

El rol de las herramientas y tecnologías de la Web Semántica en las Bibliotecas Digitales
Maria Rosa Mostaccio, Facultad de Filosofía y Letras - UBA ,Púan 480, CABA,1406
subsecbibliotecas@filo.uba.ar

Resumen

El propósito de este trabajo es analizar y evaluar posibles aplicaciones de las herramientas y tecnologías de la web semántica en las bibliotecas digitales que potencian las posibilidades de búsquedas mediante el descubrimiento y la navegación por los datos de los recursos, que permiten utilizar conceptos bien definidos e inferir datos implícitos a partir de descripciones bibliográficas, y fundamentalmente que posibilitan la interoperabilidad entre distintos sistemas y usuarios describiendo y organizando los recursos en un lenguaje comprensible por las computadoras o agentes.

En la primera parte de dicho trabajo, se analiza el rol de las bibliotecas digitales en la web semántica y de que manera utilizando herramientas de la web semántica puedan proveer una mayor interoperabilidad, normalización en las descripciones bibliográficas y una mejor recuperación de la información. Se plantean los beneficios de la biblioteca digital semántica en relación a la biblioteca digital actual.

Luego en una segunda parte, se presenta una aplicación específica de ontología asociada a la descripción bibliográfica, es la iniciativa MarcOnt prototipo utilizado en la Biblioteca Digital Jerome, del Digital Enterprise Research Institute (DERI) y la Gdansk University of Technology (GUT).

En una tercera parte, se incluye el modelo FRBR utilizado en Greenstone 3, que si bien se encuentra en etapa de investigación, resulta interesante observar la interrelación establecida entre una interfase de servicio de alerta para sus usuarios basado en la FRBR que mejora la representación y recuperación de la información como así también la interoperabilidad con otros sistemas.

Mediante la utilización de estas tecnologías se brindará mejores servicios a los usuarios, permitiendo que amplíen sus expectativas sobre la precisión de la información recuperada y afinen en la obtención de resultados satisfactorios, y también se logrará una mayor interoperabilidad integrando metadatos desde diferentes fuentes heterogéneas y entre diferentes sistemas como así también la integración y participación activa de sus usuarios.

Introducción

Si bien las bibliotecas digitales ofrecen funcionalidades de referencias cruzadas, de búsquedas por texto completo y por metadatos asignados que mejoran las técnicas de recuperación pero estas funcionalidades son implementadas en un contexto determinado para una audiencia específica en una institución determinada, y si bien se publican en la web y comparten el mismo esquema de metadatos, como Dublin Core u otros, utilizan lenguajes documentales diferentes para la representación del contenido y están escritos en diferentes idiomas, le faltan métodos genéricos para la integración armoniosa de los contenidos generados por diferentes fuentes y una infraestructura flexible para implementar nuevas funcionalidades.

Con la ayuda de las herramientas semánticas y en especial de las ontologías se pueden combinar consultas estructuradas de metadatos con búsquedas en texto completo de los recursos, como así también lograr una interoperabilidad entre sistemas.

Se analiza el rol de las bibliotecas digitales en la web semántica y de que manera utilizando herramientas de la web semántica puedan proveer una mayor interoperabilidad, normalización en las descripciones bibliográficas y una mejor recuperación de la información. Se plantean los beneficios de la biblioteca digital semántica en relación a la biblioteca digital actual.

La aplicación de tecnologías de la web semántica en las bibliotecas digitales permitirá describir y organizar los recursos en un lenguaje legible por máquina o agentes ,presentar la semántica de los metadatos en términos de ontologías ,utilizando un lenguaje formal, interoperabilidad con otros sistemas y una ampliación de los recursos no ya limitado a contenidos ,objetos digitales , organización de objetos (por. ej. colecciones) , vocabularios controlados, tesauros, taxonomías sino también integrar como recursos a usuarios y comunidades de usuarios.

Bibliotecas Digitales y la Web Semántica

“El concepto de biblioteca digital no es únicamente el equivalente de repertorios digitalizados con métodos de gestión de la información. Es más bien, un entorno donde se reúnen colecciones, servicios, y personal que favorece el ciclo completo de la creación, difusión, uso y preservación de los datos, para la información y el conocimiento". (Santa Fe Workshop on Distributed Knowledge Work Environments: Digital Libraries, 1997)

La biblioteca digital no pretende "copiar" la producción impresa, sino que debe generar una nueva estructura de la información; el uso de documentos hipertextuales hacen que un "libro electrónico" ya no sea lineal, como ocurre con la confección en papel, sino que el libro ahora también posee profundidad", el usuario tiene acceso a la información de formas muy variadas, cuenta con la posibilidad de proveer vínculos no solo a documentos textuales, sino también a imagen y vídeo, de modo instantáneo, lo cual permite explicar dichos contenidos de formas diversas.

Si bien las bibliotecas digitales utilizan metadatos y vocabularios controlados o tesauros ,pero lo que se requiere ahora es el cambio hacia un nuevo modelo que sea comprensible para las computadoras o agentes y proporcionar semántica a los metadatos.

Las nuevas herramientas de la web semántica están produciendo un gran cambio en la arquitectura de las bibliotecas digitales, como así también contribuyendo a un mejoramiento en la recuperación y en la organización de la información.

