

DESARROLLO DE UNA ONTOLOGÍA EN EL CONTEXTO DE LA WEB SEMÁNTICA A PARTIR DE UN TESAURO DOCUMENTAL TRADICIONAL*

Jaime A. Guzmán Luna* *
Durley Torres Pardo* **
Alba Nubia López García* ***

RESUMEN

Un tesauro es una herramienta documental que se emplea para catalogar y recuperar información en entornos especializados. En la búsqueda de mejores alternativas para la representación de los recursos de información, las Ontologías en el contexto de la web semántica están siendo vistas como una nueva opción para representar el conocimiento existente en los tesauros, ya que se ha demostrado que permiten mejorar la precisión y reutilización de todo tipo de conocimiento. El objetivo de este trabajo es presentar una opción para la migración de un tesauro documental tradicional a una ontología en el entorno de la web semántica, con el fin de permitir su uso en la recuperación de documentos.

PALABRAS CLAVE: Tesauro/ Ontología/ Web Semántica/ Migración de Tesauros/ Recuperación de Información

GUZMÁN LUNA, Jaime, TORRES PARDO, Durley, LOPEZ GARCÍA, Alba Nubia. *Desarrollo de una ontología en el contexto de la web semántica a partir de un tesauro documental tradicional.* En: *Revista Interamericana de Bibliotecología*. Vol. 29, No. 2 (jul.-dic. 2006); p.79-94.

* Artículo derivado de la investigación: Desarrollo de un sistema de recuperación de información semántico especializado en artes plásticas y hábitat. Universidad Nacional de Colombia sede Medellín. Facultad de Arquitectura. Inicio: noviembre de 2005, finalización: noviembre de 2006. Artículo recibido el 22 de agosto, evaluado el 31 de octubre y aceptado con modificaciones el 8 de noviembre de 2006

** Magíster en Ingeniería de Sistemas. Docente Escuela de Sistemas, Universidad Nacional de Colombia Sede Medellín. jaguzman@unal.edu.co

*** Ingeniera de Sistemas. Auxiliar de Investigación Universidad Nacional de Colombia, sede Medellín. durleytp@gmail.com

**** Bibliotecóloga Unidad de Documentación CEHAP, Universidad Nacional de Colombia, Sede Medellín. unicehap@unalmed.edu.co

ABSTRACT

A thesaurus is a documentary tool that is used to catalog and retrieve information in specialized environments. In the search for better alternatives in the representation of information resources, ontologies—in the context of the semantic web—are being seen as a new means of representing the knowledge that currently exists in thesauri, now that it has been demonstrated that ontologies make it possible to improve the precision and the reutilization of all kinds of knowledge. The objective of this work is to present an option for the migration of the traditional documentary thesaurus to ontology in the environment of the semantic web, with the objective of permitting its use in document retrieval.

KEY WORDS: Thesaurus/ Ontology/ Semantic Web/ Migration of Thesaurus/ Information Retrieval

GUZMÁN LUNA, Jaime, TORRES PARDO, Durley, LOPEZ GARCÍA, Alba Nubia. *Development of an ontology in the context of the semantic web: the migration of the traditional documentary thesaurus* En: *Revista Interamericana de Bibliotecología*. Vol. 29, No. 2 (jul.-dic. 2006); p.79-94.

1. INTRODUCCION

“Un tesoro es un vocabulario controlado y dinámico, compuesto por términos que tienen entre ellos relaciones semánticas y genéricas y que se aplica a un dominio particular del conocimiento” [1], aunque también puede ser considerado como “un vocabulario de un lenguaje controlado de indexación, organizado formalmente, de manera que las relaciones entre los conceptos se establezcan a priori, para ser usadas en sistemas de recuperación de información” [2]. No obstante, diversos investigadores han defendido la necesidad tanto de cambios estructurales como de adaptaciones tecnológicas por parte de los tesauros [3] y avanzar hacia una técnica más moderna que se adapte mejor a la descripción de un recurso, pero sin perder los valiosos conocimientos registrados en los propios tesauros.

Una oportunidad para alcanzar este objetivo ha sido abierta hoy en día con el éxito que ha tenido la web semántica [4], una línea de investigación que busca extender la web actual mediante el marcado de sus recursos con metadatos semánticos [5] y el uso de ontologías [6] con el fin de permitir a las máquinas su entendimiento para lograr así la automatización de diversas tareas que requieran de tales recursos. Si partimos de la base que todos los tesauros tradicionalmente utilizados en centros de documentación y bibliotecas contienen siempre algún tipo de marcado propio en su definición, es lógico pensar que para lograr una primera adaptación de estos tesauros al ámbito de la web semántica es necesario codificar este marcado de forma que se adapte a los mecanismos y estándares de este ámbito. De esta forma podemos aprovechar gran parte del trabajo ya hecho registrado en los tesauros existentes.

El presente documento detalla la conversión de una adaptación del tesauro del Centro de Naciones Unidas sobre Asentamientos Humanos [7], en una ontología especializada en el hábitat. Con este trabajo se busca alcanzar una mayor precisión en la definición de las relaciones entre los términos, permitiendo la retoma de todo el conocimiento en el área del hábitat almacenada en el tesauro original, recuperando para ello su lista de términos y sus relaciones. En general, se trata de convertir la información almacenada en este tesauro en “conocimiento”, mediante estructuras de conocimiento formalizadas en una ontología, el cual podrá ser utilizado en la realización de tareas específicas tales como la indexación y la recuperación de la información semánticas [8, 9].

