

# SISDEON: Un sistema de información sobre el desplazamiento forzado en Colombia mediante el uso de ontologías

## SISDEON: An information system on displacement in Colombia using ontology

Nestor Eduardo Rivera Vidal (c) I.S.<sup>1</sup>, Grissa Vianney Maturana González. (c) I.S.<sup>1</sup>, Sandra Arenas, M.Sc.<sup>2</sup>, Ana María Tangarife. Bibl.<sup>2</sup>, y Jaime Alberto Guzman. M.Sc.<sup>1</sup>.

1. Grupo Sintelweb, Universidad Nacional de Colombia - Sede Medellín, Colombia

2. Escuela interamericana de Bibliotecología, Universidad de Antioquia, Colombia  
jaguzman@unal.edu.co; comunicaciones@bibliotecologia.udea.edu.co

Recibido para revisión 20 de Agosto de 2008, aceptado 23 de Octubre de 2009, versión final 1 de Diciembre de 2009

**Resumen**—Para resolver el problema de la ambigüedad, redundancia y poca consistencia del lenguaje natural utilizado por los investigadores, se busca modelar el conocimiento por medio de estructuras como los tesauros, este es el caso del fenómeno del Desplazamiento Forzado en Colombia. Y no solo modelar información organizada y concisa es hoy uno de los objetivos de la Web semántica, al elaborar dichas estructuras del conocimiento, sino también el libre acceso a la información es su gran pretensión. Para lograr dicho objetivo Sisdeon se apoya en el uso de ontologías y de herramientas actualizadas que permiten el manejo de estas a través de la Web y que buscan ampliar el rango de posibilidades usando la lógica descriptiva como apoyo y estableciendo así consultas más complejas; además de ampliar el manejo la recuperación de documentos a través de razonadores semánticos. Es allí donde radica la importancia a de un sistema de información basado en el uso de ontologías: herramientas modernas que dejan profundizar, manejar y recuperar de manera eficiente los documentos.

**Palabras Clave**—Recuperación de Información, Ontologías, Web semántica, Desplazamiento forzado en Colombia., SKOS-Core.

**Abstract**—To solve the problem of ambiguity, redundancy and little consistency of natural language used by researchers to model the knowledge, many people use special structures such as thesauri, that is the case of the phenomenon of forced displacement in Colombia. Now one model of organized and concise information is not the only objective of the Semantic Web to develop these structures of knowledge, but also free access to information is one of its principal goals. To achieve this goal, Sisdeon relies on the use of ontologies and updated tools that enable the handling of these ontologies through the Web and enable to widen the range of possibilities with the support of the Description Logic and establishing more complex queries with goal to expand the documents retrieval using semantic reasoning. This is where lies the importance of an information system based on the use of ontologies: modern tools that permit deeper, manage and retrieve documents efficiently.

**Keywords**—Information Retrieval , Ontologies, Semantic Web, forced displacement in Colombia, SKOS-Core.

### 1. INTRODUCCIÓN

Actualmente los tesauros son de gran utilidad en organizaciones del conocimiento sobre temas concretos, por ser una herramienta de control terminológico muy útil para el análisis, descripción y recuperación automatizados. Pero son de mucha más utilidad cuando estos permiten tener acceso desde la red, ya sea Internet o intranet, ya que al ser procesado como un documento digital es más amplia su frontera. Para la automatización de tesauros a través de la Web semántica, se han desarrollado lenguajes que en conjunción con XML ofrecen una potente y rica estructuración semántica, como RDF, RDFS y lenguajes semánticos como DAML-OIL y OWL para todo tipo de esquemas y ontologías. Para estos lenguajes, también se han desarrollado modelos especializados en el manejo de estas estructuras, siendo representativas de estos Topic Maps, Zthes y SKOS Core [1].