Existen actualmente aplicaciones que están utilizando estas herramientas de la web semántica y hay varios autores, entre ellos Kruk, S.R. et al (2006) que las citan como Bibliotecas Digitales Semánticas y en algunos casos se las citan como Bibliotecas Digitales Sociales Semánticas , utilizando la ontología FOAF (Friend -of- a -Friend) para integrar las redes sociales a las bibliotecas.

Una biblioteca digital semántica es una biblioteca digital cuyas colecciones están formadas por "documentos funcionales". Un documento funcional, comienza con un objeto digital, como en una colección de biblioteca digital, documento que consta de una "reproducción digital fiel", con sus características de calidad, permanencia, integridad, interoperabilidad, etc., pero no se queda en sólo algunos simples metadatos agregados para identificarlo. Un “documento funcional” es un objeto digital de calidad, semánticamente rico, el cual, -además de los cuatro tipos de metadatos estándar-y gracias a otros metadatos de hipervínculo ha sido definido y ligado de tal forma con otros documentos funcionales de manera tal que permite su óptimo reuso, búsqueda, integración, minería de datos e interoperabilidad proveyendo una lista de sus contenidos permitiendo las interfaces con otros programas externos que puedan explotarlos. No se trata de una inteligencia artificial mágica que permita a las máquinas entender las palabras de los usuarios, es sólo la habilidad programada en una máquina para resolver problemas bien definidos, a través de operaciones bien definidas que se llevarán a cabo sobre datos existentes bien definidos.

Además hay una creciente necesidad de las bibliotecas e instituciones culturales a que cooperen unos con otros y que puedan ofrecer sus objetos culturales y contenido digital a un público más amplio mediante un sistema común de biblioteca digital. Para lo cuál es esencial establecer un acceso uniforme de metadatos proporcionados por las distintas instituciones implicadas para brindar servicios de búsqueda y descubrimiento de recursos a sus usuarios finales. Existen aplicaciones de

Bibliotecas digitales sociales semánticas, que además de almacenar el contenido y los metadatos, también realizan un seguimiento de sus usuarios, sus interacciones, y permiten que los mismos incorporen sus conocimientos al sistema.

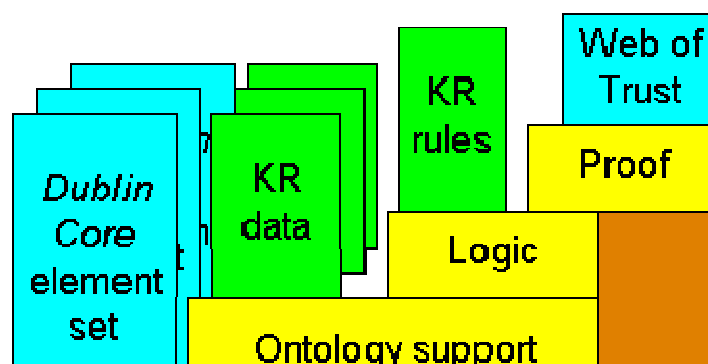
Aspectos básicos de la Web Semántica

Para dar sentido a la información que contiene la web es necesario basarse en la capacidad de asociar semánticamente datos independientes entre sí, pero relacionados en un determinado contexto. Uno de los objetivos que se procuran con la Web Semántica consiste en definir y describir los datos (no los documentos como la web 1.0) contenidos en la Web, de tal forma que puedan ser utilizados y comprendidos por las máquinas sin necesidad de intervención humana. De este modo, la Web se puede convertir en un espacio navegable y comprensible, donde es posible la relación entre términos independientes de una misma búsqueda. Es necesario establecer los aspectos básicos de la web semántica para luego poder aplicarlos en el mejoramiento de la organización y recuperación de la información en las bibliotecas digitales.

La Web semántica se basa en la idea de añadir metadatos semánticos y ontológicos a la World Wide Web. Esas informaciones adicionales —que describen el contenido, el significado y la relación de los datos— se deben proporcionar de manera formal, para que así sea posible evaluarlas automáticamente por máquinas de procesamiento. El objetivo es mejorar Internet ampliando la interoperabilidad entre los sistemas informáticos usando "agentes inteligentes"

La Web semántica identifica un conjunto de tecnologías, herramientas y estándares que forman los bloques básicos de una infraestructura que dan soporte a la visión de la Web asociada con el significado. La arquitectura de la Web semántica se compone de una serie de estándares organizados en una cierta estructura que es una expresión de sus interrelaciones. Esta arquitectura es a menudo representada usando el diagrama propuesto por primera vez por Tim Berners-Lee (Berners - Lee, Hendler et al. 2001).

Se presenta a continuación el gráfico extraído de: Tim Berners.Lee 2000
<http://www.w3.org/2000/Talks/0906-xmlweb-tbl/slide9-0.html>



Si se analiza dicho gráfico pueden considerarse que los principales componentes de la Web Semántica son los siguientes:

- **Unicode:** el alfabeto. Se trata de una codificación del texto que permite utilizar los símbolos de diferentes idiomas sin que aparezcan caracteres extraños. De esta forma, se puede expresar información en la Web Semántica en cualquier idioma.
- **URI:** las referencias. URI es el acrónimo de "Uniform Resource Identifier" o Identificador Uniforme de Recursos, identificador único que permite la localización de un recurso que puede ser accedido vía Internet. Se trata del URL (descripción de la ubicación) más el URN (descripción del espacio de nombre).
- **XML + NS + xmlschema:** se trata de la capa más técnica de la Web Semántica. En esta capa se agrupan las diferentes tecnologías que hacen posible que los agentes puedan entenderse entre ellos. XML ofrece un formato común para intercambio de documentos, NL (namespaces) sirve para cualificar elementos y atributos de nombres usados en XML asociándolos con los espacios de nombre identificados por referencias URI y XML Schema ofrece una plantilla para elaborar documentos estándar. De esta forma, aunque se utilicen diferentes fuentes, se crean documentos uniformes en un formato común y no propietario.
- **RDF + rdfschema:** basada y apoyada en la capa anterior, esta capa define el lenguaje universal con el cual podemos expresar diferentes ideas en la Web Semántica. RDF es un estándar propuesto por el W3C para definir la arquitectura necesaria para soportar metadatos (descripciones) sobre recursos web. Es la aplicación XML recomendada por el consorcio para codificar, intercambiar y reutilizar metadatos estructurados. Es una base para procesar metadatos, proporciona interoperabilidad entre aplicaciones que intercambian información legible por máquina en la web. Por su parte RDF Schema provee un vocabulario definido sobre RDF que permite el modelo de objetos con una semántica claramente definida. Esta capa no sólo ofrece descripción de los datos, sino también cierta información semántica.
- **Lenguaje de Ontologías:** ofrece un criterio para catalogar y clasificar la información. El uso de ontologías permite describir objetos y sus relaciones con otros objetos ya que una ontología es la especificación formal de una conceptualización de un dominio concreto del conocimiento. Esta capa permite extender la funcionalidad de la Web Semántica, agregando nuevas clases y propiedades para describir los recursos.

- **Lógica:** además de ontologías se precisan también reglas de inferencia. Una ontología puede expresar la regla "Si un código de ciudad está asociado a un código de estado, y si una dirección es el código de ciudad, entonces esa dirección tiene el código de estado asociado". De esta forma, un programa podría deducir que una dirección de la Universidad Complutense, al estar en la ciudad de Madrid, debe estar situada en España, y debería por lo tanto estar formateado según los estándares españoles. La computadora no "entiende" nada de lo que está procesando, pero puede manipular los términos de modo mucho mas eficiente beneficiando la inteligibilidad humana.
- **Pruebas:** será necesario el intercambio de "pruebas" escritas en el lenguaje unificador (se trata del lenguaje que hace posible las inferencias lógicas hecha posibles a través del uso de reglas de inferencia tal como es especificado por las ontologías) de la Web Semántica.
- **Confianza:** los agentes deberían ser muy escépticos acerca de lo que leen en la Web Semántica hasta que hayan podido comprobar de forma exhaustiva las fuentes de información. (Web Of Trust RDF Ontology -WOT- <http://xmlns.com/wot/0.1/> y FOAF <http://xmlns.com/foaf/0.1/>)
- **Firma digital:** bloque encriptado de datos que serán utilizados por los computadoras y los agentes para verificar que la información adjunta ha sido ofrecida por una fuente específica confiable. (XML Signature WG: <http://www.w3.org/Signature/>)

En el gráfico presentado se describe el papel de las ontologías junto con la utilización de un lenguaje semántico como RDF, pero pueden utilizarse otros lenguajes semánticos específicos. RDF carece de poder expresivo (negación, implicación, cardinalidad, etc). Por ejemplo, no es posible especificar las condiciones necesarias y suficientes para definir la pertenencia a una clase. Para lograr una mayor expresividad para el procesamiento semántico, se ha creado el lenguaje OWL(*Ontology Web Language*) , una especificación del W3C para especificar ontologías. Por ejemplo, OWL sí permite definir las condiciones necesarias y suficientes para definir la pertenencia a una clase, luego, aunque basado en el lenguaje RDF, va más allá que RDF Schema.

Pero para que la Web Semántica sea una realidad, precisa tanto de un lenguaje de consulta estándar y de un protocolo de recuperación. Con este fin, el W3C ha desarrollado los siguientes lenguajes, que habría que incluirlos en el gráfico anterior:

- SPARQL Query Language for RDF. <http://www.w3.org/TR/rdf-sparql-query/> Se trata de la especificación para el lenguaje de consulta.
- SPARQL Protocol for RDF. <http://www.w3.org/TR/rdf-sparql-protocol/> Es la especificación del protocolo de recuperación.

Ontologías en las bibliotecas digitales

Como pudo observarse en el gráfico anterior las ontologías están en una de las capas superiores pero se basan en las capas inferiores.

Una ontología es 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. Esto quiere decir que una ontología especifica una conceptualización o una forma de ver el mundo, por lo que cada ontología incorpora un punto de vista. Además, una ontología contiene definiciones que nos

proveen del vocabulario para referirse a un dominio. Estas definiciones dependen del lenguaje que usemos para describirlas. Todas las conceptualizaciones (definiciones, categorizaciones, jerarquías, propiedades, herencia, etc.) de una ontología pueden ser procesables por máquina.

Según Gruber, las ontologías se componen de:

- **conceptos:** son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.
- **relaciones:** representan la interacción y enlace entre los conceptos de un dominio. Suelen formar la taxonomía del dominio. Por ejemplo: subclase-de, parte-de, parte-exhaustiva-de, conectado-a, etc.
- **funciones:** son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como: asignar-fecha, categorizar-clase, etc.
- **instancias:** se utilizan para representar objetos determinados de un concepto.
- **reglas de restricción o axiomas:** son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: "Si A y B son de la clase C, entonces A no es subclase de B", "Para todo A que cumpla la condición B1, A es C", etc. Los axiomas, junto con la herencia de conceptos, permiten inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos.

