

APLICABILIDAD DE ONTOLOGÍAS EN BIBLIOTECAS DIGITALES

*Rogério Aparecido Sá Ramalho**

Universidad Federal de São Carlos. Brasil.

*Mariângela Spotti Lopes Fujita***

Universidad Estadual Paulista. Brasil.

Resumen: A lo largo de los últimos años, la creciente utilización de las Tecnologías de la Información y Comunicación (TICs) ha impulsado una nueva gama de necesidades informacionales. Los ambientes digitales favorecen la transposición de los límites convencionales de representación y diseminación del conocimiento registrado, incorporando nuevos elementos a los procesos de producción, organización y recuperación de informaciones. Considerando la ontología como una nueva e incitante categoría de sistemas de representación del conocimiento, analizan las aproximaciones de este nuevo abordaje computacional con los instrumentos y métodos tradicionales de organización del conocimiento, a partir de la presentación de la ontología OIR, concebida para la representación y intercambio de informaciones referenciales de documentos académicos, se investiga en qué medida el área de Ciencia de la Información puede contribuir para el desenvolvimiento de ontologías y analiza los reflejos de su utilización en bibliotecas digitales.

Palabras clave: Ontologías; sistemas de organización del conocimiento; bibliotecas digitales; organización de la información; tecnologías semánticas.

Title: APPLICABILITY OF ONTOLOGY IN DIGITAL LIBRARIES.

Abstract: Throughout the last years, the increasing use of Technologies of Information and Communication (TICs) have stimulated a new gamma of informational necessities. The digital environments favor the transposition of the conventional limits of representation and dissemination of the registered knowledge, incorporating new elements to the processes of production, organization and recovery of information. Considering the ontologies as a new and riveting category of systems of representation of knowledge, we searched bases that allow to analyze the approaches of this new computational approach with the instruments and methods of Knowledge Organization, starting from presentation of the ontology OIR, designed for the representation and exchange of referential informations of academic papers, investigating in what scale the area of Information Science can contribute to the development of ontologies and analyzing the consequences of its use in digital libraries.

Keywords: Ontologies; systems of knowledge organization; digital libraries; organization of information; semantic technologies.

* ramalho@ufscar.br

** goldstar@flash.tv.br

Recibido: 7/12/2010; 2ª revisión: 24/01/2011; aceptado: 3/02/2011.

1. INTRODUCCIÓN

Tradicionalmente las bibliotecas se caracterizan como el medio más importante de conservar el saber colectivo, representando una especie de cerebro universal que posibilita recuperar lo que olvidamos y lo que aún no sabemos (Eco, 2003, p. 1). Tal afirmación destaca el papel que las bibliotecas siempre ejercieron en la historia de la humanidad, en su constante búsqueda para atender las necesidades informacionales de cada periodo de la historia.

A lo largo de los últimos años, el aumento exponencial del volumen de informaciones y las especificidades inherentes a los contenidos digitales han ocasionado grandes impactos en lo que hay de más concreto y significativo en las bibliotecas: los acervos; antes formados solamente por objetos táctiles y ahora constituidos también por documentos almacenados en formato digital.

Los contenidos informacionales almacenados en bibliotecas digitales se presentan en diversas formas (fotos, películas, *sites*, libros, documentos de archivos, piezas de museo) y se han multiplicado cada día tornando la organización aun más necesaria, pues individualmente nadie es capaz de recordar donde cada fragmento de información puede ser encontrado, y cuáles son sus relaciones con otras partes (Gnoli, 2009).

Considerando la organización de la información como elemento fundamental para la garantía de calidad en la recuperación, se presenta como desafío contemporáneo la necesidad de singularización contextual en la reconstrucción del conocimiento, buscando proporcionar la recuperación y el uso de las informaciones a partir de la identificación, requisitos de pertinencia y relevancia en contextos específicos.

Con el desenvolvimiento de las Tecnologías de la Información y Comunicación (TIC) nuevas demandas informacionales son impuestas a las bibliotecas, pues lo que podría ser realizado por medio de tarjetas de papel en bibliotecas del siglo XIX, no es suficiente para atender a las actuales necesidades informacionales de los usuarios contemporáneos. (Dabrowski; Synak y Kruk, 2009). En este sentido, es necesario repensar los modelos clásicos de organización del conocimiento sobre diferentes perspectivas, de modo que las ontologías han sido apuntadas como un abordaje computacional capaz de engendrar nuevos tipos de servicios y productos informacionales, favoreciendo mejoras en los procesos de representación, organización y recuperación de informaciones en bibliotecas digitales.

En la literatura del área de Ciencia de la Información el término ontología comenzó a ser utilizado a partir del final de la década de 1990, cuando los instrumentos de representación y clasificación pasaron a despertar un mayor interés de investigadores de las comunidades de Ciencia de la Computación y Lingüística (Vickery, 1997; Soergel, 1999). Desde entonces, el desenvolvimiento y utilización de ontología ha figurado como objeto de investigación en las más variadas áreas del conocimiento. No obstante, todavía es posible identificar una divergencia entre la producción literaria y el número de ontologías existentes. (Sánchez Cuadrado *et al.*, 2007).

De esta manera, el presente trabajo tiene como objetivo demostrar la aplicabilidad de ontologías en bibliotecas digitales, contextualizando sus componentes de acuerdo con el enfoque del área de Ciencia de la Información, posibilitando un mayor esclarecimiento terminológico, y destacando los desafíos y beneficios de tal abordaje en el ámbito de

desenvolvimiento de bibliotecas digitales a partir de la presentación de la ontología OIR, concebida para la representación y intercambio de informaciones referenciales de documentos académicos.

2. REPRESENTACIÓN DEL CONOCIMIENTO Y ONTOLOGÍAS

El área de Ciencia de la Información posee tradición en el desenvolvimiento y uso de instrumentos de representación del conocimiento, conforme puede ser constatado en las prácticas profesionales identificadas en su campo de actuación y en su *corpus* teórico, caracterizándose como un área de naturaleza mediadora, ya que propicia la interlocución entre los contextos de producción y de uso de la información, en especial en aquello que corresponde a la dimensión de los contenidos informacionales, en la mayoría de las veces denominada como tratamiento temático de la información (Guimarães; Sales, 2010).