Al tener estructurado el conocimiento usando los modelos mencionados, buscamos ahora guardar la información de los tesauros en bases de datos para su posterior manipulación, es entonces cuando encontramos herramientas que nos facilitan dicha operación como es el API de Jena, basado en Java. Luego se pretende recuperar la información almacenada haciendo consultas semánticas que permitan al usuario encontrar la información requerida, para ello es útil los lenguajes de consulta que nos permiten manipular los recursos digitales y nos abre la posibilidad de realizar razonamientos basados en la lógica descriptiva como el RDQL , SPARQL y el Pellet. Y finalmente para lograr nuestro objetivo de permitir que estas bases del conocimiento sean asequibles desde la Web, usamos como creador de páginas interactivas la tecnología JSP.

Todo este proceso y todas estas herramientas pretenden finalmente modelar el tesoro que motivo esta investigación, un tesoro sobre el desplazamiento forzado en Colombia para resolver el problema de ambigüedad, redundancia y poca consistencia que se tiene de dicho fenómeno y que además permita su acceso desde la Web.

## II. OBJETIVOS

- Realizar un tesoro usando ontologías para su representación y que pueda ser visualizado a través de la web.
- Usar en el sistema de información tecnología basada en la web semántica para el pleno razonamiento entre los conceptos.
- Realizar una aplicación dinámica que permita insertar términos nuevos al tesoro

## III. MATERIALES Y MÉTODOS

En áreas especializadas de conocimiento se hacen necesarias estructuras que ayuden a modelar dicho conocimiento, entre ellas encontramos los tesauros y las ontologías; definiendo por tesoro un tipo de lenguaje combinatorio que consta de listas de términos que representan un ámbito científico y técnico determinado y que posee una serie de relaciones semánticas entre los términos que lo conforman. Estas relaciones semánticas son de tres tipos concretos: equivalencia, asociación y jerarquía y se basa en expresiones conceptuales llamadas descriptores, provistas de las citadas relaciones semánticas [2]. Las ontologías son un sistema de representación del conocimiento que resulta de seleccionar un dominio o ámbito del conocimiento, y aplicar sobre él un método con el fin de obtener una representación formal de los conceptos que contiene y de las relaciones que existen entre dichos conceptos. Además, una ontología se construye en relación a un contexto de utilización [3].

### A. Representación de Tesauros en la Web Semántica

Para implementar un tesoro en la Web se requiere estudiar las distintas formas de codificación que existen en la Internet, los más conocidos son RDF/XML, OWL, SKOS-Core y otros menos usados son Zthes y Topic Maps. XML crea una infraestructura semántica que da soporte a la información que existe en Internet mientras que RDF (Resource Description Framework), se basa en idea de que podamos identificar los elementos a partir de URIs escribiendo los recursos en términos de propiedades simples o pares propiedad/valor, lo cual permite representar las declaraciones simples sobre recursos como un grafo de nodos y arcos que representan los recursos, sus propiedades y sus valores. Si XML es un punto inflexión para RDF, este a su vez lo es para OWL que representa las ontologías a partir de un vocabulario más amplio y una sintaxis más fuerte; a su vez OWL es basado en las experiencias previas realizadas por DAML-OIL. OWL, desarrollado por el consorcio W3C se divide en tres sublenguajes OWL Lite que provee una clasificación jerárquica y limitadas restricciones, OWL DL proporciona máxima expresividad mientras asegura completa computacionalidad y decidibilidad y OWL Full proporciona máxima libertad sintáctica.

Es importante ahora mencionar que la ontología en la que se describen cada uno de los conceptos del área del conocimiento es creada basada en el lenguaje OWL[4] y como

complemento Skos-core[14], el cual proporciona un marco básico para la construcción de esquemas de conceptos pero sin la definición semántica tan estricta que exige la utilización de OWL. SKOS-Core es un schema RDF que permite la representación de tesauros y sistemas para la organización del conocimiento; a su vez, sirve para la construcción de esquemas de conceptos simples para su utilización en la Web. Las relaciones básicas que proporciona Skos-core se pueden observar en la Figura 1.