Aplicaciones de Ontologías en las bibliotecas digitales

- Una de las aplicaciones de ontologías es el prototipo MarcOnt que esta aplicando el proyecto JeromeDL como plataforma de pruebas. La ontología MarcOnt, se basa en un acuerdo social en el que se combinan las descripciones de MARC21 junto con DublinCore y hace uso de todo el potencial de las tecnologías de Web Semántica. Esto incluye traducciones a / de otras ontologías, mejora el soporte para L2L, y es más eficiente la búsqueda de recursos (es decir, los usuarios pueden tener un impacto en el proceso de búsqueda).

Uno de los estándares para descripción bibliográfica es MARC21 que si bien permite describir la mayoría de las características de los recursos de la biblioteca, su contenido semántico es bajo. Esto significa que en una búsqueda de un recurso, se puede buscar por palabras clave en campos determinados pero no se puede realizar una búsqueda por conceptos y la comunicación de datos entre sistemas no es extensible. Con la Web Semántica surgió el esquema de metadatos de Dublin Core para la descripción de los recursos de la biblioteca, por lo tanto la mayoría de la información cubierta por MARC21 se pierde, por tal razón para aprovechar el potencial de la web semántica es que se ha desarrollado una ontología de MARC21.

El prototipo Jerome DL presenta una arquitectura de tres capas para la gestión de metadatos, donde cada capa enriquece la información básica recopilada en la biblioteca con anotaciones semánticas y ofrece capacidades adicionales para navegación y búsqueda. La capa inferior se encarga tareas típicas y necesarias de un repositorio de objetos digitales, es decir, realiza un seguimiento de la representación física de los recursos, su estructura y la procedencia.

La capa media presenta las descripciones bibliográficas a un nivel semántico; ya que utiliza una ontología extensible capaz de representar la información inicialmente prevista en cualquiera de los formatos estándares existentes, como Dublin Core, MARC21 o BibTEX. Los servicios de la capa media se basan en el almacenamiento, entrega y gestión de

metadatos de los documentos . Además, la capa intermedia, ofrece la recuperación de la información y gestión de identidad de los recursos. La capa superior permite a la comunidad de usuarios participar en las anotaciones o metadatos y en el filtrado de los recursos , para lo cual posee una interfaz de búsqueda en lenguaje natural con plantillas y el sistema lo traduce automáticamente en RDF.

Según la plantilla utilizada permite, por ej. :

- buscar en todas las publicaciones de la biblioteca.....
- buscar en todas las publicaciones escritas por
- buscar en publicaciones escritas por amigos de ...
- buscar en publicaciones escritas por estudiantes de.....
- encontrar artículos relativos a una materia de (en el contexto dominio)

The screenshot displays the JeromeDL web application interface. The browser window title is "JeromeDL - Búsqueda de recursos : Búsqueda semántica - Iceweasel". The address bar shows the URL: http://bleedingedge.jeromedl.org/search?lang=es_ES&type=semantic&donotexecute=false&rdfqtype=nlq&nqlc. The page title is "Resultados de búsqueda de recursos RDF y interfaz NL".

On the left, there is a sidebar with navigation links: "Navegar en la biblioteca" (This Digital Library's Homepage, Informes técnicos(22), En libros(4), Tesis(18), Diversos(236), Artículos(34), Libros(9), Memorias de conferencias(8), En colecciones(1), En memorias de conferencias(245), Remitibles(61)), "Browse Collections", "Semantic browse", "Búsqueda de recursos", and "Investigación en DL".

The main content area shows a search template selection interface. The selected template is "Show me all publications written by ...". The input field contains the text "Show me all publications written by Sebastian Kruk...". Below this, there are buttons for "Buscar" and "Limpiar el formulario".

The search results section, titled "Resultados de búsqueda", shows "Resultados 1 - 10 de 60". The first result is "Sebastian Ryszard Kruk, Slawomir Grzonkowski, Adam Gzella, Mariusz Cygan :DigiMe - Ubiquitous Search and Browsing for Digital Libraries". Other results include "Sebastian Ryszard Kruk, Bernhard Haslhofer, Piotr Piotrowski, Adam Westerski, Tomasz Woronicki :The Role of Ontologies in Semantic Digital Libraries" and "Sebastian Ryszard Kruk, Adam Gzella, Filip Czaja, Władysław Bultrowicz, Ewelina Kruk :MultiBeeBrowse - Accessible Browsing on Unstructured Metadata".

On the right, there is a sidebar with "JeromeDL abreviatura", "Perfil del lector" (Login, Registro), "Preseleccionar estilo" (Por defecto), "Mostrar metadatos sin formatear" (RDF, RDF/XML, RDF/NS, RDF/TURTLE, RDF/NTriples, RDF/JSON), and "Sindicar" (RSS 1.0, RSS 2.0, ATOM, FEED).

En el ejemplo de búsqueda se utilizó la opción de buscar por todas las publicaciones escritas por Sebastian Kruk

Como resultado se visualizan todas las obras del autor, pero para limitar o afinar los resultados , permite una búsqueda semántica propuesta por el usuario , por ej. palabras claves “marcont” y por tipo de documento :

JeromeDL - JeromeDL with Exhibit {Select only marccont in facet Palabras claves:} - Iceweasel

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

http://bleedingedge.jeromedl.org/exhibit/index.jsp?state=search&ref=http%3A%2F%2Fbleedingedge.jeromedl.org

Facultad de Filosofía ... Campus Virtual de Fi... Bienvenido a Webma... IRM - Recursos Infor...