El trabajo aquí mostrado hace parte del proyecto de investigación “Desarrollo de un sistema de recuperación de información semántico especializado en artes plásticas y hábitat” realizado en la Facultad de Arquitectura de la Universidad Nacional de Colombia, sede Medellín, con el cual se busca mejorar la recuperación de información en una base documental al implementar tecnologías relacionadas con la web semántica. Al desarrollar la búsqueda documental, sobre una ontología se tratan de referenciar los datos a metadatos, con un esquema común sobre el dominio del hábitat, permitiendo además que todo estos datos sean entendidos y utilizados por el computador, con esto se mejora la búsqueda de información, así el agente buscador no sólo encontrará la información de forma precisa sino que le será posible la realización de inferencias de forma automática, buscando información relacionada con las especificaciones hechas por el usuario en la consulta. Sin embargo, los resultados presentados en el actual documento únicamente se ocuparán de la fase inicial que comprende el desarrollo de las ontologías de forma concreta, la que fue generada en el dominio del hábitat. Este documento esta organizado de la siguiente manera: La sección 2 recopila algunos aspectos importantes de las ontologías y la web semántica; en la sección 3 se describe lo que es un tesauro y presenta un breve resumen del *SKOS-Core*, una propuesta orientada a la clasificación temática en la web semántica; en la sección 4 se detalla el diseño de la ontología especializada en hábitat, realizada a partir del tesauro documental especializado en este tema haciendo uso del *SKOS-Core*; en la sección 5 se da una breve descripción de la implementación de la ontología y, finalmente, en la sección 6 se recopilan las conclusiones y trabajos futuros.

2. LAS ONTOLOGÍAS Y LA WEB SEMÁNTICA

Las ontologías han sido tradicionalmente usadas como modelo de representación del conocimiento en inteligencia artificial. Existen diferentes definiciones de lo que es una ontología siendo la más conocida la ofrecida por Gruber [10], donde una

ontología es “una especificación explícita y formal de una conceptualización”. De manera más simple, una ontología ha de concebirse como un entendimiento común y compartido de un dominio, que puede comunicarse entre humanos y sistemas computacionales. Ésta última característica y el hecho de que puedan compartirse y reutilizarse en diferentes aplicaciones, explica en parte el gran interés suscitado en los últimos años en la creación e integración de ontologías [11]. Las ontologías se están empleando en todo tipo de aplicaciones informáticas en las que sea necesario definir concretamente el conjunto de entidades relevantes en un campo de aplicación determinado, así como las interacciones entre las mismas.

En general, los objetivos de una ontología se pueden recopilar en los siguientes [12]:

- (a) *Compartir la comprensión común de la estructura de información entre personas o agentes del software*
- (b) *Permitir la reutilización del conocimiento perteneciente a un dominio,*
- (c) *Permitir representar de manera explícita los supuestos de un dominio*
- (d) *Separar el conocimiento de un dominio, del conocimiento que se pueda denominar operacional y*
- (e) *Hacer posible analizar el conocimiento de un campo.*

De cara a la web y retomando la definición técnica de ontología dada anteriormente, se encuentra que la web es mucho más extensa que una base de conocimiento, por lo que es necesario el uso de lenguajes estándares que permitan implementar ontologías en este ambiente. Esto obligó el desarrollo de diversos trabajos destacándose en los últimos años los trabajos realizados por la web Semántica, entre los que se encuentran el lenguaje *RDF (Resource Description Framework)* [12, 13] que implementa un modelo básico para establecer propiedades sobre los recursos y el lenguaje de ontologías *OWL* [14, 15] cual tiene como punto de partida experiencias previas realizadas en el lenguaje *DAML-OIL* [16]. El *OWL*, es un lenguaje de marcado con lógicas descriptivas, el cual se divide en tres sublenguajes: el *OWL-Lite*, el *OWL-DL* y el *OWL-Full*, siendo el primero el más sencillo. Todos estos lenguajes se sustentan para su formalización en el *XML* [17], un metalenguaje que nos permite definir y validar nuestros propios lenguajes de etiquetado que se usan en la Web.

3. LOS TESAURUS Y EL SKOS-CORE

Los tesauros documentales son instrumentos de control terminológico utilizado para convertir a un lenguaje normalizado el idioma natural de los documentos. Se

estructuran en listados de términos que tienen entre ellos relaciones genéricas, y que se aplican a un dominio concreto del conocimiento. Asociadas a los tesauros existen diferentes normas internacionales ANSI Z39:19:2003 [18, 19], ISO 2788-1986 [20], ISO 5964-1985 [21], las cuales definen lineamientos para su construcción y mantenimiento.

En términos generales, la forma de presentación de los tesauros tiene los siguientes elementos, agrupados en cuatro secciones [22, 23]:

- 1 *Lista Alfabética de Descriptores*. Es decir es una lista de términos autorizados y formalizados (Descriptores). Un gran número de descriptores manejan características particulares de los términos tales como: Nota de alcance (NA), si es un término más genérico (MG), si es un término relacionado (VT) ó un término más Específico (ME).
- 2 *Lista Temática de Descriptores*. Relaciona los descriptores y sus sinónimos (USE) o cuasi-sinónimos (UP) de los términos preferentes. Aparecen agrupados en temas y sub-temas y están ordenados en cada sub-tema alfabéticamente.
- 3 *Lista Jerárquica de Descriptores*. Muestra en orden descendiente un conjunto de términos considerados menos amplios que el término más genérico (TGS), sin embargo no todos los términos están acá, ya que cada descriptor no siempre va acompañado de un MG.
- 4 *Índice Permutado de Descriptores*. Acá cada descriptor y sinónimo se encuentra ordenado alfabéticamente según su palabra clave.