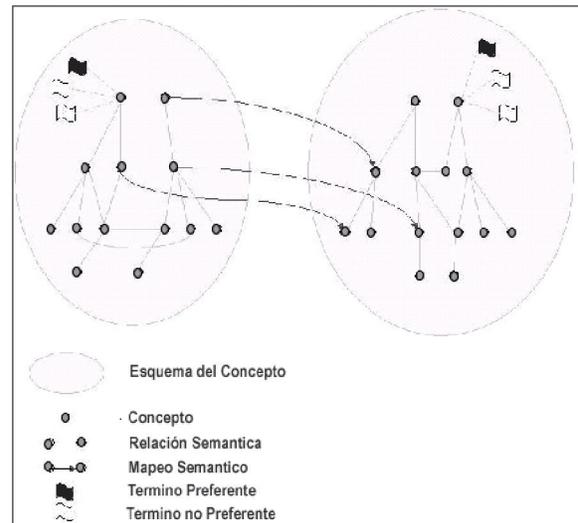


Figura 1. Relaciones básicas contempladas por Skos-core.

Skos-core provee una serie de propiedades que permiten la representación del conocimiento de una forma flexible, describiendo así el contenido y la estructura organizacional de los conceptos. Las propiedades más importantes son descritas a continuación (Véase Tabla 1):

### B. Manipulación y edición de ontologías

Cuando se tienen ontologías de gran tamaño (cientos o miles de conceptos y relaciones), se hace necesario ofrecer al usuario herramientas que no solo permitan obtener vistas útiles de la ontología sino también que estas permitan facilitar su manipulación y validación. Algunas de estas herramientas de manipulación y edición de ontologías permiten a través de una interfaz grafica crear nuevos conceptos y relaciones entre estos, y finalmente posibilitan exportar las ontologías a alguno de los lenguajes de representación semántica que lo soporte.

En el mercado se encuentran varias herramientas de visualización de ontologías, unas más completas que otras, más accesibles y útiles de acuerdo a las necesidades que se tengan: el OntoStudio (antes conocido como OntoEdit) es una herramienta de edición de ontologías que apoya el desarrollo y mantenimiento de las mismas utilizando medios gráficos en un entorno Web, pretende además que las ontologías puedan ser almacenadas y posteriormente ser manipuladas por una base de datos relacional. OntoEdit no es una herramienta libre,

Tabla 1. Propiedades Skos-Core

	Propiedad	Definición
<b>Clase Concepto</b>	skos:Concept	Unidad de conocimiento
<b>Etiquetas</b>	skos:prefLabel	Termino preferente
<b>Léxicas</b>	skos:altLabel	Termino no Preferente
	skos:hiddenLabel	Termino oculto
<b>Etiquetas</b>	skos:prefSymbol	Simbolo ó imagen preferente
<b>Simbólicas</b>	skos:altSymbol	Simbolo ó imagen no preferente
<b>Propiedades</b>	skos:definition	Definición de Termino
<b>documentales</b>	skos:scopeNote	Información parcial del Termino
<b>Relaciones</b>	skos:related	Terminos relacionados
<b>Semánticas</b>	skos:narrower	Terminos específicos
	skos:broader	Terminos generales
<b>Etiquetas de</b>	skos:Collection	Colección de terminos
<b>Colección</b>	skos:member	Miembro de una colección

por ello es limitada en uso [5]. Ontolingua provee de un entorno de colaboración distribuido para navegar, crear, editar, modificar y utilizar ontologías. Es una herramienta desarrollada por Knowledge Systems Laboratories in Stanford. Uno de sus aspectos más importantes es tener la habilidad de designar una ontología colaborativa, aspecto que la diferencia de otras herramientas que solo ofrecen el trabajo con una sola ontología.

WebOnto es un JAVA applet con un servidor Web personalizado que permite a los usuarios navegar y editar modelos de conocimiento sobre la Web. WebOnto provee una manipulación directa de la ontología por interfaces ricas y de fácil acceso [6]. El Protégé es un programa informático de tipo open source. Su función consiste en la creación, edición y mantenimiento de ontologías. Además de ser un programa exquisitamente diseñado en cuanto a interfaz y facilidad de funcionamiento posee la capacidad de generar de manera automática el código fuente de cualquier ontología creada con Protégé en formato RDF/OWL [7].