Idioma es_ES Buscar RSS

Resultados de búsqueda de recursos RDF y Interfaz NL

(Browsing with Exhibit)

TILES • DATES • PLACES

2 Books filtered from 60 originally (Reset All Filters)

ordenados por: bookType; luego por... • ☒ agrupar según orden

Inproceedings (2)

- The Role of Ontologies in Semantic Digital Libraries**
 Author: Sebastian Ryszard Kruk, Bernhard Haslhofer, Piotr Piotrowski, Adam Westerski, and Tomasz Woroniecki
 Keywords: ontology, foaf, foaf, marccont, digital libraries, and semantic web
 Event: Related to: The Role of Ontologies in Semantic Digital Libraries
 Conference: The 10th European Conference on Research and Advanced Technology for Digital Libraries
- Collaborative Ontology Development with MarcOnt Portal**
 Author: Sebastian Ryszard Kruk and Maciej Dąbrowski
 Keywords: marccont portal, corrib, marccont, ontology, and collaboration
 Event: Related to: Collaborative Ontology Development with MarcOnt Portal
 Conference: Semantic Technology Conference 2007

JeromeDL abreviatura

Perfil del lector

- Login
- Registro

Exhibit filter panel 1

Palabras claves:

- learning objects
- library
- marc21 bibtext
- marcont

Categories:

- (missing this field)

Redactores:

- (missing this field)

Editores:

- (missing this field)

Publication types:

- inproceedings

Lista

También permite limitar los resultados de búsqueda de acuerdo a filtros determinados por el usuario para recuperar los documentos pertinentes y relevantes de acuerdo al perfil del usuario.

Se visualiza a continuación un acceso al recurso, donde se encuentra el pdf para acceder al texto completo, descripción adicional, recomendaciones y opciones para traducir automáticamente dicho recurso en RDF para MarcOnt, Dublin Core y BibTex.

- Otra aplicación es el modelo FRBR en **Greenstone 3**, que se encuentra en etapa de investigación, aún no está accesible para su visualización, pero está desarrollando una nueva plataforma para bibliotecas digitales semánticas. Este cambio tendrá profundos efectos en la recuperación del contenido para los usuarios finales e introducirá también un impacto significativo en la arquitectura de la biblioteca digital actual.

Los elementos claves de esta nueva plataforma son: un servicio de alerta para sus usuarios basado en la ontología FRBR, que mejora la representación y recuperación, y la interoperabilidad con otros sistemas.

La infraestructura de Greenstone 3 se desarrolla en dos fases, para los fines de este trabajo se analizará la *fase 1: la semántica de los documentos y de las colecciones*. Se describen y analizan las posibles acciones requeridas por los usuarios en el servicio de alerta de Greenstone. Estas tareas se realizan mediante un modelo semántico para las colecciones y contenidos de la BD. Estos requerimientos pueden lograrse mediante una ontología en la interfaz bibliotecaria, un subsistema interactivo para crear y mantener colecciones de la biblioteca digital.

El Servicio de alerta de Greenstone (Greenstone-AS) es un sistema de alerta genérico a través de varias bibliotecas, incluyendo Fedora y DSpace. Permite la creación de perfiles de usuarios de acuerdo a sus necesidades. Estos perfiles se encuentran en una red de servidores de bibliotecas digitales y el usuario puede mantenerse informado de cualquier cambio de documentos en cualquier servidor. Además, Greenstone envía otro tipo de notificaciones, tales como la creación de una nueva clasificación dentro de un tema, o la adición de una nueva colección de material existente en un servidor.

Las acciones básicas consideradas para la creación de perfiles de usuarios fueron las siguientes:

Un usuario puede querer ser notificado cuando :

