están haciendo uso de esos sistemas de organización del conocimiento para recuperar documentos que den respuesta a sus necesidades de información.

Dicha importancia puede sustentarse si se observa que con la ayuda de las asociaciones terminológicas que se establezcan en los tesauros y las ontologías,

⁹⁶ CODINA, Lluís y PEDRAZA-JIMÉNEZ, Rafael. Óp. Cit., p. 559 - 560

los usuarios pueden profundizar más en la temática que están consultando, identificar otros tópicos con los que tal tema tenga relación o determinar qué aspectos quiere excluir de su búsqueda para resolver las inquietudes que tenga.

Teniendo claridad en la importancia que tienen las relaciones en los procesos de búsqueda y recuperación de información, se puede dar inicio a su comparación, cuando se presentan en los tesauros y las ontologías.

Uno de los aspectos a partir de los cuales se puede establecer la comparación mencionada es el tipo de relaciones que prima en los tesauros y las ontologías, y la forma como estas afectan el uso de los sistemas de organización del conocimiento en la búsqueda y recuperación de la información.

Para empezar, se puede afirmar que las relaciones que se destacan en los tesauros, como lenguajes documentales clásicos, son las relaciones de carácter jerárquico, que se encargan de reflejar las taxonomías que los estructuran y de hacer visibles las relaciones de generalidad, especificidad y asociación existentes entre los términos que los componen⁹⁷, y de orientar la búsqueda y recuperación de información hacia la profundización o generalización de tal proceso, a partir de las identificación de los TG, TE y TR que concuerden con la necesidad a la que se quiere responder.

Caso contrario se presenta en las ontologías, pues son sistemas de organización del conocimiento que tienden a obviar las relaciones jerárquicas, que luego de ser empleadas en la estructuración de las ontologías, son omitidas cuando han sido terminadas⁹⁸, debido al carácter asociativo que requieren para cumplir con su función en el entorno digital.

⁹⁷ GARCÍA JIMÉNEZ, Antonio. Óp. Cit., p. 86

⁹⁸ Ídem.

En concordancia con lo anterior, las relaciones que se presentan en las ontologías son las de tipo asociativo, que permiten un nivel más alto de articulación terminológica, que puede impactar el proceso de recuperación de información⁹⁹, al posibilitar el establecimiento de nuevas asociaciones que amplíen el número de resultados obtenidos, con contenidos de calidad que respondan a la necesidad de información que dio origen a la búsqueda.

Teniendo en cuenta las diferencias establecidas en cuanto a las relaciones que se presentan en los tesauros y en las ontologías, se puede concluir que debido a la riqueza semántica que adquieren las ontologías al contar con un mayor número de relaciones de carácter asociativo, la recuperación de información a partir del uso de ellas se hace más efectiva.

Además, es válido afirmar que la efectividad en la recuperación de la información que alcanza una ontología se debe al establecimiento de relaciones asociativas que reflejan el alto nivel de representación conceptual que alcanzan¹⁰⁰, ya que hacen que estos sistemas de organización del conocimiento sean más cercanos a las asociaciones naturales que puede hacer el hombre, al tiempo que se hacen comprensibles por los computadores y adaptables a los sistemas de información que las pueden alojar, por el grado de formalización que alcanzan.

5.3. Semejanzas y diferencias entre tesauros y ontologías

Concluyendo este capítulo dedicado a la comparación de las características y aspectos relacionados con las ontologías y los tesauros, a continuación se presenta un listado de las diferencias y semejanzas presentes entre estos dos sistemas de organización del conocimiento, que de acuerdo con la propuesta de

⁹⁹ CODINA, Lluís y PEDRAZA-JIMÉNEZ, Rafael. Óp. Cit., p. 560

¹⁰⁰ ARANO, Silvia. Los tesauros y las ontologías en la Biblioteconomía y la Documentación. En: Anuario Hipertext.net, no. 3 (2005) (Consultado el 15 de noviembre de 2012) Disponible en: <http://www.upf.edu/hipertextnet/numero-3/tesauros.html>

Qin y Paling¹⁰¹, son las que dejan claros los aspectos que deben aprovecharse para generar una convergencia entre ellos.

Entre los aspectos comunes y diferentes entre las ontologías y los tesauros que identifican estos autores, se encuentran los siguientes:

- La representación conceptual que se alcanza en una ontología es superior a la que alcanzan los tesauros puesto que incluye atributos o *slots* que crean agrupaciones conceptuales que difieren de la estructura jerárquica que manejan los lenguajes documentales tradicionales, aspecto que consigue que en la representación se identifiquen e incluyan diferentes tipos de relaciones que enriquecen la definición de objetos digitales en sistemas de información, con su descripción detallada¹⁰².
- El establecimiento de las relaciones en una ontología se relaciona con la definición de los vínculos que existen entre los términos que componen cada clase y subclase, así como entre las clases y subclases que se han identificado, de manera que se crean enlaces que permiten la asociación de todos los conceptos que estos sistemas de organización del conocimiento contienen¹⁰³.

Con el establecimiento de relaciones adicionales a las de tipo jerárquico que se presentan en los tesauros, se alcanza un nivel semántico mucho más alto, puesto que generan asociaciones que hacen que la representación conceptual sea mucho más completa y que la definición construida del objeto digital sea más específica.