El tratamiento temático de la información tiene como foco cuestiones referente al análisis, descripción y representación del contenido de los documentos, así como sus interfaces con las teorías y sistemas de almacenamiento y recuperación de la información (Barité, 1997). Históricamente, ese campo de estudios se presenta a partir de tres vertientes teóricas: la catalogación de asunto (*subject cataloguing*) de matriz norteamericana, la indexación (*indexing*) de matriz inglesa y el análisis documental (*analyse documentaire*) de matriz francesa (Guimarães, 2009).

Siendo así, considerando el análisis documental como una de las corrientes teóricas del tratamiento temático de la información, se verifica la necesidad de instrumentos que posibiliten representar formalmente los relacionamientos existentes entre los contenidos informacionales.

En este contexto, las ontologías figuran como un nuevo abordaje computacional de representación y organización de contenidos informacionales, favoreciendo la aproximación de los modelos clásicos de representación, utilizados tradicionalmente en el ámbito del área de Ciencia de la Información, con las nuevas tecnologías semánticas desarrolladas en el ámbito del ambiente Web.

Para García Marco (2007) la investigación sobre ontologías y el movimiento de convergencia con los instrumentos convencionales de representación renuevan la importancia de las herramientas de control conceptual y terminológico, situando el área de Organización del Conocimiento en una posición privilegiada, a partir de la aproximación de un conjunto de ciencias que vienen, históricamente, manteniendo una relación estrecha. El autor destaca también que uno de los principales peligros de tal aproximación es el de confundir diferentes instrumentos de organización conceptual y terminológica, haciendo sinónimos conceptos próximos, pero que son claramente diferentes.

En el ámbito del área de Ciencia de la Información una ontología puede ser definida como un sistema de representación del conocimiento que posibilita describir formalmente las propiedades y relacionamientos de un determinado modelo conceptual, favoreciendo la realización de inferencias automáticas en los procesos de organización y recuperación de recursos informacionales (Ramalho, 2010).

De acuerdo con los moldes convencionales, la representación de un documento consiste en describir sus puntos de acceso a partir de la catalogación y realizar la

indexación de su contenido, identificando los asuntos abordados y traduciéndolos por medio de un lenguaje documentario.

A partir de tales procedimientos se torna posible la identificación de forma única de los documentos y la realización de búsquedas por asunto. Pero, ¿Qué tal consultar la experiencia profesional del autor? O ¿Verificar detalles sobre su formación académica?, ¿Cuánto facilitaría recuperar de modo automático la dirección electrónica de todos los eventos sobre determinada temática realizados en los últimos años?, son en estas perspectivas que residen las potencialidades del uso de ontologías.

El punto principal es que bien construida una ontología puede representar una visión de mundo, potencializando representaciones semánticas que no podrían ser obtenidas a partir de descripciones textuales. Las representaciones semánticas pueden ofrecer informaciones sobre los objetos en la manera como las personas tienden a pensar. Nosotros usamos asociaciones libres, agrupamiento de objetos por semejanzas y permitiendo muchos tipos de contextos, diferenciando la importancia de los factores de acuerdo con la situación. Una foto aérea va a significar algo diferente para un turista que para un militar (Dabrowski; Synak y Kruk, 2009).

En esta perspectiva, el estudio de ontología ha despertado el interés principalmente de investigadores relacionados al área de Organización del Conocimiento (OC), el cual se ha consolidado como importante campo de investigación para el área de Ciencia de la Información, debido a sus diversidades conceptuales teóricas y metodológicas (Fujita, 2008).

Las ontologías se presentan como una nueva categoría de instrumentos de representación, de tal modo que a pesar de las nuevas potencialidades ofrecidas, poseen aspectos comunes a los modelos y métodos empleados tradicionalmente en el ámbito del área de Ciencia de la Información. Siendo así, se puede describir sucintamente los componentes de una ontología:

- Clases y subclases: las clases y subclases de una ontología agrupan un conjunto de elementos, “cosas”, del “mundo real”, que son representadas y categorizadas de acuerdo con sus similitudes, llevándose en consideración un dominio concreto. Los elementos pueden representar cosas físicas o conceptuales, desde objetos inanimados hasta teorías científicas o corrientes teóricas.
- Propiedades Descriptivas: describen las características, adjetivos y/o cualidades de las clases.
- Propiedades Relacionales: se trata de las relaciones entre clases pertenecientes o no a una misma jerarquía, describiendo y rotulando los tipos de relaciones existentes en el dominio representado.
- Reglas y axiomas: enunciados lógicos que posibilitan imponer condiciones como tipos de valores aceptados, describiendo formalmente las reglas de la ontología y posibilitando la realización de inferencias automáticas a partir de informaciones que no necesariamente fueran explícitas en el dominio, pero que pueden estar implícitas en la estructura de la ontología.
- Instancias: Indican los valores de las clases y subclases, constituyendo una representación de objetos o individuos pertenecientes al dominio modelado, de

acuerdo con las características de las clases, relacionamientos y restricciones definidas.

- Valores: atribuyen valores concretos a las propiedades descriptivas, indicando los formatos y tipos de valores aceptados en cada clase.

En esta perspectiva, se puede considerar como componentes estructurales de las ontologías a las Clases/Subclases; Propiedades Descriptivas; Propiedades Relacionales y Reglas/Axiomas. Tales componentes forman una “espinas dorsal” de la ontología, permitiendo la representación formal de los elementos y relacionamientos del dominio modelado, así mismo las Instancias y Valores posibilitan la atribución de informaciones concretas a las Clases y Propiedades Descriptivas, respectivamente.

Las ontologías potencializan la representación de los relacionamientos existentes entre los conceptos que “traducen” los contenidos informacionales, caracterizándose como un tipo de fuente secundaria, instituyendo nuevas formas de acceso y favoreciendo mayor flexibilidad y calidad en la recuperación de informaciones, por medio de la realización de inferencias.

De este modo, se verifica que el área de Ciencia de la Información puede contribuir para el desenvolvimiento de ontologías a partir del fornecimiento de substratos teóricos y metodológicos, posibilitando el aprovechamiento de instrumentos y métodos ya sedimentados en su campo de actuación.

En contrapartida el desenvolvimiento de ontologías en el ámbito del área de Ciencia de la Información incorpora nuevos subsidios computacionales para la representación de las relaciones entre conceptos, contribuyendo para mejorar las actividades de representación, clasificación e indexación.

A lo largo de los últimos años, innumerables metodologías han sido propuestas en el área de Ciencia de la Computación para el desenvolvimiento de ontologías, pudiéndose destacar la “metodología inicial” presentada por Uschold y King (1995); METHONTOLOGY propuesta por Fernández y colaboradores (1997); la metodología desarrollada a partir del proyecto *Toronto Virtual Enterprise* (Uschold y Gruninger, 1996); y el método 101 propuesto por Noy y McGuinness (2001).