### C. Búsqueda de información organizada semánticamente

Al tener una ontología que modele datos, se pretende realizar búsquedas de términos a través de esta, y para esto usamos herramientas que faciliten dicha actividad: RQL es un lenguaje de consulta declarativo para RDF/S que captura la semántica en su diseño, Permite consultar esquemas RDF y descripciones RDF (instancias) en la misma consulta [8]. RDQL es un lenguaje de consulta declarativo para RDF. Considera un modelo RDF como un conjunto de tripletas: (objeto, propiedad, valor). Permite especificar patrones que son mapeados contra las tripletas del modelo para retornar un resultado. Se pueden realizar consultas en RDQL desde una aplicación java (por medio de Jena) [9]. SPARQL es otra herramienta que puede representar y utilizar los resultados obtenidos en las búsquedas a través de una gran variedad de

información como son datos personales, redes sociales y metadatos sobre recursos digitales.

### D. Consultas y razonamiento

Actualmente el razonador mas completo que existe es el Pellet, que al ser un razonador de OWL-DL basado en java, permite acoplarse perfectamente con APIs de java como el Jena, a su vez con herramientas editoras de ontologías como el SWOOP, Protégé o con herramientas que soportan reglas para ontologías como el SWRL. No solo esto hace de Pellet una herramienta muy fuerte en el mercado, ya que además de ofrecer las características básicas de cualquier razonador (verificar restricciones de cardinalidad, propiedades complejas de axiomas, clasificación, realización, entre otras), ofrece una caja completa de preguntas y respuestas, un razonador de tipos de datos, soporte de reglas, análisis y reparación de ontologías, depuración de la ontología y un razonamiento incremental. Este razonamiento incremental viene íntimamente ligado a la lógica descriptiva que caracteriza a los lenguajes ontológicos. Esta lógica consiste en conceptos (o clases), propiedades o roles (relaciones entre conceptos) e individuos (elementos del dominio); la base del conocimiento consta de dos partes: TBox (términos) que es la descripción de los conceptos y ABox (aserciones) que es la descripción de los individuos. Finalmente al tener toda esta base del conocimiento basado en la lógica descriptiva podemos entrar a razonar sobre la ontología para conocer la satisficibilidad de conceptos, su instanciación, recuperación de información y finalmente definir nuevas reglas para el interés específico [12].

Un ejemplo claro del para que existen todas estas herramientas que nos permiten manejar los tesauros y las bases del conocimientos que estos pretenden modelar es el tesoro sobre el Desplazamiento Forzado en Colombia, tesoro que surgió tras ser un fenómeno del que se habla deliberadamente pero que no alcanza ser plenamente clarificado ni explicado y no posee un enfoque teórico conceptual que permita su interpretación. Además del problema de la ambigüedad, se plantea también el de la dispersión, invisibilización y dificultad de acceso a la información lo que conlleva a una inversión de tiempo considerable en el rastreo bibliográfico al momento de iniciar una investigación, así como también la poca difusión de los estudios o informes de investigación de universidad o centros de investigación. Para resolver estos problemas se planteo todo un proyecto de bibliotecología de los estudiantes de la Universidad de Antioquia para realizar un tesoro e implementarlo en Web para el acceso de la comunidad en general.

## IV. RESULTADOS: SISTEMA SISDEON

En el desarrollo de la aplicación se hace uso de varios componentes necesarios para la correcta manipulación y posterior extracción de conocimiento de la ontología, que en principio es la base del software a desarrollar.