¹⁰¹ QIN, Jian y PALING, Stephen. Converting a controlled vocabulary into an ontology: The case of GEM. En: Information Research, v. 6, no. 2 (ene., 2001) (Consultado el 15 de noviembre de 2012) Disponible en: <http://informationr.net/ir/6-2/paper94.html>

¹⁰² Idem.

¹⁰³ Ídem.

- El nivel de formalidad que se alcanza en una ontología permite que la representación conceptual que se desarrolla en su interior sea usable por las personas y comprensible por las máquinas en los procesos de búsqueda y recuperación de información, al tiempo que posibilita que el conocimiento que almacena sea compartido y reutilizado para la construcción de otras ontologías, que enriquezcan su estructura y complementen la representación de un área particular del conocimiento, que se organice en sistemas de información.

Por su parte, los tesauros son usados por personas y en algunos casos en los que se ha logrado su convergencia hacia lo digital, por algunos sistemas de información a los que han podido adaptarse, teniendo en común con las ontologías la característica de poder ser reutilizado para generar nuevos tesauros desde la actualización de sus estructuras.

6. INDICADORES PARA LA EVALUACIÓN DE TESAUROS

Teniendo en cuenta la anterior comparación realizada entre tesauros y ontologías, a continuación se plantearán los indicadores que se generaron con el objeto de evaluar si un tesauro tiene potencial para converger hacia una ontología, partiendo de las diferencias que se identificaron en relación con la representación conceptual y con las relaciones terminológicas.

Durante la explicación de estos indicadores, se irá realizando su aplicación al tesauro de la National Agricultural Library (NAL), de acuerdo con lo que se indicó en el capítulo dedicado a la definición del problema al que busca dar solución este trabajo de grado, para determinar si es adecuado o no para la implementación de la metodología que se presentará más adelante.

I. Capacidad de representación conceptual de la terminología del tesauro

Con este indicador se quiere medir la capacidad que tienen los términos incluidos en el tesauro analizado, de hacer una representación conceptual completa cuando estos son transferidos hacia un esquema ontológico.

En el caso del tesauro de la NAL, en la categoría de Tierra, Ambiente y Recursos naturales, se pudo identificar qué términos específicos como Ecología y Recursos Naturales pueden realizar representaciones conceptuales más completas que las presentes en el tesauro, al agrupar de manera diferente la terminología, consiguiendo un mayor desarrollo de la temática y un mejor cubrimiento de la misma, como se observa en la siguiente figura:

ecología

Términos Genéricos

Tierra, Ambiente y Recursos Naturales

Términos Específicos

agroecología
ecología animal
ecología comunitaria
ecología del fuego
ecología del paisaje
ecología del suelo
ecología forestal
ecología microbiana
ecología poblacional
ecología química
ecología vegetal
ecosistemas
ecotonas
paleoecología
procesos y fenómenos ecológicos
relaciones tróficas
zonas ecológicas

Términos Relacionados

ambiente
ecofisiología
ecólogos

English

ecology

Ecología

agroecología
alimentos ecológicos
ambiente acuático
biocontención
biogeografía
biología del suelo
biomas
bioprotección
bosques
cadena alimentaria
calentamiento global
cambio climático
características de la población
carga de combustible (forestal)
ciencia ambiental
Ciencia de Animales y Productos de Animales
ciencia del suelo
ciencia del suelo
Ciencia Forestal y Productos Forestales
ciencias atmosféricas

Figura 6. Comparación entre la terminología asociada a la Categoría Ecología, en el tesoro de la NAL y una pequeña muestra de la estructura ontológica que puede crearse si se trabajan las relaciones presentes en el tesoro.

II. Capacidad de la estructura jerárquica de agrupar conceptos en categorías o clases

Con la aplicación de este indicador se comprueba la existencia de una organización terminológica jerárquica en el tesoro que se estudie, que sea fuerte y sirva como estructura inicial de una ontología como la que se quiere generar con la metodología que se planteará más adelante.

Al revisar el tesoro desarrollado por la NAL, se identifica el desarrollo de una estructura jerárquica que agrupa términos en categorías y subcategorías, y que realiza un desarrollo temático que inicia con términos generales y se orienta hacia la identificación de términos específicos que lo representan a profundidad.

Se puede observar esto si se analiza su estructura desde el punto de vista jerárquico, como se ve a continuación:

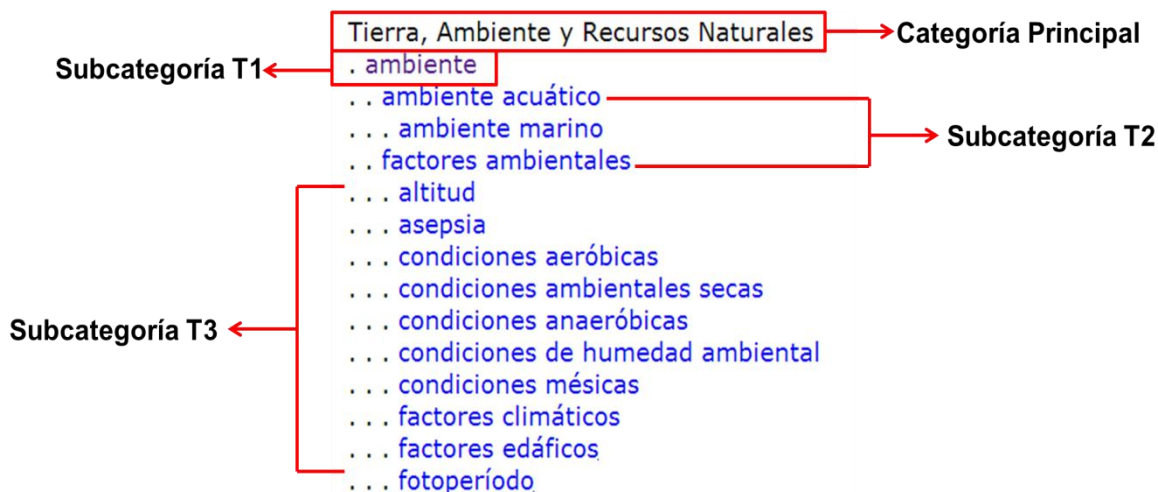


Figura 7. Relaciones jerárquicas definidas en el tesoro de la National Agricultural Library

III. Capacidad de vinculación de los términos del tesoro mediante relaciones de asociación

Este indicador se encuentra relacionado con la identificación de la posibilidad de crear relaciones de asociación entre los términos del tesoro, de manera que se adicione conocimiento a tales vínculos con la inclusión de categorías léxicas como verbos, adjetivos y sustantivos en la definición de las relaciones terminológicas.

Cuando se quiso comprobar si los términos del tesoro de la NAL tenían la capacidad de relacionarse a partir de asociaciones, que agregaran conocimiento a la vinculación jerárquica que ya existe entre ellos, se encontró que:

Cuando se revisa la subcategoría Recursos Naturales, se encuentran las siguientes relaciones:

recursos naturales

Términos Genéricos

Tierra, Ambiente y Recursos Naturales

Términos Específicos

recursos agrícolas
recursos biológicos
recursos de la tierra
recursos energéticos
recursos forestales
recursos hídricos
recursos marinos
recursos minerales
recursos no renovables
recursos pesqueros
recursos renovables

Términos Relacionados

capital natural
infraestructura verde

Figura 8. Relaciones presentes en la subcategoría Recursos Naturales, de la categoría Tierra, Ambiente y Recursos naturales

Si se tienen en cuenta otros términos que se relacionarían con la subcategoría Recursos naturales por la temática que aborda, y que no se encuentran contemplados en los términos asociados directamente con dicha categoría, se podrían identificar los siguientes vínculos:

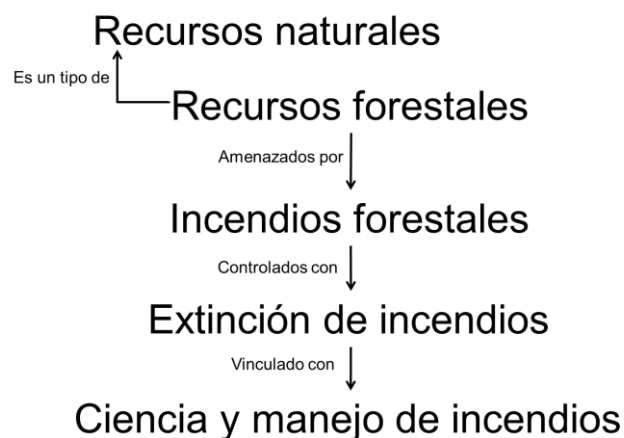


Figura 9. Relaciones de la subcategoría Recursos naturales con otros términos que no se vinculan directamente en ella, y que se pueden asociar haciendo uso de verbos que enriquecen su vínculo.

Luego de aplicar estos indicadores para evaluar si es posible que el tesoro de la NAL converja hacia una ontología, se pudo concluir que sí se puede trabajar con él para el diseño de la metodología que se presenta en el siguiente capítulo, debido a que:

- Sus términos pueden reorganizarse para generar representaciones conceptuales más amplias, que consiguen un mejor tratamiento de los aspectos vinculados con el área de conocimiento o universo en el que se especializa.
- Posee una estructura jerárquica que agrupa los términos en categorías principales y los organiza de acuerdo con el nivel de especificidad que alcanzan en relación con la temática con la que se relacionan, facilitando la identificación de aquellos términos genéricos y específicos que resultan importantes a la hora de mejorar la representación conceptual del tema en el que se especializa.
- Los términos que contiene son susceptibles de asociarse entre sí por la estructura jerárquica en la que se encuentran organizados, de manera que puede agregárseles valor semántico que hace que la representación conceptual alcance un nivel mayor de precisión cuando se transfiera hacia una ontología.

7. METODOLOGÍA PARA LA CONVERSIÓN DE TESAuros EN ONTOLOGÍAS

Posterior a todo el recorrido teórico que se ha hecho hasta este momento, de la evaluación del tesoro de la NAL a la luz de los indicadores que se crearon para definir si, desde la perspectiva que se ha manejado a lo largo de esta investigación, tiene potencial para converger hacia una ontología y de la identificación de este tesoro como herramienta para el desarrollo de esta metodología, a continuación se presenta el producto de este trabajo de grado.

7.1. Diseño de la metodología

Como resultado de la investigación que se realizó sobre los tesauros y las ontologías como sistemas de organización del conocimiento y de la comparación que se hizo de ellos para la identificación de las características comunes y diferentes que presentan, se pudo establecer que los aspectos en los que más difieren son las relaciones terminológicas que presentan y el nivel de representación conceptual que alcanzan.