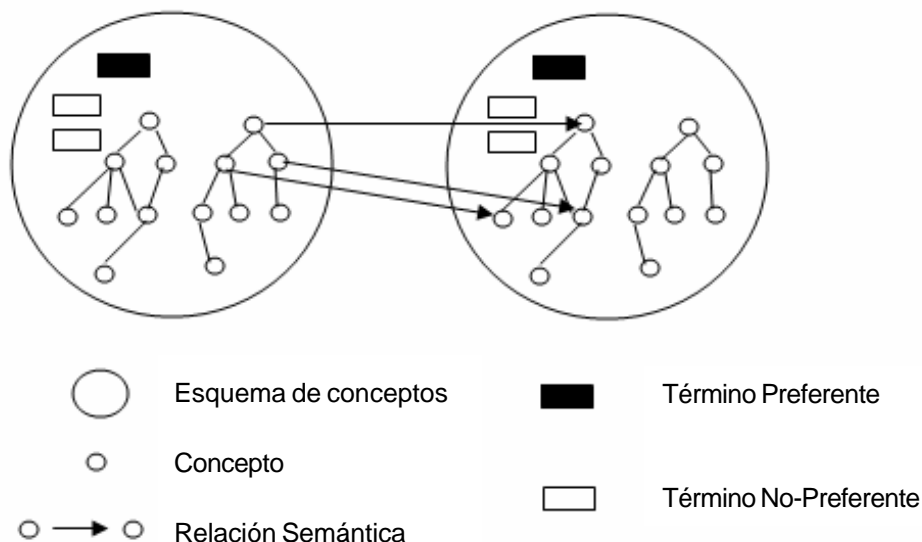
En un tesoro también existe el manejo de relaciones entre los descriptores. Es importante resaltar tres tipos de relaciones básicas [23]: (a) Relaciones de equivalencia, es la relación entre descriptores o términos preferidos y los no descriptores o términos no usados en la indización referentes a un mismo concepto (b) Relación jerárquica, es la relación vertical entre todos los descriptores de una misma clase, expresada en términos de subordinación de los conceptos en un doble sentido (c) Relaciones de Asociación, indican la relación o uniones en la significación de los descriptores (dos descriptores que son susceptibles de evocarse mutuamente por asociación de ideas). Todas las relaciones son recíprocas (*A* es a *B* lo que *B* es a *A*).

Si bien la aparición del *XML* marca el comienzo de una tendencia en la implementación de tesauros, la aparición del *RDF* supone un punto de inflexión en cuanto a la creación de una infraestructura semántica. Por ello no es raro suponer que la primera propuesta para la conversión de un tesoro se hubiera orientado a utilizar el *RDF* como mecanismo básico para su implementación, siendo luego utilizado el *OWL*, donde su versión *OWL-LITE* ha sido la más recomendada para transformar tesauros en el ambiente de la web semántica. Finalmente, una propuesta

más concreta para la representación de tesauros documentales en la web semántica ha sido el *SKOS-Core* [24, 25] el cual consiste en un *RDF-Schema* [26] para la representación de tesauros y sistemas similares de organización del conocimiento. Además, de ser una propuesta formal validada por la *W3C* [27], esta pensada para ser un complemento al *OWL*, ya que proporciona un marco básico para la construcción de conceptos. Estas dos características hacen del *SKOS-Core*, una herramienta valiosa y útil para recopilar el conocimiento almacenado en un tesauro.

En el *SKOS-Core*, la idea base del uso de un *RDF Schema* reside en su capacidad para permitir la definición de *conceptos* y *esquemas de conceptos* [24]. Un concepto se define como una unidad de pensamiento que puede ser definida o descrita. A su vez un esquema de conceptos, no es otra cosa que una colección de conceptos. Cada concepto solo puede tener una etiqueta preferente [24], lo que los documentalistas denominan *Descriptor* o *Término Preferente* y un número limitado de etiquetas alternativas, denominado *Término No-descriptor* o *No-preferente* (ver figura 1).

FIGURA 1
Relaciones básicas del SKOS-Core



Fuente: <http://www.w3.org/2001/sw/Europe/reports/thes/1.=/guide/>

Las relaciones que encontramos entre los descriptores de los tesauros tradicionales pueden ser precisadas entre los conceptos del esquema de conceptos, definidos en el *SKOS-Core*, con la salvedad de que tales relaciones pasan a ser de carácter semántico.

En la tabla 1, se detallan las propiedades y relaciones manejadas por el SKOS-Core, donde el prefijo Skos es una forma de identificar tal definición como un elemento de esta ontología.

TABLA 1
Descripción de las propiedades
y relaciones definidas en la ontología SKOS-Core.

Propiedad SKOS-Core	Tipo / Actua	Definición	Relación
Skos:prefLabel	Terminos	Preferente o Descriptor	Equivalencia
Skos:altLabel	Terminos	No-preferente o No-descriptor	
Skos:externalID	Conceptos	Identificadores para los conceptos	
Skos:scopeNote	Conceptos	Definición de concepto, notas aclaratorias	
Skos:definition	Conceptos	Definición de concepto	
Relaciones Semanticas	Tipo / Actua	Definición	Relación
Skos:narrower	Terminos	Terminos especificos	
Skos:broader	Terminos	Terminos generales	
Skos:related	Terminos	Terminos relacionados	
Generic	Tipo / Actua	Hereda de subclassof	Subsuncion
Skos:broaderGeneric	Terminos	Herencia de terminos especificos	
Skos:narrowerGeneric	Terminos	Herencia de terminos generales	
Instative	Tipo / Actua	Hereda de type	Un concepto es instancia de otro
Skos:broaderInstative	Conceptos	Un concepto es instancia de un concepto especifico	Skos:relatedHasPart
Skos:narrowerInstative	Conceptos	Un concepto es instancia de un concepto general	Skos:relatedOf

Fuente: Cuadro resumen generado <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-200510>.