Entre las diversas teorías y normas presentes en la literatura del área de Ciencia de la Información que pueden ser aprovechadas en el desenvolvimiento de ontologías, de acuerdo con el enfoque de este trabajo, se destaca la teoría del concepto y la teoría de la clasificación facetada utilizadas para la elaboración de tesauros (Campos y Gomes, 2006; Gomes *et al.*, 1990) y la norma ANSI/NISO Z39.19-2005 que presenta directrices para la elaboración de vocabularios controlados.

Cuando son comparados los abordajes es posible identificar los diferentes enfoques característicos en cada una de las áreas del conocimiento. En cuanto el área de Ciencia de la Computación tiene como principal objetivo el desenvolvimiento de artefactos de software, presentando un ciclo de actividades fundamentales en los principios de Ingeniería de Software, la Ciencia de la Información se enfoca en la definición de recomendaciones y procedimientos para la elaboración de vocabularios controlados, enfatizando aspectos terminológicos, buscando favorecer la organización y uso de las informaciones.

Basándose en la concepción de la unidad de conocimiento sobre las fronteras de las Ciencias, se torna deseable el desenvolvimiento de abordajes que favorezcan la aproximación entre áreas, contribuyendo para la evolución del conocimiento humano de forma integral.

En esta perspectiva, se buscó sistematizar un método para el desenvolvimiento de ontologías con el objetivo de facilitar la comprensión de las etapas y procesos necesarios, minimizando detalles técnicos relacionados a la especificación como artefacto de software y, valorizando los puntos de convergencia con los demás instrumentos de representación tradicionalmente concebidos en el ámbito del área de Ciencia de la Información, conforme descrito en las etapas a seguir:

- Identificación del dominio y alcance de la ontología: en esta etapa se busca identificar claramente el propósito de la ontología y las potencialidades ofrecidas a partir de su utilización, delineando el dominio pretendido y niveles de representación.
- Verificación de posibilidades de reutilización: analizar posibles ontologías y demás instrumentos de representación que puedan ser reutilizados o servir como substratos teóricos para el modelado de la ontología, potencializando aspectos de interoperabilidad.
- Levantamiento de términos y conceptos: identificar la terminología utilizada para la representación de los componentes en la ontología, evitando problemas con polisemia y redundancias.
- Definición de los componentes estructurales: definir los componentes estructurales de la ontología, categorizando la jerarquía de clases e identificando las respectivas propiedades descriptivas y relacionales.
- Formalización de axiomas y reglas: describir formalmente los axiomas y reglas de la ontología, posibilitando el procesamiento de las restricciones y la realización de inferencias de forma automática.
- Implementación: definir las tecnologías y lenguajes computacionales que serán utilizadas para implementar y dar soporte al desenvolvimiento de la ontología.
- Evaluación y documentación: hacer juzgamientos de cada una de las etapas del modelaje y de los resultados potenciales de la utilización de la ontología, de acuerdo con los objetivos propuestos, y documentar todos los componentes y tareas desarrolladas. En esta etapa los criterios del OntoClean¹ pueden auxiliar en la evaluación de los componentes estructurales.

De este modo, el desenvolvimiento de ontologías se presenta como un proceso no lineal, siendo necesarios refinamientos e interacciones sucesivas a lo largo del proceso, hasta la obtención del modelo deseado y así, como los instrumentos de representación convencionales, las ontologías también deben poseer una política de actualización permanente.

Conforme destacan Hu y Zhao (2007), la utilización de ontologías para describir conceptos, basándose en su nivel semántico, ha sido presentada como un abordaje promisor en el campo de investigación de bibliotecas digitales.

3. UTILIZACIÓN DE ONTOLOGÍAS EN BIBLIOTECAS DIGITALES

A lo largo de los últimos años, el término biblioteca digital ha sido utilizado indistintamente para designar una temática que engloba actividades de investigación muy variadas, fundamentadas a partir de diferentes concepciones y abordajes.

Hace más de dos décadas Swatson (1988) ya destacaba que las tecnologías digitales posibilitan nuevos horizontes para la creatividad humana explorar el mundo de conocimientos registrados, ofreciendo oportunidades para pesquisas de investigación innovadoras en materia de recuperación de la información.

Conforme afirma García Marco (2007) el ambiente Web se ha convertido en el nuevo ambiente de distribución, almacenamiento, publicación y acceso a la información. Fue a partir del desenvolvimiento de la Web que nuevas demandas informacionales comenzaron a ser identificadas, de modo que las pesquisas relacionadas a la Web Semántica concedieron notoriedad al desenvolvimiento de ontologías.

Greenberg (2007) destaca que existen fuertes relaciones entre la Web Semántica y las bibliotecas digitales, pues ambas poseen características comunes: son desarrolladas, en parte, como respuesta para la abundancia de información; poseen misiones fundamentadas en el servicio, acceso a la información y descubierta de conocimiento; y se han desarrollado padrones y tecnologías colaborativas.

El desenvolvimiento de herramientas que beneficien el establecimiento de relaciones semánticas entre diferentes partes de un *corpus* documental, favorece que las bibliotecas digitales puedan inferir informaciones acerca de preguntas complejas, fundamentales para el área académica, tales como las presentadas por Shum y colaboradores (2007):

- ¿Qué publicaciones sustentan y contestan este documento?
- ¿Cuál es la corriente intelectual de esta idea?
- ¿Qué datos hay para sustentar esta reivindicación específica o previsión?
- ¿Quién más está trabajando sobre este problema?
- ¿Este abordaje fue utilizado en otros campos? ¿Qué conexiones lógicas o analógicas fueron realizadas entre esas ideas?

Tales preguntas requieren evidentemente un trabajo interpretativo complejo, y además de eso, puede haber divergencias de diversos tipos. Para responderlas se torna necesario el uso de anotaciones semánticas en un nivel diferente del abordado por metadatos convencionales (Shum *et al.*, 2007).

Tradicionalmente la unidad primaria de información consiste, por ejemplo, en un libro; en la actualidad, la representación del contenido alcanza detalles antes inimaginables, pudiendo ser un mapa, una figura, una diapositiva, un capítulo o una entrada de alguna obra de referencia, y de acuerdo con la política de indexación adoptada, demandando nuevos instrumentos de representación y organización (Cunha, 2000).