En consecuencia, se planteó una metodología que retoma algunos elementos de aquellas que fueron expuestas en capítulos anteriores y se centra en el aprovechamiento de las diferencias que pudieron ser identificadas.

La metodología propuesta plantea las siguientes etapas:

1. Identificación de las categorías del tesoro en las que se va a trabajar para generar la ontología, y de los términos que las componen.

2. Extracción de los términos importantes que contribuyen con la representación conceptual de tales categorías para establecer las clases que tendrá la ontología.
3. Proposición de nuevos términos en caso de ser necesario, por la falta de términos que posibiliten la agrupación de los descriptores del tesoro en clases de la ontología que se creará
4. Eliminación de aquellos términos que no concuerdan con la temática que se desarrollará en la ontología.
5. Identificación de los niveles jerárquicos que ayudarán a estructurar la ontología con una agrupación primaria de los términos que la compondrán.
6. Definición y establecimiento de las relaciones de tipo asociativo que pueden crearse para generar la agrupación terminológica definitiva que tendrá la ontología.
7. Creación de la estructura ontológica teniendo en cuenta los componentes estructurales que se identificaron en el capítulo relacionado con las ontologías, entre los que se encuentran las clases o conceptos, atributos o propiedades, instancias, facetas y relaciones.

7.2. Implementación de la metodología

Al probar la metodología planteada en el numeral anterior en el tesoro de la NAL se siguió el siguiente proceso:

1. Se identificaron tres subcategorías de la categoría Tierra, ambiente y recursos naturales del Tesoro mencionado. Estas son Ambiente, Ecología y Recursos naturales.

Posterior a la identificación de estas subcategorías, se realizó la identificación de los términos que están asociados a ellas, generando una perspectiva general de las temáticas con las que se encuentran vinculadas y la definición de las posibles relaciones que se podían desarrollar en la ontología que se estaba creando.

A continuación se presenta el resultado de esta etapa de reconocimiento de las categorías seleccionadas:

Subcategoría	Ambiente
Término general	Tierra, ambiente y recursos naturales
Términos específicos	Ambiente acuático Factores ambientales
Términos relacionados	Agricultura sustentable Cambio climático Ecología Ley ambiental Política ambiental Sustancias tóxicas

Tabla 1. Términos asociados con la subcategoría Ambiente, dentro del tesoro de la National Agricultural Library

Subcategoría	Ecología
Término general	Tierra, ambiente y recursos naturales
Términos específicos	Agroecología Ecología animal Ecología comunitaria Ecología del fuego Ecología del paisaje Ecología del suelo Ecología forestal Ecología microbiana

	Ecología poblacional Ecología química Ecología vegetal Ecosistemas Ecotonas Paleoecología Procesos y fenómenos ecológicos Relaciones tróficas Zonas ecológicas
Términos relacionados	Ambiente Ecofisiología Ecólogos

Tabla 2. Términos asociados con la subcategoría Ecología, dentro del tesoro de la National Agricultural Library

Subcategoría	Recursos naturales
Término general	Tierra, ambiente y recursos naturales
Términos específicos	Recursos agrícolas Recursos biológicos Recursos de la tierra Recursos energéticos Recursos forestales Recursos hídricos Recursos marinos Recursos minerales Recursos no renovables Recursos pesqueros Recursos renovables
Términos relacionados	Capital natural Infraestructura verde

Tabla 3. Términos asociados con la subcategoría Recursos naturales, dentro del tesoro de la National Agricultural Library

2. Luego de la identificación de los términos que componen las tres subcategorías analizadas, se procedió a extraer aquellos descriptores que por su importancia dentro de las temáticas trabajadas, podrían ser considerados clases o subclases dentro de la ontología que se estaba construyendo con la metodología planteada.

Los términos extraídos de los listados presentados en el paso anterior y que se identificaron como posibles clases conceptuales en la ontología que se quiso desarrollar fueron:

- Ecología
- Recursos naturales
- Ecosistemas

3. En vista de la falta de términos que agruparan algunos de los descriptores que componen las subcategorías seleccionadas del tesoro, se incluyeron los siguientes términos que contribuyeron con el establecimiento de clases en la ontología que se creó:

- Problemática ambiental
- Economía ambiental
- Alternativas ecológicas
- Legislación ambiental

4. Debido a que la ontología que se desarrolló se centraba en la temática ambiental y ecológica, se identificó y eliminó el término *Paleoecología* que por su significado, no tenían mayor vinculación con el universo en el que se centró este sistema de organización del conocimiento.

5. Al contar ya con un conjunto de términos definido para la construcción de la ontología, se identificaron los niveles jerárquicos que se pudieron establecer entre ellos, con el objetivo de crear una estructura base desde la que se pudieran identificar las relaciones de tipo asociativo que enriquecieran a la ontología, desde la mejora de la representación conceptual.

Los niveles jerárquicos que se identificaron fueron:

Clases	Subclases
Ecología	<ul style="list-style-type: none"> - Ecosistemas - Recursos naturales <ul style="list-style-type: none"> o Recursos renovables o Recursos no renovables
Problemática ambiental	<ul style="list-style-type: none"> - Alternativas ecológicas - Legislación ambiental
Economía ambiental	<ul style="list-style-type: none"> - Capital verde - Infraestructura verde

Tabla 4. Identificación de las clases y subclases que agruparon los términos incluidos en la ontología desarrollada en este trabajo de grado.