La codificación de las etiquetas correspondientes a los *Términos Preferentes* y *No Preferentes* pertenecientes a un *Concepto*, *skos:Concept*, se realiza mediante las propiedades *skos:preflabel*, para los términos preferentes y *skos:altLabel*, para los no preferentes o descriptores. Esta segunda etiqueta representa la relación de sinonimia o equivalencia del tesoro. La utilización de notas aclaratorias es otra necesidad que se representa con el *skos:ScopeNote* y las definiciones de conceptos se representan mediante *skos:definition*.

Adicionalmente, el *SKOS-Core* permite representar las relaciones básicas entre los conceptos manejadas por la mayoría de los tesauros mediante la familia *skos:semanticRelation*, bajo la cual se encuentran las relaciones jerárquicas definidas por las propiedades *skos:narrower* y *skos:broader*, para términos específicos y generales respectivamente y *skos:related* para términos relacionados.

El *SKOS-Core* también posee la forma de especificar otros tipos de propiedades que representan una forma de ampliar las relaciones semánticas entre conceptos, estas propiedades sobrepasan lo actualmente considerado en un tesoro tradicional,

es así como aparecen: (a) *skos:broaderGeneric* y *skos:narrowerGeneric*, para definir únicamente relaciones de subfunción entre dos conceptos, donde un concepto es subclase de otro, (b) *skos:broaderInstative* y *skos:narrowerInstative*, donde *Instative* expresa que un concepto es una instancia de otro, y (c) *skos:broaderPartitive* y *skos:narrowerPartitive*, donde *Partitive* indica que un concepto es parte de otro (es una forma de especificar que un concepto es parte de otro).

Por último, este también permite expresar refinadamente la relación de asociación con las propiedades *skos:relatedHasPart* y *skos:relatedPartOf*, las cuales expresan que un concepto tiene parte de otro concepto o que un concepto esta relacionado con parte de otro concepto respectivamente.

4. DISEÑO DE UNA ONTOLOGIA ESPECIALIZADA EN EL HABITAT

4.1 Descripción del Tesouro Tradicional especializado en el Hábitat

El tesouro documental del Centro de Naciones Unidas sobre Asentamientos Humanos [7] contiene un repertorio de términos que intenta cubrir las numerosas preocupaciones que tienen las Naciones Unidas, relacionadas con el tema de los asentamientos humanos. Sin embargo, para el desarrollo de este proyecto se ha tomado una adaptación de este tesouro, el cual hace parte de la Unidad de Documentación de la Escuela del Hábitat CEHAP perteneciente a la Facultad de Arquitectura de la Universidad Nacional de Colombia, sede Medellín. Por pertenecer a esta área lo hemos denominado Tesouro del Hábitat. El tesouro del Hábitat del CEHAP es un tesouro monolingüe, de estructura en forma de árbol, presentado alfabéticamente, cuenta con 1263 términos descriptores, que son los términos autorizados y formalizados por el tesouro sobre Asentamientos Humanos de la UNESCO. El término descriptor se distingue dentro del tesouro por ser el primer término que aparece inmediatamente después de la numeración y del cual se van a describir sus características particulares NA (Nota de alcance), MG (Término más genérico al término descriptor), VT (Término relacionado con el término descriptor), USE (Sugiere un sinónimo con el cual se puede referenciar el termino descriptor), y UP (Indica que tal término descriptor es Usado por) , cada cambio de número indica un nuevo término descriptor.

En la figura 2, se presentan apartes de este tesouro, donde se puede apreciar el término preferente Vivienda y las abreviaciones que recopilan a otros términos con quienes el término vivienda guarda cierto significado (dependiendo de la abreviación).

FIGURA 2
Visualización de apartes del tesaur del Hábitat

001191	VIVIENDA UP Alojamiento Habitación MG HABITAT ME APARTAMENTOS CASAS VIVIENDA COLECTIVA VT ASENTAMIENTOS PRECARIOS BARRIOS DE TUGURIOS CONDICIONES DE VIVIENDA COOPERATIVAS DE VIVIENDA ENCUESTAS HABITACIONALES NECESIDADES DE VIVIENDA VIVIENDA DE ALQUILER VIVIENDA DE EMERGENCIA VIVIENDA EN ALTURA VIVIENDA EN PENDIENTE VIVIENDA ESTATAL VIVIENDA INDIGENA VIVIENDA POPULAR VIVIENDA POR AUTOCONSTRUCCION VIVIENDA PALAFITICA VIVIENDA PROGRESIVA VIVIENDA PUBLICA VIVIENDA RURAL VIVIENDA URBANA TT CNUAH (HABITAT) 1986, p.60	001199	Vivienda de bajo costo USE VIVIENDA POPULAR
		001200	VIVIENDA DE EMERGENCIA UP Refugio de emergencia MG VIVIENDA VT DESASTRES MITIGACION DE DESASTRES PLANIFICACION POSTERIOR A DESASTRES VIVIENDA PROVISORIA TT CNUAH (HABITAT) 1986, p.60
		001211	VIVIENDA DE INTERES SOCIAL UP Vivienda de bajo costo VT VIVIENDA POPULAR
		001202	Vivienda económica USE VIVIENDA POPULAR TT CNUAH (HABITAT) 1986, p.60
		001203	Vivienda elemental USE VIVIENDA PROGRESIVA TT CNUAH (HABITAT) 1986, p.60
		001204	VIVIENDA EN ALTURA NA Edificios residenciales de tres pisos... UP Edificios altos MG VIVIENDA VT APARTAMENTOS DIRECCION DE LA VIVIENDA TT CNUAH (HABITAT) 1986, p.60
001193	Vivienda ampliable USE VIVIENDA PROGRESIVA TT CNUAH (HABITAT) 1986, p.60		
001194	Vivienda básica USE VIVIENDA PROGRESIVA 05.04.1988		
001195	VIVIENDA COLECTIVA		