Conforme destacan Dabrowski, Synak y Kruk (2009) en el escenario contemporáneo la utilización de palabra-clave no se presenta como un abordaje suficientemente significativo para describir los contenidos informacionales, siendo necesario el uso de relaciones semánticas que posibiliten representar la manera como las personas tienden a considerar tales contenidos.

Weinstein y Birmingham (1998) afirman que el uso de ontologías en bibliotecas digitales favorece los siguientes beneficios:

- Consultas más precisas: como todos los conceptos de una ontología pueden ser utilizados como puntos de acceso para consultas, posibilitan expresiones de búsqueda más adecuadas y precisas.
- Explicitar búsquedas imprecisas para consultas precisas: en ontologías, los atributos son definidos por sus relaciones con otros atributos, ofreciendo numerosas secuencias intuitivamente naturales para auxiliar los usuarios en la articulación de sus necesidades y refinamiento de las consultas, tornándolas cada vez más específicas.
- Integración de catálogos: el mismo tipo de raciocinio usado para agrupar descripciones bibliográficas puede ser aplicado para integrar múltiples catálogos.
- Cálculo de las necesidades de licencia por servicio: los relacionamientos de la ontología pueden proporcionar un lenguaje para expresar contratos de licencia, así como también medios para computar tasas de servicio en tiempo de ejecución.

De esta manera, se observa que las ontologías posibilitan ir más allá de la representación de los aspectos descriptivos y temáticos de los documentos, ofreciendo subsidios computacionales para la representación de los propios dominios, contribuyendo para la contextualización de las informaciones y apuntando nuevas posibilidades en el actuar profesional del área de Ciencia de la Información.

En esta perspectiva, innumerables autores han presentado las potencialidades de la utilización de ontologías en bibliotecas digitales. Deng y Tang (2002) hacen uso de las ontologías y la tecnología de agentes para proponer un nuevo modelo de descubrimiento del conocimiento, utilizando ontologías para dos funciones principales: favorecer una mejor clasificación y mejorar la renovación y tasa de precisión de recuperación por medio de búsquedas semánticas.

Li y Sun (2003) proponen un nuevo modelo de servicio de biblioteca digital, que integra una serie de ontologías, desempeñando un papel central en la comprensión semántica de los recursos y servicios digitales.

En medio del creciente número de proyectos relacionados al desenvolvimiento de bibliotecas digitales basados en ontologías, y de acuerdo con el ámbito de aplicación de esta investigación, se puede destacar también los proyectos JeromeDL y ScholOnto.

El proyecto JeromeDL² consiste en una biblioteca digital de código abierto que posibilita la descripción de recursos informacionales a partir del lenguaje computacional RDF y la realización de búsquedas semánticas basadas en ontologías (Kruk; Decker y Zieborak, 2005).

ScholOnto es un servidor de biblioteca digital que a partir de la aplicación de ontologías fornece una plataforma semántica para la discusión e interpretación del discurso académico, focalizando alternativas de interface y colaboración. (Shum; Motta y Domingue, 2000).

En el ámbito de esta investigación, buscando evaluar la eficacia del método propuesto para el desenvolvimiento de ontologías basado en los principios del área de Ciencia de la Información y comprobar la aplicabilidad de ontologías en bibliotecas digitales, fue

elaborada una ontología para la representación e intercambio de informaciones referenciales de documentos académicos, la cual fue denominada como OIR.

4. OIR: ONTOLOGÍA PARA INFORMACIONES REFERENCIALES

Construir una ontología compatible con las normas ya existentes y hábitos de la comunidad, al mismo tiempo que posibilite disfrutar de las capacidades ofrecidas por las tecnologías semánticas, es el primer paso para que en el futuro las bibliotecas digitales puedan cambiar nuestra forma de investigación, acceso y utilización de los recursos. (Dabrowski; Synak y Kruk, 2009).

Teniendo como base el método presentado anteriormente, se desarrolló la ontología OIR, con el objetivo de favorecer una mejor representación y intercambio de informaciones referenciales de documentos académicos, por medio de la descripción formal de los relacionamientos semánticos inherentes a los recursos informacionales.

Así, después de la definición del propósito y alcance de la ontología, fue realizado un estudio con la finalidad de identificar instrumentos y métodos de organización y representación de informaciones, tradicionalmente utilizados en el ámbito del área de Ciencia de la Información, que pudieran servir de substratos teóricos para el modelado de la OIR.

Para la definición de las propiedades descriptivas de la ontología cuestiones de interoperabilidad de vocabularios fueron considerados como un factor preponderante, buscándose fundamentación en los principales formatos de descripción bibliográfica (MARC; Dublin Core y ETD-MS) y en el padrón Brasileiro de Metadatos para Tesis y Disertaciones (MTD-BR), posibilitando la compatibilización de registros, frente a la creciente necesidad de coparticipación de recursos informacionales.

Además del padrón MTD-BR, con la finalidad de potencializar la integración de las informaciones disponibles en OIR con otros repositorios brasileiros, también fue utilizado en la definición de los metadatos el padrón de la plataforma Lattes³, favoreciendo la compatibilización y el relacionamiento con informaciones académicas de investigadores y programas de postgrado.

En la perspectiva de reutilización de ontologías fue utilizada la ontología *Friend Of A Friend*⁴ (FOAF), difundida en el ambiente Web para describir informaciones sobre redes sociales, posibilitando representar redes de colaboración.

El lo que respecta al levantamiento terminológico para la definición de los términos, esta etapa fue favorecida a partir del aprovechamiento de padrones ya sedimentados, optándose por la reutilización de términos ya empleados en los padrones y formatos convencionales de representación.

Para la inclusión de nuevos términos, de acuerdo con los objetivos propuestos para la ontología, se buscó evitar expresiones técnicas, contextualizándolas a partir de un vocabulario adecuado en el contexto brasileño.

En relación a la nominación de los componentes de la ontología, con el objetivo de facilitar la comprensión y evitar posibles problemas de interoperabilidad, fue seguida la convención de no utilizar acentuación, espacios y caracteres especiales. Se definió también, la utilización de letras mayúsculas en el inicio y en todas las palabras

subsecuentes contenidas en el nombre de las clases y subclases, notación comúnmente denominada como *CamelBack*.