6. Teniendo claridad con respecto a las clases y subclases que se identificaron dentro de la estructura jerárquica que dio soporte a la ontología que se generó, se empezaron a definir las relaciones asociativas entre las clases y subclases seleccionadas, con el objetivo de entender la manera como se vinculan.

El resultado de esta etapa fue el siguiente:

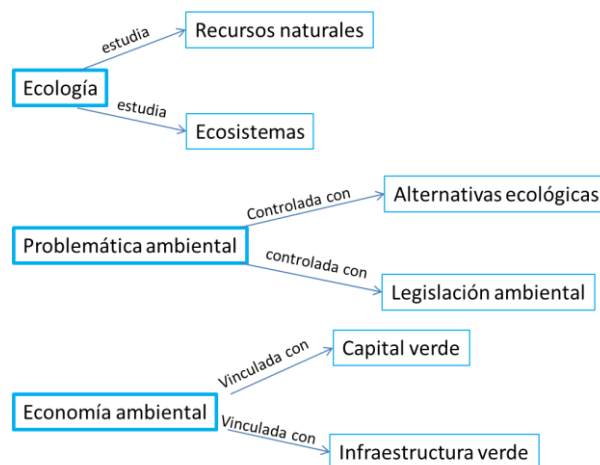


Figura 10. Relaciones asociativas entre las clases y subclases identificadas en la etapa anterior

7. Finalmente, teniendo claridad con respecto a la forma como se vinculaban las clases y subclases definidas, se inició el proceso de creación de la estructura ontológica, teniendo en cuenta los componentes identificados en capítulos anteriores.

Las estructuras ontológicas construidas para cada una de las clases que se identificaron fueron:

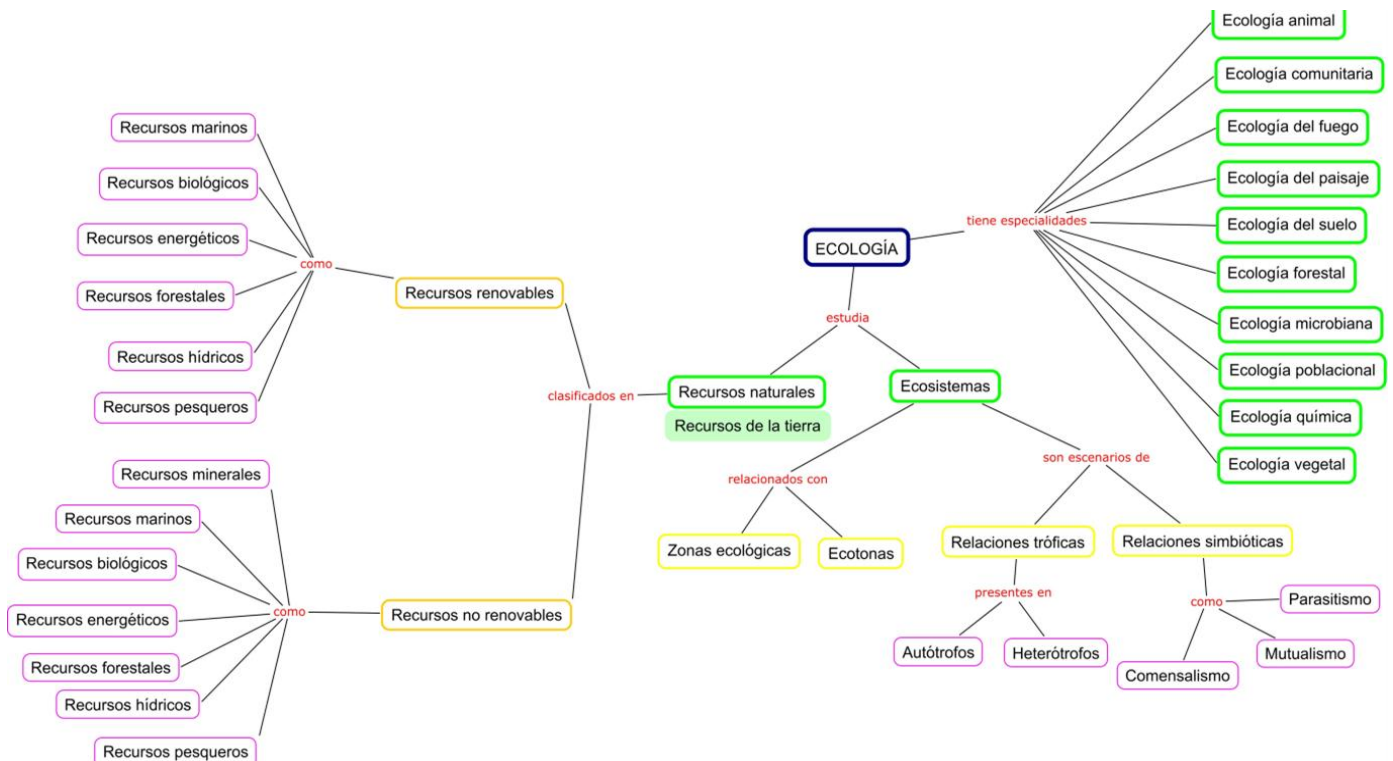


Figura 11. Estructura ontológica asociada con la clase Ecología

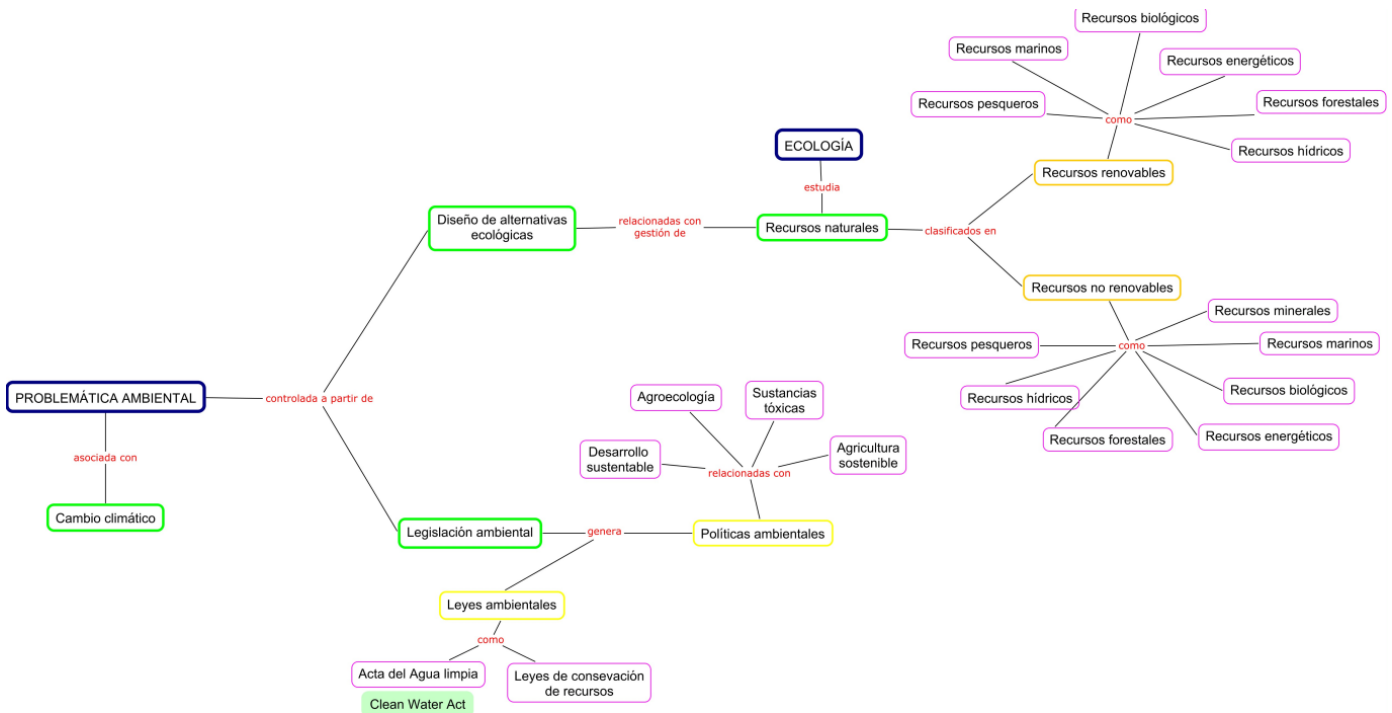


Figura 12. Estructura ontológica asociada con la clase Problemática ambiental

Para el análisis de estos esquemas ontológicos se debe tener en cuenta que:

- Las clases o conceptos principales son los que están en los cuadros con borde de color azul, en letra mayúscula
- Los atributos o *slots* que facilitan creación de agrupaciones terminológicas dentro de la ontología son las palabras escritas en letras rojas
- Las subclases de los conceptos principales son aquellas que están en los cuadros con borde de color verde
- Los términos equivalentes que pudieron establecerse son aquellos que están en cuadros con relleno de color verde

- Las instancias primarias son aquellas que se encuentran en cuadros de borde de color amarillo y las secundarias, en cuadros de borde de color fucsia.
- Las relaciones están representadas por las líneas de color negro

Como pudo verse en las estructuras ontológicas presentadas anteriormente, la metodología planteada para la conversión de tesauros en ontologías fue implementada con algunas categorías del tesoro de la NAL y puesta a prueba de manera positiva, ya que haciendo uso de las diferencias establecidas entre estos sistemas de organización del conocimiento se consiguió:

- Identificar las clases o conceptos que facilitaron la agrupación de los diferentes términos que componen las categorías analizadas.
- Determinar la estructura jerárquica a partir de la cual se pudo generar los esquemas ontológicos que se presentaron.
- Establecer los atributos o propiedades con las que se pudieron definir las relaciones asociativas entre las clases y la terminología asociada a cada una de ellas.
- Hacer uso de categorías léxicas como los verbos, para agregar conocimiento a las asociaciones hechas entre los términos asociados a la estructura ontológica generada.