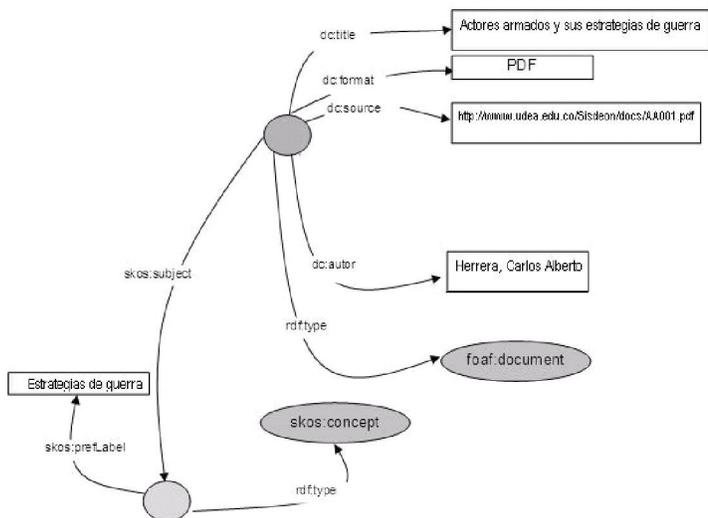


Figura 4. Representación de un documento en Sisdeon.

**B. Recuperación de Documentos**

En el sistema Sisdeon el usuario tiene la posibilidad de realizar consultas sencillas en las que simplemente se busca por un termino determinado, con lo que el sistema con la ayuda de las etiquetas de cada concepto, preflabe, altlabels y hiddenlabels, busca el concepto que mas se acomode a la consulta.

Una segunda y un poco mas compleja estrategia de búsqueda es la que permite múltiples parámetros, es decir, permite que el usuario escoja entre un conjunto de términos, los que le interesa tener en la búsqueda y así realizar consultas por documentos que hagan referencia a varios términos a la vez (búsqueda tipo AND) y por documentos que puedan o no hacer referencia a los diferentes conceptos buscados (búsqueda tipo OR).

Es de vital importancia mencionar que las consultas y las búsquedas son realizadas con la ayuda de SPARQL y Pellet respectivamente. SPARQL utilizado para formular la consulta y pellet para razonar sobre la ontología y encontrar los documentos que mejor se acomoden a las condiciones establecidas.

**C. Visualización**

Para mostrar al usuario a través de la web la ontología, se ha escogido JSP(JavaServer Pages), por ser una tecnología del servidor eficaz y una base excelente para desarrollar aplicaciones Web, ya que este se complementa con Java para implementar la lógica compleja que requiere el manejo de bases de datos. Gracias a los servlets que son una estructura que sirve a JSP [21], podemos hacer páginas dinámicas, además que las paginas JSP actúan como interfaces de aplicaciones que también procesan el almacenamiento de datos, estas junto con Java y su tecnología JDBC para el manejo de sistemas de administración de bases de datos son una buena alternativa para que el cliente pueda visualizar los razonamientos hechos por el pellet. Para tener una idea mas clara de la presentación que se tiene al momento de consultar por un termino específico

se presenta en la figura 5 el termino «estrategias de guerra».

Finalizando de esta manera el proceso, para, en ultima instancia presentar al usuario lo requerido de una forma agradable y fácil de utilizar, al mismo modo que permite una cómoda navegación a través de toda la aplicación Web.

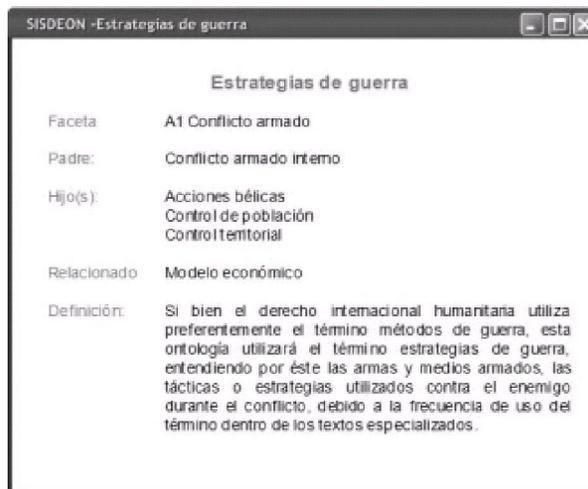


Figura 5. Visualización del termino «Estrategias de guerra».