El mismo procedimiento fue adoptado para representar las propiedades relacionadas de la ontología, con excepción del uso de letras minúsculas en el inicio de sus nombres, diferenciándolas de las clases. En cuanto a las propiedades descriptivas, con la intención de facilitar la asimilación de los componentes de la ontología, se utilizó como padrón la inserción del nombre de la clase en el inicio del nombre propiedad, seguida por el símbolo de *underline* “_” y el nombre de la propiedad descriptiva, iniciando en minúscula.

La etapa siguiente se define por la estructuración de la jerarquía de clases e identificación de las respectivas propiedades. De esta manera, teniendo como base los principios clasificatorios utilizados en la construcción de tesauros y taxonomías fueron especificados los relacionamientos jerárquicos existentes entre clases, así como los otros tipos de relacionamientos (propiedades relacionales), y propiedades descriptivas de cada una de las clases existentes en el dominio.

Conforme se destacó anteriormente, para que todo el potencial de las ontologías pueda ser explorado, se torna necesaria su formalización a partir de lenguajes que puedan ser procesadas por máquinas. De este modo, después de las definiciones conceptuales de las etapas anteriores, se realizó la implementación de la ontología y formalización de las reglas y axiomas, con el objetivo de posibilitar la realización de inferencias automáticas.

De acuerdo con los objetivos de la ontología propuesta se optó por el lenguaje computacional OWL para su implementación, por ser el lenguaje recomendado por el *World Wide Web Consortium*⁵ (W3C) para el desarrollo de ontologías y debido a su amplia aceptación. La especificación completa del lenguaje OWL se presenta en McGuinness e Harmelen (2004).

Como herramienta de apoyo para el modelado y formalización de las reglas y axiomas de la ontología se utilizó la herramienta Protégé⁶, un software de código abierto Java, desarrollado inicialmente por el departamento de Informática Médica de la Universidad de Stanford, que fornece una arquitectura extensible para edición de ontologías.

Después de la implementación, fue realizada la evaluación y documentación de la ontología. Para documentar OIR se utilizó recursos propios de la herramienta Protégé, a partir de la inserción de anotaciones en sus componentes. En lo concerniente a los procedimientos de evaluación, estos pueden ser divididos en dos categorías:

1. Evaluación de la jerarquía de clases y componentes de la ontología: realizada a partir de los criterios de la metodología OntoClean, adecuando ontológicamente los componentes estructurales de la ontología hasta la obtención del modelo deseado.
2. Evaluación de las reglas y axiomas: efectuada por medio del juzgamiento del resultado de búsquedas que se aproximen de las necesidades informacionales de usuarios en potencial, simuladas para evaluar la capacidad de recuperación de informaciones de la ontología y realización de inferencias automáticas.

Dicho proceso fue realizado interactivamente y de forma no lineal, por medio de refinamientos sucesivos, para alcanzar el resultado esperado. En esta perspectiva, se presente a seguir algunos trechos de los componentes estructurales que forman la especificación de OIR.

4.1 Clases y Subclases

Las clases y subclases son estructuradas por medio de una taxonomía de conceptos, organizadas a partir de las características esenciales de las instancias representadas en la ontología e interconectadas por relaciones jerárquicas. La relación de subordinación *is-a* (es_un) es la primitiva de estructuración más utilizada en el ámbito del área de Ciencia de la Información y se presenta como una forma bastante útil de representar a la jerarquía de clases de una ontología.

Tal representación posibilita una visión general de la ontología y todas las entidades del dominio modelado, de modo que todas las propiedades existentes en una clase son “heredadas” por sus subclases, las subclases se constituyen como especializaciones de las clases, a partir de la definición de propiedades que no pueden ser aplicadas a todas las instancias de la clase superior, motivo por el cual gráficamente se representa una única flecha direccionada para la clase superior, Figura 1.



Figura 1. Jerarquía de Clases de la OIR.

Conforme puede ser observado, en el plano del lenguaje, las ontologías posibilitan representar relaciones de equivalencia, generalmente utilizadas para la integración de ontologías, representadas por medio de dos flechas de “doble mano” de la relación *is-a*, indicando una relación de equivalencia entre dos clases, y no entre clases y subclases.

4.2 Propiedades Descriptivas

Las propiedades descriptivas son utilizadas para describir las características y/o cualidades de las instancias de la ontología, posibilitando la atribución de valores concretos, y su aprovechamiento se aproxima de los instrumentos y normas de representación descriptiva utilizados en el ámbito del área de Ciencia de la Información.

Para efecto de ejemplificación, el cuadro I presenta las propiedades descriptivas de la clase Documento de la ontología OIR y una sucinta descripción de cada propiedad.

Documento	
Propiedad Descriptiva	Descripción
Area_nome	Nombre del área de conocimiento
Documento_abstract	Abstract del documento
Documento_ano	Año de publicación del documento
Documento_assunto	Tópicos temáticos tratados en el documento
Documento_citacaoAbnt	Citación del documento, de acuerdo con las normas
Documento_dataAtualizacao	Data en que fue realizada la última actualización de
Documento_dataPublicacao	Data de publicación del documento
Documento_descricao	Descripción del documento
Documento_direitos	Informa las condiciones de distribución, reproducción y utilización del documento
Documento_doi	<i>Digital Object Identifier</i> ⁷ (Identificador de Objeto)
Documento_formato	Formato del documento
Documento_idioma	Idioma del documento
Documento_meioDivulgacao	Medio de divulgación del documento
Documento_numPaginas	Número total de páginas del documento
Documento_paginaFinal	Número de la página final del documento
Documento_paginaInicial	Número de la página inicial del documento
Documento_resumo	Resumen del documento
Documento_tipo	Tipo de documento (Tesis; Disertación; ArtículoEnPeriodico; TrabajoEnEvento; Libro; CapituloDelLibro)
Documento_titulo	Título del documento
Documento_uri	Dirección electrónica del documento

Cuadro I. Propiedades Descriptivas de la Clase Documento.