8. CONCLUSIONES

Al concluir el desarrollo de este trabajo de grado, se pueden deducir diferentes aspectos respecto a la investigación realizada, la implementación de la metodología desarrollada y las estructuras ontológicas que se construyeron. Entre estas se encuentran las siguientes:

- Durante el desarrollo de la investigación que permitió el desarrollo del marco teórico de este trabajo de grado, se evidenció que existe una gran cantidad de literatura relacionada con tesauros y con ontologías, en la que se exponen sus componentes, características, funciones y uso, así como su relación con otros sistemas de organización de conocimiento, en el caso de los vocabularios controlados, y con software e instrumentos tecnológicos, en cuanto a las ontologías.

Sin embargo, se pudo concluir que aún hace falta desarrollo de investigaciones que se concentren en tratar la relación que existe entre las ontologías y los lenguajes documentales clásicos, como los tesauros, para identificar las relaciones que existen entre estos sistemas de organización del conocimiento en cuanto a su estructura y la forma como realizan la representación conceptual de las áreas en las que se especializan.

- A medida que se avanzó en el desarrollo de la investigación que dio sustento teórico a la metodología planteada se evidenció que cuando los autores hablan acerca de ontologías, en algunos casos se le da más énfasis a las herramientas e instrumentos tecnológicos que a la estructura semántica misma de este sistema de organización del conocimiento.
- Es necesario que desde el campo de la Ciencia de la Información se empiece a dar importancia a la estructuración conceptual de la ontología por encima de las

herramientas tecnológicas o software, ya que si bien las herramientas permiten el desarrollo de procesos de búsqueda y recuperación de información, una adecuada representación conceptual es la que los hace posibles.

- A la hora de empezar a trabajar en la conversión de un tesoro como el de la NAL, es importante tener en cuenta que para conseguir el resultado esperado, es decir, la generación de una ontología, se debe conocer el área en la que se especializan estos sistemas de organización de conocimiento para poder desarrollarlos correctamente, de acuerdo con las lógicas de las disciplinas en las que se especialicen.
- A medida que se va desarrollando la metodología y se va aplicando en un tesoro definido, resulta necesario contar con otros vocabularios que se especialicen en temáticas similares a la del escogido, con el objeto de tener un punto de comparación que posibilite la identificación de fallas y aciertos en la definición de las relaciones conceptuales, el nivel de representación conceptual y el manejo de las estructuras jerárquicas que les dan soporte.

BIBLIOGRAFÍA

ARANO, Silvia. Los tesauros y las ontologías en la Biblioteconomía y la Documentación. En: Anuario Hipertext.net, no. 3 (2005) (Consultado el 15 de noviembre de 2012) Disponible en: <http://www.upf.edu/hipertextnet/numero-3/tesauros.html>

CODINA, Lluís y PEDRAZA-JIMÉNEZ, Rafael. Tesauros y ontologías en sistemas de información documental. En: El profesional de la información, v. 20, no. 5 (sept.-oct, 2011) p. 555 – 563 (Consultado el 12 de agosto de 2012) Disponible en: http://www.lluiscodina.com/ontologiaTesauros_2011.pdf

DE SALES, Rodrigo y CAFÉ, Ligia. Diferencias entre tesauros e ontologías. En: Perspectivas em Ciencia da Informacao, v.14, no. 1 (ene.-abril, 2009) p. 99 – 116 (Consultado el 12 de agosto de 2012) Disponible en: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362009000100008

FERNÁNDEZ BREIS, Jesualdo Tomás. Un entorno de integración de ontologías para el desarrollo de sistemas de gestión de conocimiento. 386 p. 2003. Tesis doctoral (Ciencia de la computación e inteligencia artificial) Universidad de Murcia. Departamento de Ingeniería de la Información y las Comunicaciones. (Consultado el 20 de octubre de 2012) Disponible en: <http://hdl.handle.net/10803/10921>

GARCÍA JIMÉNEZ, Antonio. Instrumentos de representación del conocimiento: tesauros versus ontologías. En: Anales de documentación, v. 7 (2004) p. 79 - 95 (Consultado el 30 de octubre de 2012) Disponible en: <http://revistas.um.es/analesdoc/article/view/1691/1741>

GARCÍA TORRES, Alberto, et al. Reutilización de tesauros: El documentalista frente al reto de la web semántica. En: El profesional de la información, v.17, n. 1, (ene.-feb., 2008) p. 8 – 21 (Consultado el 12 de agosto de 2012) Disponible en: <http://www.elprofesionaldelainformacion.com/contenidos/2008/enero/02.pdf>

GIL URDICIÁIN, Blanca. Manual de lenguajes documentales. Gijón (Asturias): Ediciones Trea, 2004. 280 p. ISBN: 84-9704-138-0

GILCHRIST, Alan. Thesauri, taxonomies and ontologies - an etymological note. En: Journal of Documentation, v. 59, no. 1 (2003) p. 7 - 18

GIMENO PERELLÓ, Javier. Sistemas de indización aplicados en bibliotecas: Clasificaciones, tesauros y encabezamientos de materias. En: Tratado básico de Biblioteconomía. Madrid: Editorial Complutense, 2004. 484 p. ISBN: 84-7491-750

GRUBER, Thomas. A translation approach to portable ontology specification. En: Knowledge acquisition, v. 5, no. 2 (abril, 1993) p. 199 - 220. ISSN: 1042-8143 (Consultado el 20 de octubre de 2012) Disponible en: <http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>