**D. Base del conocimiento**

La ontología sobre el Desplazamiento Forzado en Colombia se encuentra dividida en facetas

- Faceta A: causas
  - Conflicto armado
  - Modelo económico
  - Actores
- Faceta B: características
  - Población
  - Trayectorias
  - Tipologías
- Faceta C: consecuencias
  - Crisis humanitaria
  - Desintegración y rearticulación
  - Impactos socioeconómicos
  - Recomposición del territorio
- Faceta D: atención a la población
  - Prevención
  - Asistencia humanitaria
  - Retorno
  - Estabilización y consolidación socioeconómica
- Faceta E: metodologías de investigación
  - Cuantitativa
  - Cualitativa

Donde cada término posee la siguiente estructura:

Término	Violencia política
Faceta	Causas
Definición	Se entenderá por violencia política aquella ejercida como medio de lucha político-social, ya sea con el fin de mantener, de modificar, de sustituir o de destruir un modelo de Estado o sociedad; o también de destruir o reprimir a un grupo humano con identidad dentro de la sociedad por su afinidad social, política, gremial, étnica, racial, religiosa, cultural o ideológica, esté o no organizado

T amplio	Estrategias de guerra
T específico	Amenazas

T relacionados

- Asesinato político
- Masacre
- Victimas de la violencia política
- Actores armados
- Crímenes de guerra
- Derecho internacional Humanitario
- Población civil
- Acciones bélicas
- Justicia privada

## E. Otros trabajos

Existen tesauros que abordan el desplazamiento desde distintos enfoques: El Tesoro de Derechos Económicos, Sociales y Culturales, elaborado por el Programa de Ciencias y Derechos Humanos de la Asociación Americana para el Avance de las Ciencias y el Sistema Internacional de Información y Documentación sobre Derechos Humanos, es un documento de referencia sobre los derechos económicos, sociales y culturales y analiza el tema del desplazamiento como una violación a alguno de estos derechos [17]. Igual sucede con el Tesoro de Derechos Humanos elaborado por Ivanna Caccia, Classification Scheme for Human Rights Documentation [18].

Una perspectiva un poco más cercana al tema del desplazamiento en Colombia sería la del tesoro de UNHCR, Internacional Thesaurus of Refugee Terminology [19], elaborado inicialmente por Piers Campbell en 1986 para la Oficina de Refugiados de las Naciones Unidas, posteriores a las versiones desarrolladas a través del Centro de Documentación en Refugiados que incluye actualmente una edición electrónica en varios idiomas. No obstante la cercanía temática, el tesoro sobre refugiados analiza el fenómeno del desplazamiento en relación con la migración interna causada por la implementación de modelos de desarrollo.

Otra aplicación basada en la Web semántica es el SABIOS que propone la introducción de las tecnologías emergentes basadas en semántica, combinadas con sistemas multiagentes y técnicas de recuperación de información para permitir mejorar los procesos de inserción, catalogación y recuperación de documentos digitales, implementando para ello un sistema compuesto por tres módulos: un módulo de conocimiento, un módulo de búsqueda semántica y finalmente un módulo de visualización y navegación de resultados [20].

## V. CONCLUSIONES Y TRABAJOS FUTUROS

Al abordar un sistema de información mediante el uso de ontologías, se ha desarrollado un marco teórico completo de las herramientas y modelos que permiten el manejo de documentos a lo largo de la Web semántica, pretendiendo mostrar no solo las herramientas más avanzadas sobre el tema, sino haciendo notar la evolución que estas han tenido a través del tiempo y viendo así su potencial adquirido.

Se pretende no solo crear estructuras dinámicas para el manejo de la información a través de la Web, sino también

una estructura completa que incluya ontologías, bases de datos, conexiones dinámicas, recuperación de información, búsquedas complejas definidas por las necesidades de la base del conocimiento que se pretende modelar, por medio del pellet y la lógica descriptiva, herramientas muy nuevas y poco exploradas pero de gran ayuda para el modelado de tesauros con amplia dispersión de información como lo es el tesoro del Desplazamiento Forzado en Colombia.