Fuente: Tesaur de la Unidad de Documentación de la Escuela del Hábitat, Facultad de Arquitectura, Universidad Nacional sede Medellín.

4.2 Diseño de la Ontología Mediante el SKOS-Core

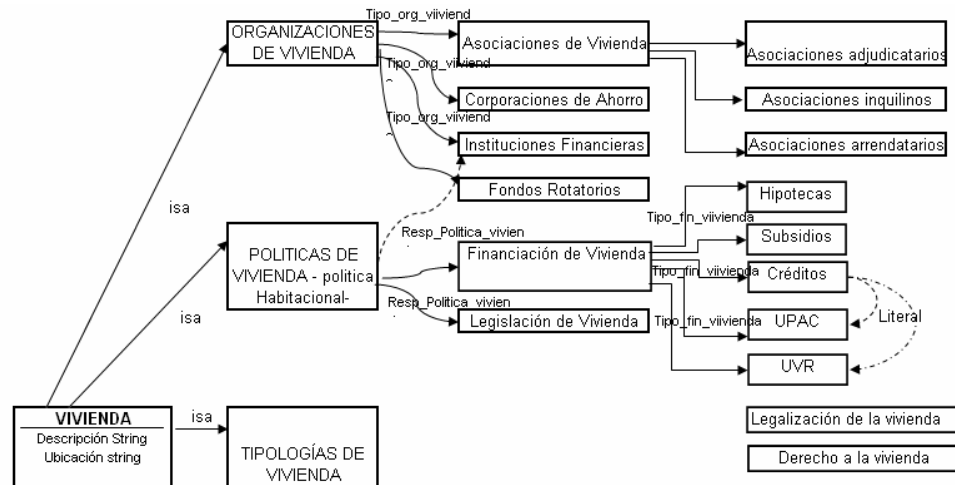
Dadas las características descritas en la sección 3 sobre el SKOS-Core, se seleccionó este como mecanismo para definir la ontología que representaría el conocimiento de las relaciones entre los conceptos contenidos en el Tesaur del Hábitat. Para que la representación realizada en el SKOS-Core fuera de utilidad en la ontología del Hábitat que se deseaba implementar, fue necesario representar de modo explícito y formalizado, no sólo las características que definen los términos incluidos en el tesaur, sino también establecer las mismas relaciones que se encuentran en los tesauros tradicionales (jerarquía, equivalencia y asociación) además de proporcionar a cada término importantes capacidades de expresión de las relaciones semánticas.

La conversión del conocimiento almacenado en el tesoro original a la ontología deseada, partió de las siguientes modificaciones: (a) representar cada término en un único concepto identificador, para lo cual es importante tener en cuenta que cada término se va a convertir en concepto, pero que dos conceptos no pueden llamarse exactamente igual; (b) convertir la jerarquía existente en una jerarquía de conceptos consiste en agrupar por categorías los conceptos, lo cual se puede hacer a partir de cualquiera de los dos principios, de lo general a lo particular o de lo particular a lo general; (c) adicionar el conocimiento acerca de la relación entre cada concepto, para lo cual se deben adicionar relaciones como asociación entre conceptos ó equivalencia.

Para llevar a cabo la conversión se acudió a la experiencia que el documentalista experto en el tema podía aportar. En primer lugar, se procedió a extraer del tesoro los términos más relevantes y cuyo orden jerárquico dependía directamente del término Hábitat (término más general en el esquema de conceptos). En la adaptación de este tesoro de Hábitat, se establecieron siete categorías principales que cubren sus temáticas más generales de acuerdo con el documentalista, donde cada una de estas categorías representa un concepto: Asentamientos, Ciudad, Desarrollo, Arquitectura y Construcción, Vivienda, Medio Ambiente, Administración Pública. A cada categoría (concepto), una a una, se le realizó un mapeo de las relaciones de herencia que tenían, es decir se efectuó el recorrido a todo el tesoro y se fueron extrayendo otros conceptos que pudieran estar referenciados como conceptos subordinados de ese concepto, pero que a su vez pudieran categorizar otros conceptos cada vez más específicos. El proceso anterior se repitió con cada una de las siete categorías, generándose un árbol jerárquico por cada una de ellas. Con el fin de brindar una idea clara del proceso realizado en el presente trabajo, se seguirán apartes de los pasos realizados tomando como ejemplo la categoría Vivienda.

En la figura 3, se detalla el resultado del procedimiento anterior para el caso de la categoría Vivienda. Bajo esta categoría, que es uno de los conceptos directamente subordinado del concepto hábitat, se agruparon otras subcategorías aun más específicas como por ejemplo Organización de Vivienda, Políticas de Vivienda, Tipologías de Vivienda, etc. Bajo cada una de las anteriores se agrupan otras categorías. En el caso de Políticas de Vivienda, se asociaron conceptos como Financiación de Vivienda y Legislación de Vivienda, donde Políticas de Vivienda asociaba conceptos como Subsidios, Créditos, UPAC y UVR, el proceso se repitió para cada categoría, hasta haber recorrido las siete inicialmente definidas.