GUARINO, Nicola. Formal Ontology, Conceptual Analysis and Knowledge Representation. En: International Journal of Human-Computer Studies, v. 43, no. 5/6 (nov., 1995) p. 625 – 640. ISSN: 1071-5819 (Consultado el 20 de octubre de 2012) Disponible en: <http://www.loa.istc.cnr.it/Papers/FormOntKR.pdf>

GUINCHAT, Claire y MENO, Michel citado por MARTÍN GAVILÁN, César. Lenguajes documentales. Principales tipos de clasificación. Encabezamientos de materia. Descriptores y tesauros. (Consultado el 8 de septiembre de 2012) Disponible en: <http://eprints.rclis.org/bitstream/10760/14817/1/lendoc.pdf>

HILERA R., José, et al. An evolutive process to convert glossaries into ontologies. En: Information Technology and Libraries, v. 29 no. 4 (dic, 2010) p. 195-204

IBEKWE-SANJUAN, Fidelia. Constructing and maintaining knowledge organization tools: a symbolic approach. En: Journal of Documentation, v. 62, no. 2, p.229 – 250

LUCAS DA SILVA, Daniela. ROCHA SOUZA, Renato y BARCELLOS ALMEIDA, Mauricio. Ontologias e vocabularios controlados: comparação de metodologías para construção. En: Ciência da Informação, v. 37, no. 3(sept.-dic., 2008) p. 60 - 75. ISSN: 0100-1965. Disponible en: <http://www.scielo.br/pdf/ci/v37n3/v37n3a05.pdf>

MOCHON BEZARES, Gonzalo y SORLI ROJO, Ángela. Tesauros de Ciencias Naturales en Internet. En: Revista Española de Documentación Científica, v. 31. no. 4 (oct.-dic., 2008) p. 647 - 658 (Consultado el 8 de septiembre de 2012) Disponible en: <http://redc.revistas.csic.es/index.php/redc/article/view/448/460>

MOURA, Maria Aparecida. Information, ontological tools and ad hoc social nets: the interoperability in the construction of thesaurus and ontologies. En: Informacao & Sociedade-Estudos, v. 19, no.1, (2009) p. 59-73.

NATIONAL AGRICULTURAL LIBRARY. Biblioteca Nacional de Agricultura (Consultado el 8 de septiembre de 2012) Disponible en: <http://www.nal.usda.gov/about/spanfly.pdf>

NECHES ROBERT, et al. Enabling technology for knowledge sharing. En: A.I. Magazine, v. 12, no. 3 (jul.-sept., 1991) p. 36 - 56. ISSN: 0738-4602 (Consultado el 20 de octubre de 2012) Disponible en: <http://www.aaai.org/ojs/index.php/aimagazine/article/view/902/820>

NOY, Natalia F. y McGUINNESS, Deborah L. Desarrollo de Ontologías-101: Guía para crear tu primera ontología. Standford University: Standford, 2005. 29 p. (Consultado el 29 de octubre de 2012) Disponible en: http://protege.stanford.edu/publications/ontology_development/ontology101-es.pdf

OSPINA RODRÍGUEZ, Martha Carolina. Guía metodológica para la elaboración de una ontología a partir de un tesoro: estudio analítico comparativo (Fase 1). En: Códices, v. 7, no. 1 (ene.-jun., 2011) p. 113 – 130

PINTO MOLINA, María. Análisis y síntesis del Proceso Documental (Consultado el 8 de septiembre de 2012) Disponible en: http://www.mariapinto.es/alfintra/contenidos/ANALISIS_SINTESIS.pdf

POPOVICI, Dorin-Mircea, et al. Applying of an ontology based modeling approach to cultural heritage systems. En: Advances in Electrical and Computer Engineering v. 11, no. 3 (2011) p. 105 – 110

QIN, Jian y PALING, Stephen. Converting a controlled vocabulary into an ontology: The case of GEM. En: Information Research, v. 6, no. 2 (ene., 2001) (Consultado el 15 de noviembre de 2012) Disponible en: <http://informationr.net/ir/6-2/paper94.html>

SANCHEZ-ALONSO, Salvador y GARCIA-BARRIOCANAL, Elena. Making use of upper ontologies to foster interoperability between SKOS concept schemes. En: Online InformationReview, v. 30, no. 3 (2006) p. 263-277

SÁNCHEZ-JIMÉNEZ, Rodrigo y GIL-URDICIÁIN, Blanca. Lenguajes documentales y ontologías. En: El profesional de la información, v. 16, no. 6, (nov.-dic., 2007) pp. 551 – 561. ISSN 1386-6710 (Consultado el 27 de octubre de 2012) Disponible en: <http://www.elprofesionalde lainformacion.com/contenidos/2007/noviembre/02.pdf>

SLYPE, Georges Van. Los lenguajes de indización: Concepción, construcción y utilización en los sistemas documentales. Madrid: Fundación Germán Sánchez Ruipérez, 1991. 198 p. ISBN: 84-86168-60-0

TEMATRES. Gestión de términos. Notas. (Consultado el 8 de septiembre de 2012) Disponible en: <http://r020.com.ar/tematres/manual/notas.html>

USCHOLD, Mike y KING, Martin. Towards a methodology for building ontologies. Artificial Intelligence Applications Institute. University of Edimburgh, 1995. 13 p. (Consultado el 3 de noviembre de 2012) Disponible en: http://www1.cs.unicam.it/insegnamenti/reti_2008/Readings/Uschold95.pdf