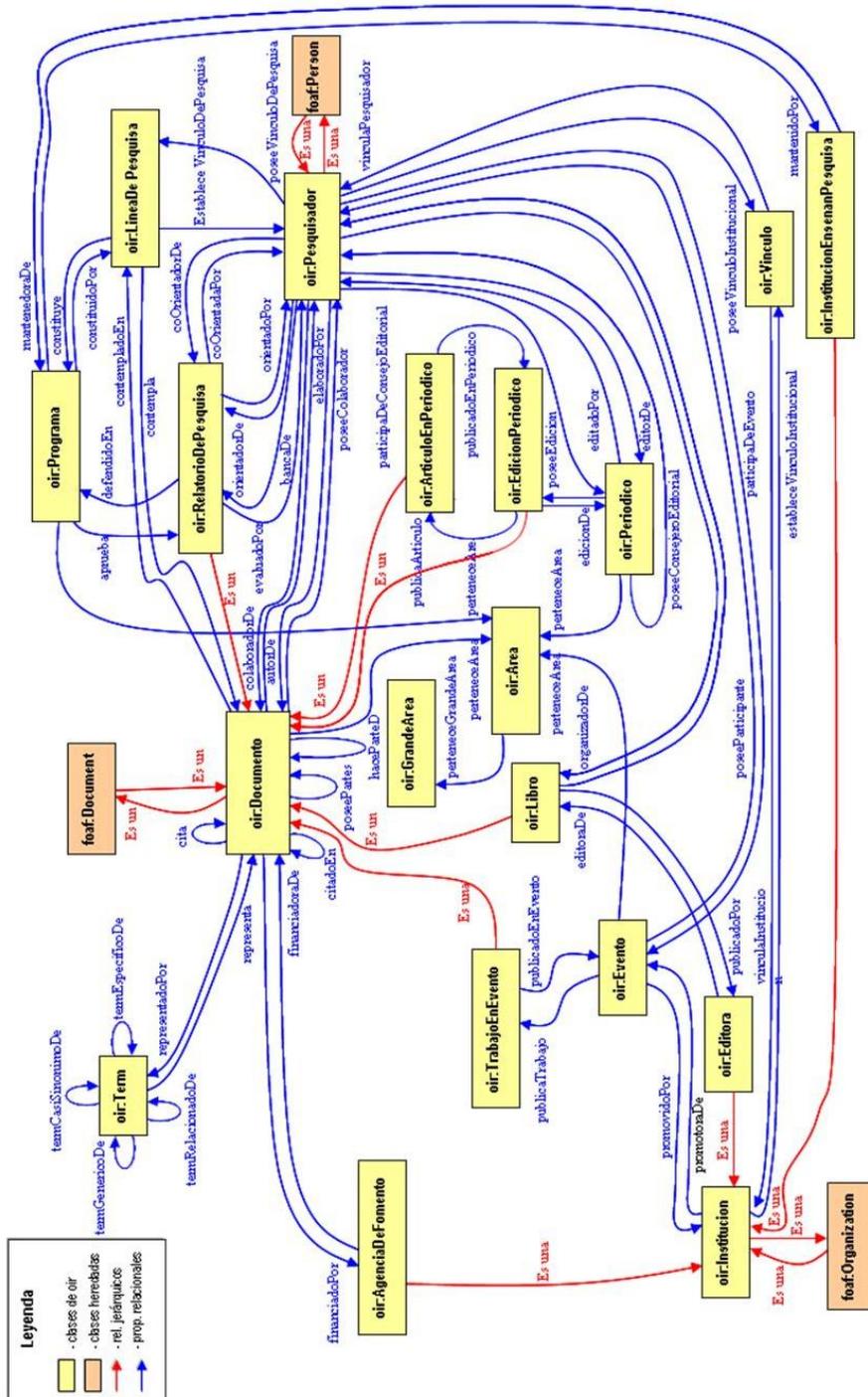


Figura 2: Diagrama de Clases y Propiedades Relacionales de la OIR.

4.3 Propiedades Relacionales

Las Propiedades Relacionales permiten la definición de relacionamientos más avanzados que pueden ser incorporados a la jerarquía de clases, presentándose como la forma más eficiente para la representación de relacionamientos en ontologías.

Tales propiedades se caracterizan como relaciones ónticas y no jerárquicas, posibilitando el establecimiento de relacionamientos entre las instancias de clases de la ontología y mayor libertad de representación.

Las relaciones ónticas pueden ser subdivididas en diversos tipos, pudiendo variar de acuerdo con el contexto (Dahlberg, 1978). De esa forma, se verifica que las propiedades relacionales representan los relacionamientos existentes entre objetos, siendo necesario indicar la clase origen, la clase destino y el rótulo del relacionamiento, el cual normalmente es definido por medio de verbos (ej.: representa), o verbos preposicionados (ej.: publicadoPor).

En esta perspectiva visando posibilitar una mejor comprensión de la ontología OIR e identificación de las potencialidades ofrecidas, se presenta en la figura 2 una representación gráfica de todas sus clases y propiedades relacionadas.

4.4 Reglas y Axiomas

Las reglas y axiomas son enunciados lógicos que posibilitan imponer condiciones como tipos de valores aceptados, favoreciendo la realización de inferencias automáticas. Tales componentes se presentan como la principal característica de las ontologías frente a los demás instrumentos de representación del conocimiento.

Para su definición son utilizadas sentencias descritas en Lógica de Primer Orden (LPO), conocida también como cálculo de predicados de primer orden (CPPO), posibilitando definir formalmente la semántica de los formalismos. Buscando ejemplificar la descripción de reglas y axiomas se presenta en el cuadro II algunas reglas básicas definidas en la ontología OIR.

Identificador	Reglas / Axiomas
1	$\text{ArticuloEnPeriodico}(?x) \rightarrow \text{Documento}(?x)$
2	$\text{InformeDePesquisa}(?x) \wedge \text{EvaluadoPor}(?x, ?y) \rightarrow \text{poseeColaborador}(?x, ?y)$
3	$\text{InformeDePesquisa}(?x) \wedge \text{contempladoEn}(?x, ?y) \wedge \text{elaboradoPor}(?x, ?z) \rightarrow \text{poseeVinculoDePesquisa}(?z, ?y)$

Cuadro II. Ejemplos de Reglas y Axiomas de la OIR.

Conforme puede ser observado, las Clases y Propiedades representan los predicados para la definición de las reglas y Axiomas. La Regla 1 indica que un Artículo publicado en un Periódico también es una instancia de la clase “Documento”, así hubiera sido registrado apenas en la clase “ArticuloEnPeriodico”; la Regla 2 infiere que un investigador que participó como un jurado de Maestría o Doctorado puede ser considerado como colaborador de la versión final del trabajo presentado; la Regla 3 infiere que un

Investigador que sustentó una disertación de Maestría o tesis de doctorado en determinada línea de pesquisa, posee vínculo con esta línea de pesquisa.

Siendo así, a partir de la unión de una serie de reglas elementares, como las demostradas anteriormente, es posible elaborar reglas complejas, por ejemplo, la ontología OIR fue utilizada con éxito para la definición automática de la referencia completa de documentos, identificando nuevos relacionamientos como redes de colaboración entre autores y deduciendo centenas de nuevas informaciones de modo automatizado, posibilitando mejoras significativas en el ámbito de la organización y recuperación de contenidos informacionales.