FIGURA 3.
Jerarquía de conceptos para el caso de “Vivienda”



Fuente: Gráfico extractado del tesoro del hábitat donde se relacionan apartes de los términos más específicos del término descriptor vivienda.

El segundo paso consistió en definir las relaciones de asociación entre los términos, las cuales, a diferencia de la relación de jerarquía, indican si un término tiene una similitud temática con otro término. En la figura 3, la relación de asociación está determinada por las líneas punteadas, pese a que este proceso se realizó con todos los términos descriptores de cada categoría, varios términos no fueron relacionados al no existir otros que pudieran asociárseles. La obtención de estas relaciones se basó en la experiencia que tiene el documentalista sobre el tema y el conocimiento que manifestaron varios usuarios sobre los temas.

Para ampliar un poco más la idea, la figura 4 detalla algunos términos descriptores de vivienda, que pueden tener asociados a otros conceptos, sin importar que pertenezcan o no a la misma categoría vivienda. Se han conservado algunas notas de alcance, para aclarar aún más a lo que hace referencia exactamente ese concepto descriptor.

FIGURA 4.
Relaciones de asociación de algunos descriptores del término Vivienda

ORGANIZACIONES DE VIVIENDA NA Comprende todo tipo de organización relacionada con la solución de la vivienda y de hábitat tales como asociaciones de arrendatarios de vivienda, etc. adjudicatarios. PROCESOS HABITACIONALES POPULARES CORPORACIONES DE AHORRO Y VIVIENDA FINANCIACION DE VIVIENDA	REFORMA URBANA POLITICA URBANA
INSTITUCIONES FINANCIERAS FINANCIACION DE VIVIENDA	TIPOLOGIAS DE VIVIENDA
COOPERATIVAS DE VIVIENDA NA Incluye la vivienda adquirida o construida eg. forma cooperativa, y los organismos y organizaciones responsables. FINANCIACION DE VIVIENDA FONDOS ROTATORIOS FINANCIACION DE VIVIENDA	VIVIENDA EN ALTURA NA Edificios residenciales de tres pisos o más. CONJUNTOS HABITACIONALES
POLITICA DE VIVIENDA LEGISLACION DE LA VIVIENDA LEGALIZACION DE VIVIENDA PROGRAMAS DE VIVIENDA POLITICA DE ASENTAMIENTOS VIVIENDA POPULAR	VIVIENDA INDIGENA
FINANCIACION DE VIVIENDA NA Métodos para financiar la construcción, la adquisición o la compra hipotecaria de viviendas. CORPORACIONES DE AHORRO Y VIVIENDA INSTITUCIONES FINANCIERAS COOPERATIVAS DE VIVIENDA FONDOS ROTATORIOS	VIVIENDA PALAFÍTICA
LEGISLACION DE LA VIVIENDA POLITICA DE VIVIENDA	VIVIENDA POPULAR LOTES Y SERVICIOS POLITICA DE VIVIENDA PROGRAMAS DE VIVIENDA VIVIENDA PUBLICA
DERECHO A LA VIVIENDA DERECHO URBANO	VIVIENDA TRADICIONAL ARQUITECTURA TRADICIONAL TECNOLOGIA APROPIADA
LEGALIZACION DE VIVIENDA POLÍTICA DE VIVIENDA	VIVIENDA RURAL ASENTAMIENTOS RURALES
	VIVIENDA DE EMERGENCIA ALBERGUES PLANIFICACION POSTERIOR A DESASTRES
	VIVIENDA PREFABRICADA COORDINACION MODULAR
	CONJUNTOS HABITACIONALES VIVIENDA EN ALTURA
	VIVIENDA EN PENDIENTE VULNERABILIDAD
	VIVIENDA PROGRESIVA NA Se refiere tanto a las etapas de parcelación, loteo, adecuación del terreno, etc., como a las relaciones con la construcción de la vivienda a partir de un núcleo básico. LOTES Y SERVICIOS

Los dos pasos, desarrollados anteriormente, facilitan el proceso de codificación del tesoro con las propiedades *SKOS-Core*, puesto que la jerarquización en árbol de donde se precisa la relación de herencia, proporciona una forma más rápida de asociar los conceptos a la propiedad jerárquica del *skos:SemanticRelation*, más exactamente *skos:broader* o *skos:narrower*. Para ello, los conceptos son tomados únicamente de los términos descriptores, razón por la cual cada uno de ellos estará asociado con la propiedad *skos:prefLabel*. Así mismo, el *skos:altlabel* y el *skos:ScopeNote*, estarían definidos directamente por las abreviaciones ya definidas del tesoro USE y Nota de Alcance, respectivamente. También en la definición de la propiedad *skos:altlabel*, por ser la que establece la relación de equivalencia,

se tuvo especial cuidado de asignar otros conceptos que podían representar las variaciones lingüísticas y el uso del lenguaje natural. Por último, la propiedad *skos:related*, se implementó con las relaciones definidas anteriormente en el proceso de asociación.