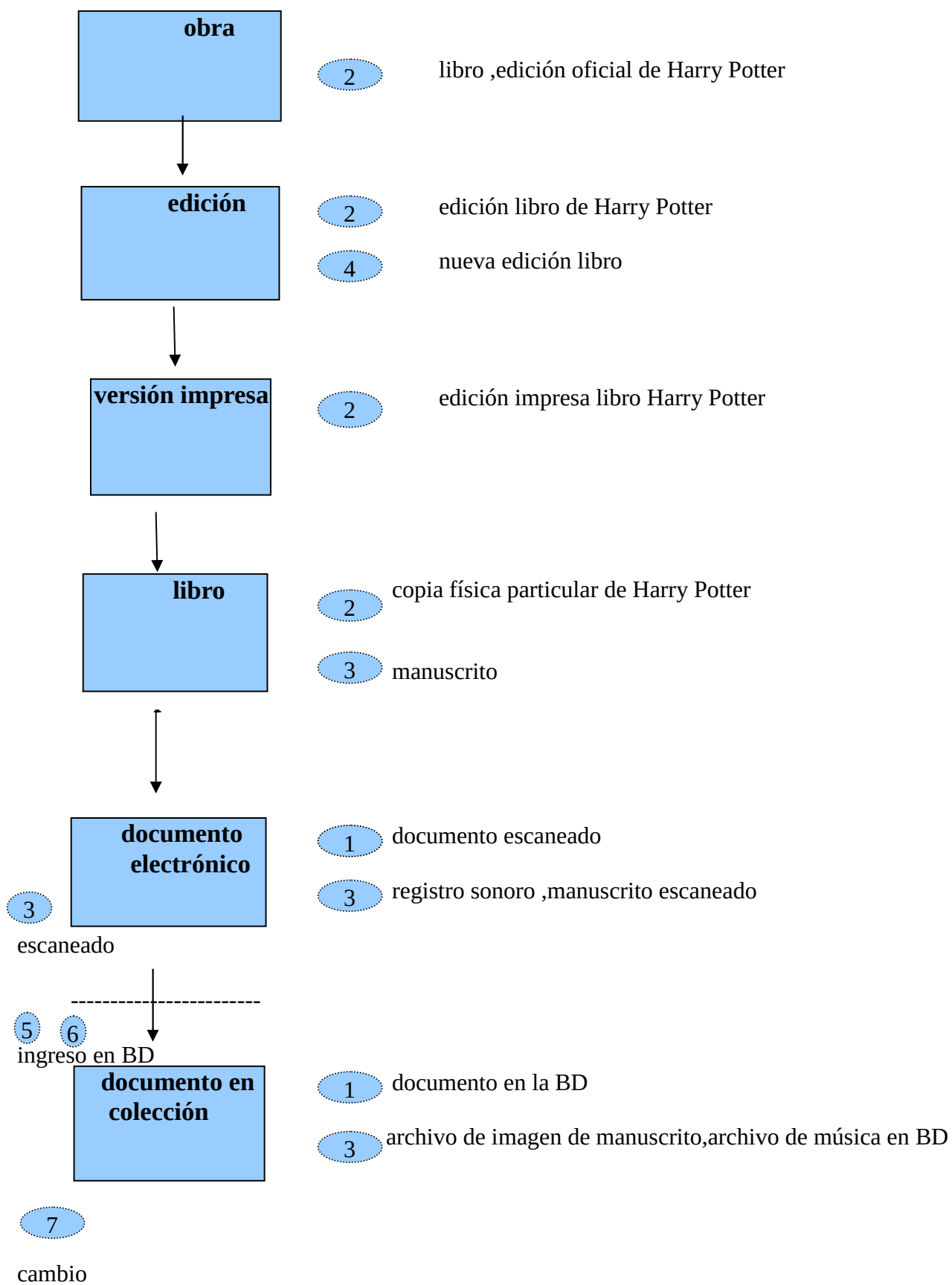
1. Una vez que un nuevo documento electrónico se ha incorporado y se encuentra disponible en la colección de la biblioteca digital.
2. Una nueva obra ha sido publicada.
3. Un manuscrito ha sido nuevamente digitalizado, en una resolución mayor, o en un formato diferente.
4. Otra edición del mismo libro se publica.
5. Un documento electrónico ha sido recientemente publicado en la biblioteca digital. Si bien el documento no es nuevo, pero se ha incorporado recientemente a la colección.
6. Un documento se elimina de la biblioteca digital.
7. Un documento electrónico ha cambiado: por ejemplo, la documentación de software en línea puede ser escrito y adaptado continuamente. Propiedades similares para blogs y wikis.

Estos diferentes perfiles de usuario tienen implicaciones para el modelo semántico utilizado por la biblioteca digital.

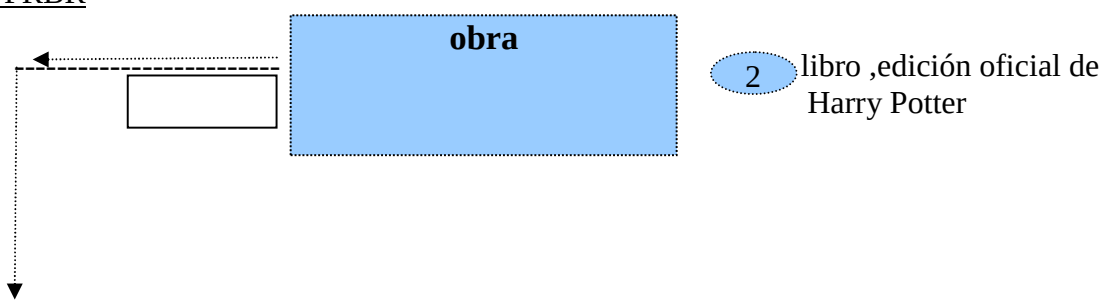
Como puede observarse en el gráfico Visualización perfil servicio de alerta, por ejemplo, el caso de una obra de Harry Potter, puede observarse los distintos formatos o soportes de una misma obra, que se encuentran disponibles en una biblioteca digital, utilizando los mismos números de las acciones mencionadas antes, para mostrar las jerarquías semánticas que hay que considerar para responder a las necesidades de los usuarios y en paralelo puede observarse las mismas acciones dentro de la ontología FRBR. Esta comparación es para mostrar que a partir de las jerarquías semánticas establecidas con la ontología FRBR es posible establecer un perfil de servicio de alerta de acuerdo a necesidades de los usuarios.