5. CONCLUSIONES

Consideramos que a partir del estudio realizado fue posible verificar como las ontologías pueden ser aplicadas en el ámbito del desenvolvimiento de bibliotecas digitales. De modo que las ontologías se presentan como una nueva categoría de instrumentos de representación del conocimiento y de este modo, pese a sus diferentes vinculaciones operativas y teóricas, poseen aspectos comunes que las aproximan de los demás modelos de representación utilizados tradicionalmente en el ámbito del área de Ciencia de la Información.

Se constató que a partir de la aproximación de las metodologías de desenvolvimiento de ontologías presentes en la literatura de Ciencia de la Computación, y de los substratos teóricos del área de Ciencia de la Información es posible la elaboración de métodos más adecuados para el desarrollo de ontologías, minimizando detalles técnicos relacionados a la especificación como artefacto de software y, valorizando los puntos de convergencia con los demás instrumentos de representación tradicionalmente concebidos en el ámbito del área de Ciencia de la Información.

A partir del modelo propuesto fue descrito sucintamente los principales componentes de la Ontología de Informaciones Referenciales (OIR), concebidos con el objetivo de posibilitar un modelo conceptual para la representación de informaciones referenciales de documentos académicos, contribuyendo para una mejor organización y recuperación de informaciones, a partir de la descripción formal de los relacionamientos semánticos identificados en el dominio modelado.

En esta perspectiva, se verifica que la utilización de ontologías promete dar abundantes frutos en el ámbito de desenvolvimiento de bibliotecas digitales, favoreciendo la transposición de los métodos convencionales de representación, contribuyendo para la realización de operaciones lógicas de forma automática, como, por ejemplo, localizar un término en diversas fuentes de información, detectar términos genéricos y equivalentes, eliminar ambigüedades, identificar hipónimos, entre otros. Favoreciendo así una representación y contextualización más adecuada de los contenidos documentales.

Se verificó también, que a partir de la utilización de ontologías los profesionales de información ven multiplicadas las posibilidades de actuación profesional, por medio de la incorporación de nuevas potencialidades a los procesos de representación, organización, disseminación y recuperación de informaciones.

Las ontologías favorecen ir más allá de la representación de los aspectos descriptivos y temáticos de los documentos, facilitando subsidios computacionales que posibilitan

representar los próximos dominios, permitiendo la recuperación de informaciones contextualizadas. De este modo, se puede considerar que para el área de Organización del Conocimiento las ontologías representan un avance compatible al que los tesauros proporcionaron para la recuperación en texto libre en los primordios de la Ciencia de la Información.

Lo que se puede prever, con alto grado de certeza, es que las bibliotecas digitales del futuro no serán como las identificadas en el momento actual, de modo que es necesario el desenvolvimiento de directrices teórico- metodológicas que favorezcan las enseñanzas de las nuevas competencias profesionales para representación de contenidos en ambientes digitales, evitando la fragmentación del campo de actuación y el surgimiento de dicotomías entre aquellos aptos para trabajar con ontologías y los que no lo están.

De esta manera, se torna evidente la necesidad de familiarización con las tecnologías emergentes del medio digital y del fortalecimiento de la formación lógico-matemática de los profesionales de la información, contribuyendo para la enseñanza de nuevos métodos y competencias profesionales, de representación de recursos informacionales.

No obstante, el futuro promisor de las aplicaciones basadas en ontologías apenas comienza. De este modo, se espera que este trabajo pueda contribuir para el desenvolvimiento de nuevos estudios interdisciplinarios, pues conforme afirma Wiener (1970, p. 27) *“son estas regiones fronterizas de la ciencia que ofrecen las más ricas oportunidades al investigador calificado”*.

NOTAS

¹ *OntoClean* es una metodología que posibilita evaluar y adecuar ontológicamente relacionamientos taxonómicos, exponiendo modelados inadecuados y selecciones inconsistentes (GUARINO y WELTY, 2002).

² <<http://www.jeromedl.org>>.

³ <<http://lattes.cnpq.br>>.

⁴ <<http://www.foaf-project.org>>.

⁵ <<http://www.w3.org>>.

⁶ <<http://protege.stanford.edu>>.

⁷ Este identificador, compuesto de números y letras, es atribuido al objeto digital para que este sea únicamente identificado en Internet.

BIBLIOGRAFÍA

ANSI/NISO. Z39.19 – 2005. Guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO Press, 2005, p. 184. Disponible en: <<http://www.niso.org>> [Consulta: 20 de febrero de 2007].

BARITÉ, M. Referenciales teóricos vigentes en el área de tratamiento temático de la información y su expresión metodológica. Porto Alegre: ABEED, 1998. Relatório técnico do II Encontro de Dirigentes dos cursos superiores de Biblioteconomia dos países do Mercosul, Buenos Aires, nov. 1997.

CAMPOS, M.L.A.y GOMES, H.E. Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. Perspectivas em Ciência da Informação, Belo