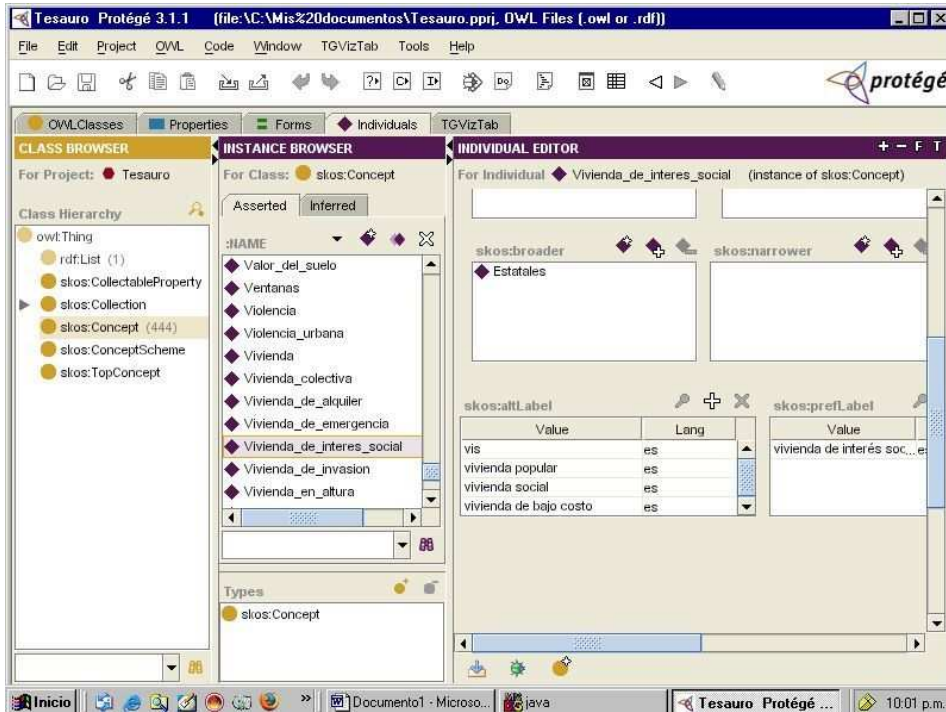
Luego de representar con el *SKOS-Core* todas las relaciones establecidas en el tesoro tradicional del hábitat, el paso que se siguió fue implementar la ontología.

5. IMPLEMENTACIÓN DE LA ONTOLOGÍA

Para implementar la ontología se selecciono Protégé v.3.1 [28] con OWL como lenguaje de representación (ver figura 5) debido a las características que presenta como son su funcionamiento bajo WINDOWS y la amigable interfaz de usuario que emplea.

En esta herramienta se importó la ontología del SKOS-Core, la cual se muestra en el lado derecho de la figura 5. Al seleccionar la opción de adicionar un concepto, Protégé abre una ventana para cada propiedad *skos* dentro de un mismo *frame*. El proceso radica entonces en ir al campo concepto, y en la propiedad *skos:prefLabel* adicionar el concepto descriptor y en la ventana *altLabel* se especifican los conceptos alternativos, que hayan sido asociados a este concepto. Adicionalmente, encontramos otra ventana con la propiedad *skos:broader*, donde se escribe el padre de este concepto, es decir el concepto inmediatamente más general y en la ventana *skos:narrower*, se anexan todos los conceptos subordinados a este concepto. En la relación de asociación se escriben los conceptos que guardan alguna relación con el concepto que se esta adicionando. Además, de las propiedades *skos*, anteriormente mencionadas, existe una denominada *skos:Hidden*, que alberga todas las otras formas en que se puede referenciar el concepto. Como la relación de equivalencia ya había sido definida con la propiedad *skos:altLabel*, entonces acá se describen errores ortográficos de digitación o se adiciona el concepto en singular o plural. En los campos *altLabel* se ingresó un número de conceptos determinados en primera instancia por los conceptos de las siete categorías consideradas principales y que fueron descritas anteriormente: Asentamientos, Ciudad, Desarrollo, Arquitectura y Construcción, Vivienda, Medio Ambiente, Administración Pública se ingresaron una a una, se realizó un mapeo de las relaciones de herencia que tenían, y a cada uno de sus conceptos mas específicos se sugirió generarlos también como conceptos. Al terminar el proceso se obtuvo una ontología especializada en hábitat con 444 conceptos.

FIGURA 5.
Representación del concepto Desarrollo, en la herramienta Protégé 3.1



Fuente: Proyecto de Investigación: Sistema de Recuperación de Información Semántico Especializado en Artes Plásticas y Hábitat

6. CONCLUSIONES Y TRABAJOS FUTUROS

Este artículo representa una experiencia de aplicación de las tecnologías de la web semántica en un escenario real y aporta una definición de una ontología en un dominio concreto como lo es el hábitat. Este documento se convierte entonces en una guía alternativa para trasladar un tesoro documental a una ontología de dominio específico.

La selección del *SKOS:Core* para el proceso de conversión facilitó dicho trabajo, ya que se pudo rescatar, casi de manera directa y de forma inmediata la mayoría de criterios propios del tesoro, tales como: la estructura de herencia y las relaciones de sinonimia. Adicionalmente, el *Skos:Core*, permitió ampliar el conocimiento del tesoro al asociar a los descriptores errores ortográficos, con la propiedad *skos:Hidden*.

Dado que actualmente la ontología solo lleva implementados 444 de los 1263 términos del tesoro original, un trabajo futuro permitirá completar los términos faltantes.

Otro trabajo futuro podrá emprender la descripción de los descriptores en otras lenguas ya que el *SKOS-Core* cuenta con mecanismos para generar una representación multilingüe.