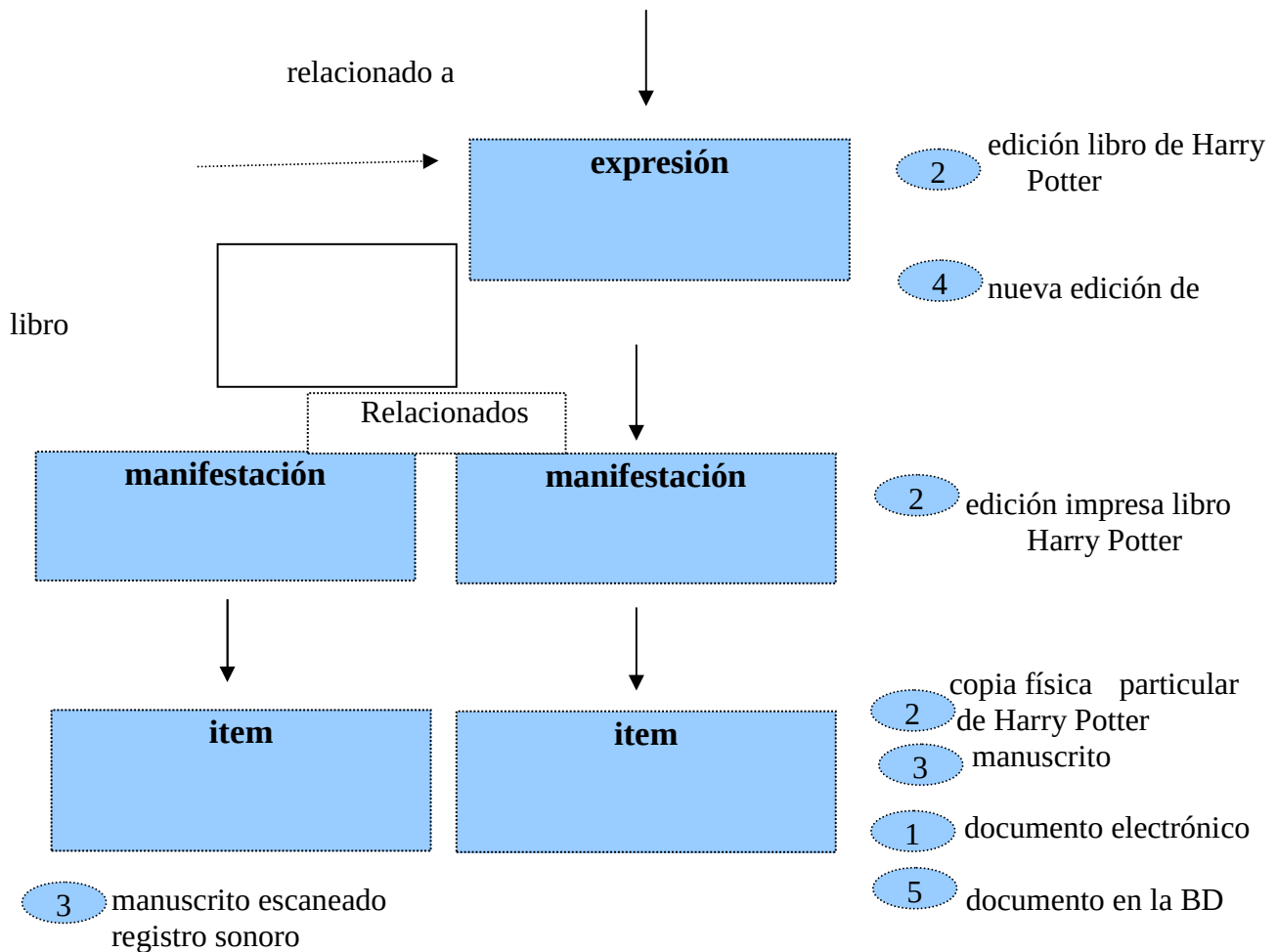
Puede observarse también que la copia escaneada se visualiza como un derivado de la copia física (impresa). Si bien en ambos casos, el modelo lógico podría ser alterado, o representado de una manera diferente, la digitalización a menudo permite nuevas posibilidades de escaneados, y un buen sistema de Biblioteca Digital debe incluir la capacidad para diferenciar estas variaciones, en particular, cuando se incluye material alojado en otros lugares.

Visualización Perfil servicio de alerta



Visualización FRBR





El manuscrito, así como el documento electrónico del manuscrito son ítems, que se refieren a diferentes manifestaciones unos de otros y pueden ser formulados de manera explícita sobre las dos manifestaciones. Las acciones 6 y 7 son propiamente de la biblioteca digital no están representadas en el modelo FRBR.

Lo que puede observarse con la aplicación del modelo FRBR es que proporciona potentes herramientas para la solución de algunas búsquedas de los usuarios.

El uso de cualquier ontología o modelo semántico, como FRBR, requiere cambios en la arquitectura de las bibliotecas digitales tradicionales, no es suficiente con la aplicación de estándar de metadatos, que tienen poca o ninguna jerarquía o interrelaciones entre conceptos y no está orientada a los objetos como el enfoque ontológico de FRBR.

Las diferencias pueden visualizarse más fácilmente si se compara el ingreso de un nuevo documento. Un paso clave es la inclusión de metadatos de autor en el documento, usando el Dublin Core en una arquitectura de biblioteca digital tradicional, simplemente hay que añadir los datos en el campo DC.creator. Pero en FRBR uno crea una relación entre una obra y un creador. El creador suele ser una persona concreta, que está representado por un objeto específico en el repositorio de FRBR. Por lo tanto, lo que en un planteamiento es representado por un campo de metadatos, en el otro es representado por una relación objeto-a-objeto. La primera se limita a exigir la introducción de texto, mientras que el último requiere una consulta para ser ejecutado en el repositorio FRBR, y tal vez, se creará un nuevo objeto.