- Horizonte, vol. 11, nº 3, set/diez. 2006. Disponible en: <<http://www.eci.ufmg.br/pcionline/index.php/pci/article/viewFile/273>> [Consulta: 20 de octubre de 2007].
- CUNHA, M.B. Construindo o futuro: a biblioteca universitária brasileira em 2010. *Ciência da Informação*, Brasília, vol. 29, nº 1, p. 71-89, jan/abr. 2000. Disponible en: <<http://revista.ibict.br/index.php/ciinf/article/view/269>> [Consulta: 12 de mayo de 2009].
- DABROWSKI, M.; SYNAK, M. y KRUK, S.R. Bibliographic Ontology. In: KRUK; MCDANIEL (Eds.). *Semantic Digital Libraries*. Berlin: Springer, 2009, p. 103-122. Disponible en: <<http://www.springerlink.com/content/j40444k3t3348l25>> [Consulta: 16 de diciembre de 2009].
- DAHLBERG, I. *A referent-oriented, analytical concept theory for Interconcept*. Frankfurt: International Classification, 1978, vol. 5, nº 3, p. 142-151.
- DENG, Z. y TANG, S. Ontology-based Multi-agent Digital Library Servings for Information Discovery on the Web. *Computer Engineering*, 2002, vol. 33, nº 2, p. 37-39, 2002.
- ECO, H. Muito além da Internet. 2003. Disponible en: <http://www.ofaj.com.br/textos_conteudo.php?cod=16> [Consulta: 16 de diciembre de 2009].
- FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A. y JURISTO, N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. 1997. *Ontological Engineering - Working Notes*, Stanford University, California.
- FUJITA, M.S.L. Organização e representação do conhecimento no Brasil: análise de aspectos conceituais e da produção científica do ENANCIB no período de 2005 a 2007. *Tendências da Pesquisa Brasileira em Ciência da Informação*, 2008, vol. 1, nº 1.
- GARCÍA MARCO, F.J. Ontologías y organización del conocimiento: retos y oportunidades para el profesional de la información. *El profesional de la información*, 2007, vol. 16, nº 6, p. 541-550. Disponible en: <<http://www.elprofesionaldelainformacion.com>> [Consulta: 10 de enero de 2008].
- GNOLI, C. The ontological approach to knowledge organization. Invited paper sent for discussion at the 2 Seminário de pesquisa em Ontologia no Brasil. Rio de Janeiro, 21-22 Setembro, 2009. Disponible en: <<http://ontobra.comp.ime.br/ontological.rtf>> [Consulta: 10 de octubre de 2009].
- GOMES, H.E. *et al.* Manual de Elaboração de Tesouros Monolíngues. Brasília: CNPq/PNBU, 1990, p. 78.
- GREENBERG, J. Advancing the Semantic Web via Library Functions. *Cataloging & Classification Quarterly*. New York, 2007, vol. 43, nº 3-4, p. 203-225. Disponible en: <<http://www.ils.unc.edu/mrc/pdf/greenberg07advancing.pdf>> [Consulta: 10 de octubre de 2008].
- GUARINO, N. y WELTY, C. Evaluating ontological decisions with OntoClean. *Communications of the ACM*, 2002, vol. 45, nº 2.
- GUIMARAES, J.A.C. Abordagens teóricas de tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental. In: GARCÍA MARCO, F. J. *Avances y perspectivas en sistemas de información y documentación*. Ibersid, Zaragoza, 2009, p. 105-117.
- GUIMARÃES, J.A.C. y SALES, R. Análise documental: concepções do universo acadêmico brasileiro em Ciência da Informação. *Datagramazero*, Rio de Janeiro, 2010,

- vol. 11, nº 1. Disponible en: <http://www.dgz.org.br/fev10/Art_02.htm> [Consulta: 20 de febrero de 2010].
- GUIZZARDI, G. *Ontological Foundations for Structural Conceptual Models*. The Netherlands: Universal Press, 2005. Disponible en: <http://doc.utwente.nl/50826/1/thesis_Guizzardi.pdf> [Consulta: 20 de octubre de 2009].
- HU, Z. An Ontology-based Framework for Knowledge Service in Digital Library. In: *Wireless Communications, Networking and Mobile Computing - WiCom 2007, 2007*. Disponible en: <<http://ieeexplore.ieee.org>> [Consulta: 5 de enero de 2009].
- KRUK, S.R.; DECKER, S. y ZIEBORAK, L. JeromeDL reconnecting digital libraries and the semantic Web, 2005. Disponible en: <http://www.marcont.org/marcont/pdf/www2005_jeromedl.pdf> [Consulta: 20 de marzo de 2008].
- LI, P. y SUN, L. The Construction of Information Resources Ontology at Digital Libraries. *Library and Information Service*, 2003, vol. 133, nº 6, p. 24-27. Disponible en: <<http://caod.oriprobe.com>> [Consulta: 10 de diciembre de 2009].
- MCGUINNESS, D.L. y HARMELEN, F. OWL Web Ontology Language Overview. W3C Recommendation, 10 february, 2004. Disponible en: <<http://www.w3.org/TR/2004/REC-owl-features-20040210>> [Consulta: 16 de diciembre de 2010].
- NOY, N.F. y McGuinness, D.L. *Ontology development 101: a guide to creating your first ontology*. Stanford: Stanford University, 2001. Disponible en: <<http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html>> [Consulta: 20 de febrero de 2009].
- RAMALHO, R.A.S. *Desenvolvimento e utilização de ontologias em Bibliotecas Digitais - uma proposta de aplicação*. 2010. 145f. Tese (Doutorado) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.
- SÁNCHEZ CUADRADO, S. *et al.* De repente, ¿todos hablamos de ontologías? *El profesional de la información*, 2007, vol. 16, nº 6, p. 562-568. Disponible en: <<http://www.elprofesionaldelainformacion.com>> [Consulta: 10 de enero de 2008].
- SHUM, *et al.* Modeling naturalistic argumentation in research literatures: Representation and interaction design issues. *International Journal of Intelligent Systems*, 2007, vol. 22, nº 1.
- SHUM, S.B.E. y MOTTA DOMINGUE, J. ScholOnto An ontology-based digital library server for research documents and discourse. *International Journal on Digital Libraries*, vol. 3, nº 3, p. 237-248, Sept. 2000. Disponible en: <<http://citeseerx.ist.psu.edu>> [Consulta: 10 de diciembre de 2008].
- SOERGEL, D. The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science*, 1999, vol. 50, nº 12, p. 1119-1120. Disponible en: <<http://www.dsoergel.com/cv/B70.pdf>> [Consulta: 10 de enero de 2010].
- SWATSON, D.R. Historical Note: Information Retrieval and the Future of an Illusion. *Journal of the American Society for Information Science*, 1988, vol. 39, nº 2, p. 92-98.
- USCHOLD, M. y KING, M. Towards a Methodology for Building Ontologies. In: *workshop on basic ontological issues in knowledge sharing*, 1995.
- USCHOLD, M. y GRUNINGER M. Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, 1996, vol. 11, nº 2, p. 93-136.

- VICKERY, B.C. Ontologies. *Journal of Information Science*, 1997, vol. 23, n° 4, p. 272-286. Disponible en: < <http://www.let.uu.nl/~paola.monachesi/personal/papers/vick97-onto.pdf> > [Consulta: 10 de diciembre de 2008].
- WEINSTEIN, P.C. y BIRMINGHAM, W.P. Creating ontological metadata for digital library content and services. *International Journal on Digital Libraries*, 1998, vol. 2, n° 1, p. 20-37. Disponible en: < <http://deepblue.lib.umich.edu/handle/2027.42/42334> > [Consulta: 15 de diciembre de 2007].
- WIENER, N. *Cibernética*. São Paulo: EDUSP: Polígono, 1970.