Por último, esta ontología será probada en una herramienta de indexación y recuperación que explote cada una de las propiedades SKOS-Core y supere procesos tradicionales. Así las cosas, los documentos estarían representados por las instancias de cada concepto de la ontología, donde un usuario podrá, no solo consultar y descargar un documento, gracias al concepto por el cual indagó, sino que también la consulta podrá sugerir otros conceptos asociados o términos más generales o más específicos, por los cuales pueda buscar, y así ampliar o precisar su búsqueda. Además de ello, si el usuario indaga por un concepto alternativo, él realizará la consulta de la misma forma que si hubiera preguntado por el concepto preferente, y podrá hacer lo mismo si la consulta se ha realizado cometiendo alguno de los errores ortográficos definidos en el *skos:Hidden*. Además proporciona la posibilidad de que el propio sistema de recuperación realice inferencias y entregue resultados alternos a una consulta formulada.

7. REFERENCIAS BIBLIOGRÁFICAS

- [1] ISO 2788:1986 - Guidelines for the establishment and development of monolingual thesauri. International Organization for Standardization (ISO), 1986.
- [2] AITCHISON, J., GILCHRIST, A., and BAWDEN, D. Thesaurus construction and use: a practical manual. 4th ed. Chicago, IL, USA: Fitzroy Dearborn. 2000.
- [3] LÓPEZ-HUERTAS, M. J. Potencialidad evolutiva del tesoro: hacia una base de conocimiento experto. En: La Representación y la Organización del Conocimiento en sus distintas perspectivas: su influencia en la Recuperación de la Información (Actas del IV Congreso ISKO-España EOCONSID99, 22-24 de abril de 1999 en Granada), 1999. Granada: ISKO-Facultad de Biblioteconomía y Documentación. 133-140.
- [4] ANTONIOU G. and VAN HARMELET, F. A Semantic Web Premier, the MIT Press, Cambridge Massachusetts, 2004, p 1-12.
- [5] Ibid., p.8.

- [6] STAAB, S. and STUDER, R. Handbook on Ontologies, International Handbooks on Information Systems. Berlin : Springer, , 2005. p. 617
- [7] UNESCO. Unesco Thesaurus. <http://www.ulcc.ac.uk/unesco/>
- [8] STAAB, S. and STUDER, R., Op Cit., p.133
- [9] FOSKETT, D. Thesaurus. En K. Spark Janes and P. Willet editors. Readings in Information Retrieval. Morgan Kaufmann Publishers, Inc., 1997- pp 111-134,.
- [10] GRUBER T. R. Toward principles for the design of ontologies used for knowledge sharing. En: Formal ontology in conceptual analysis and knowledge representation. The Netherlands: Kluwer Academic Publishers, 1993.
- [11] GOMEZ-PEREZ A., CORCHO O, FERNANDEZ-LOPEZ M., Ontological Engineering : Advanced information and knowledge processing. Berlin : Spinger, 2003.
- [12] W3C: <http://www.w3.org/TR/RDF/rdf-premier> [Consulta 10 de Julio de 2006]
- [13] ANTONIOU G. and VAN HARMELET F, Op Cit., 61-106
- [14] <http://www.w3.org/2004/OWL> [Consulta 11 de octubre de 2006]
- [15] ANTONIOU, G. and VAN HARMELET, F, Op Cit., 109-148
- [16] <http://www.w3.org/TR/daml+oil-reference> [consulta: Octubre 2 de 2006]
- [17] <http://www.xmlnews.org> [Consulta: Octubre 11 de 2006]
- [18] ANSI/NISO Z39.19 - Guidelines for the Construction, Format, and Management of Monolingual Thesauri. 2003
- [19] <http://www.hipertexto.info/documentos/tesauros.htm> [Consultado 15 de octubre de 2006]
- [20] ISO 2788:1986 - Guidelines for the establishment and development of monolingual thesauri. International Organization for Standardization (ISO), 1986.
- [21] ISO 5964:1985 - Guidelines for the establishment and development of multilingual thesauri. International Organization for Standardization (ISO), 1985.
- [22] GIL URDICIANI, B. Manual de lenguajes documentales. Madrid:Noesis. 1996

- [23] AITCHISON, Jean., GILCHRIST, Alan and BAWDEN, David. Thesaurus construction and use: a practical manual. 4th ed. Chicago, IL, USA: Fitzroy Dearborn, 2000
- [24] <http://www.w3.org/2001/sw/Europe/events/200406-esp/trabajo-final-extratesauros/node6.html> [consulta: abril de 2006]
- [25] GOMEZ-PEREZ A., M. FERNANDEZ-LOPEZ, A. de Vicente (1996), Towards a Method to Conceptualize Domain
- [26] ANTONIOU G. and VAN HARMELET F, Op Cit., 91-99
- [27] <http://www.w3c.es/> {consultado octubre 16 de 2006}
- [28] <http://protege.stanford.edu/plugins/owl/> [consultado 1 de Junio de 2006]
- [29] <http://wordnet.princeton.edu/> [Consultado 16 de octubre de 2006]
- [30] <http://www.uned.es/sel/35Simposio/Resumenes/Dones-Ortiz.doc> [Consultado 16 de octubre de 2006]
- [31] <http://www.ontopia.net/topicmaps/materials/tao.html> [Consultado 16 de octubre de 2006]
- [32] NOY, N.F. MCGUINNESS D.L. Ontology Development 101: A guide to creating your first ontology. Disponible en <http://www.Ksl.stanford.edu/people/dim/papers/ontology-tutorial-noy-mcguinnessabstract.html> [consulta: diciembre de 2005]