## REFERENCIAS

- [1] Lamarca, M. J. El Nuevo concepto de documento en la cultura de la imagen. [En línea] Disponible en Internet: <http://www.hipertexto.info/documentos/tesauros.htm> [Consulta: marzo de 2007]
- [2] Pérez, J. R. Automatización de tesauros y su utilización en la Web semántica. [En línea] Disponible en Internet: <http://www.w3.org/2001/sw/Europe/events/200406-esp/trabajo-final-extratesauros/trabajo-final-extratesauros.html> [Consulta: marzo de 2007].
- [3] Lamarca, M. J. El Nuevo concepto de documento en la cultura de la imagen. [En línea] Disponible en Internet: <http://www.hipertexto.info/documentos/ontologias.htm> [Consulta: mayo de 2007]
- [4] OWL Web Ontology Language Guide. [En línea] Disponible en Internet: <http://www.w3.org/TR/owl-guide/> [Consulta: mayo de 2007]
- [5] Herramienta de creación de ontologías OntoEdit. [En línea] Disponible en Internet: [http://personales.upv.es/ccarrasc/doc/2003-2004/OntoEdit/presentacion\\_html/definiciononto.html](http://personales.upv.es/ccarrasc/doc/2003-2004/OntoEdit/presentacion_html/definiciononto.html) [Consulta: mayo de 2007]
- [6] Wondertools. [En línea] Disponible en Internet: <http://hcs.science.uva.nl/wondertools/>
- [7] Recuperación y organización de la información. Ontologías de metadatos: función y mapeado [En línea] Disponible en Internet: <http://ontologias-metadatos.50webs.com/edicion-depuracion.html> [Consulta: mayo de 2007]
- [8] The RDF Query Language. [En línea] Disponible en Internet: <http://139.91.183.30:9090/RDF/RQL/> [Consulta: mayo de 2007]
- [9] RDQL a Query Language for RDF. [En línea] Disponible en Internet: <http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/> [Consulta: mayo de 2007]
- [10] SPARQL Query Language for RDF. [En línea] Disponible en Internet: <http://www.w3.org/TR/rdf-sparql-query/> [Consulta: mayo de 2007]
- [11] Pellet [En línea] Disponible en Internet: <http://pellet.owldl.com/> [Consulta: junio de 2007]
- [12] Tecnologías XML y Web Semántica. [En línea] Disponible en Internet: <http://www.di.uniovi.es/~labra/cursos/ver04/pres/XML1.pdf> [Consulta: junio de 2007]
- [13] OWL Web Ontology Language Overview [En línea] Disponible en Internet: <http://www.w3.org/TR/owl-features/> [Consulta: julio de 2007]
- [14] SKOS Core Guide. [En línea] Disponible en Internet: <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20050510/> [Consulta: julio de 2007]
- [15] Jena – A Semantic Web Framework for Java. [En línea] Disponible en Internet: <http://jena.sourceforge.net/> [Consulta: abril de 2007]
- [16] JavaServer Pages Technology. [En línea] Disponible en Internet: <http://java.sun.com/products/jsp/index.jsp> [Consulta: septiembre de 2007]
- [17] Tesoro de Derechos Económicos, Sociales y Culturales. [En línea] Disponible en Internet: <http://ip.aaas.org/sthesaurus.nsf> [Consulta: octubre de 2007]
- [18] Classification Scheme for Human Rights Documentation. [En línea] Disponible en Internet: <http://www.huridocs.org/clasengl.htm> [Consulta: octubre de 2007]
- [19] Thesaurus of Refugee Terminology. [En línea] Disponible en Internet: <http://www.refugeethesaurus.org> [Consulta: octubre de 2007]
- [20] Jaramillo, O., Moncada, J. D.. *La biblioteca pública y las Tecnologías e la Información y las comunicaciones*. pp. 51-71
- [21] Hanna, P., Manual de referencia JSP. McGraw-Hill. 2002
- [22] The Friend of a Friend (FOAF) project [En línea] Disponible en Internet: <http://www.foaf-project.org/> [Consulta: enero de 2008]
- [23] The Dublin Core Metadata Initiative [En línea] Disponible en Internet: <http://dublincore.org/> [Consulta: enero de 2008]