Estas diferencias tienen repercusiones claras sobre el sistema de Greenstone 3 y en la interacción del diseño. Por ejemplo, una búsqueda de autor tendrá como resultados en este sistema, uno o más

objetos autor y sus documentos correspondientes, mientras que en Greenstone 2 ,es suficiente con una búsqueda simple en los metadatos de todos los documentos

En términos de interacción, una arquitectura tradicional de biblioteca digital no tiene en el sistema ningún concepto de autor como un objeto, el autor no puede ser representado en la interfaz directamente.

En contraste, Greenstone con FRBR representa explícitamente el autor como objeto. Por consiguiente, se puede crear una página específica para cada autor con una lista de sus obras y datos biográficos. Una analogía de esto podría ser creada en una arquitectura tradicional, pero la información biográfica debe ser codificada en el documento de la colección y de manera similar, los nombres de los autores tendrían que distinguirse mediante la creación manual del autor del documento y las biografías del autor estarían mezcladas con el material original de una colección.

FRBR es sólo un ejemplo de apoyo ontológico. Otros sistemas podrían proporcionar ventajas adicionales o complementarias. Las ontologías también pueden brindar apoyo a los metadatos de las bibliotecas tradicionales. Por ejemplo, el ingreso de nuevos documentos puede ser asistido mediante la extracción de datos relevantes de una ontología. Una vez que una expresión particular o una manifestación de una obra FRBR se asocia con un nuevo documento digital, los datos de FRBR pueden ingresarse como metadatos a través del simple paso de generar los datos FRBR en un formato de metadatos compatibles como MARC.

Los metadatos utilizados en las bibliotecas digitales, tales como Greenstone 2 y DSpace, proporcionan la recuperación del documento a través de los metadatos de los documentos y / o texto completo. En el caso de la aplicación de FRBR en Greenstone , la recuperación ya no es la misma. Mientras que la sintaxis puede parecer similar, por ej. recuperación de un documento por el nombre del autor, las operaciones han cambiado. Además, la gama de las operaciones de recuperación posible, se expande más allá de lo simplemente expresadas por la búsqueda tradicional.

Es poco probable que un sólo modelo ontológico pueda cubrir todas las necesidades posibles, y FRBR no contiene ninguna representación directa de las jerarquías de materias, aunque incluye información particular sobre el tema de los documentos individuales (por ejemplo, de personas, lugares o tema) . FRBR puede proporcionar un método para mejorar el modelo de conexión de los documentos, autores y editores, clasificaciones jerárquicas y permitir un modelo estructurado de mapas de los temas del documento.

Conclusiones

Con este trabajo explorativo se intenta abrir nuevos caminos de investigación en las herramientas y tecnologías de la web semántica para aplicarlas a las bibliotecas digitales para mejorar la recuperación y organización de la información.

Si bien las bibliotecas digitales ofrecen funcionalidades de referencias cruzadas, de búsquedas por texto completo y por metadatos asignados que mejoran las técnicas de recuperación pero estas

funcionalidades son implementadas en un contexto determinado para una audiencia específica en una institución determinada , y si bien se publican en la web y comparten el mismo esquema de metadatos ,como Dublin Core u otros ,utilizan lenguajes documentales diferentes para la representación del contenido y están escritos en diferentes idiomas ,le faltan métodos genéricos para la integración armoniosa de los contenidos generados por diferentes fuentes y una infraestructura flexible para implementar nuevas funcionalidades.

Con la ayuda de las herramientas semánticas y en especial de las ontologías se pueden combinar consultas estructuradas de metadatos con búsquedas en texto completo de los recursos, como así también lograr una interoperabilidad entre sistemas.

No se trata de una inteligencia artificial mágica que permita a las máquinas entender las palabras de los usuarios, es sólo la habilidad programada en una máquina para resolver problemas bien definidos, a través de operaciones bien definidas que se llevarán a cabo sobre datos existentes bien definidos.

Las posibles aplicaciones y usos de las ontologías en las bibliotecas digitales permitirán normalizar los atributos de los metadatos aplicables a los documentos, crear una red de relaciones que aporte especificación y fiabilidad , posibilitar el trabajo cooperativo al funcionar como soporte de conocimiento entre las organizaciones , permitirá la integración de diferentes perspectivas de usuarios , interoperatividad entre sistemas distintos , construcción automatizada de mapas conceptuales y temáticos , tratamiento ponderado del conocimiento para recuperar información de forma automatizada ,establecer modelos normativos que permitan la creación de la semántica de un sistema y un modelo para poder extenderlo y transformarlo entre diferentes contextos ,servir de base para la construcción de lenguajes de representación del conocimiento , en definitiva mejorar la recuperación y organización de la información como así también compartir conocimiento con la comunidad de usuarios.

Se trata entonces de concientizar a los bibliotecarios de los nuevos desafíos que implican la utilización de estas herramientas y tecnologías de la web semántica en las implementaciones de las bibliotecas digitales para brindar servicio más eficientes a nuestros usuarios.

Bibliografía

Bainbridge, D., Buchanan, G. , McPherson,J. , Jones,S., Mahoui, J. and Witten,I. (2001). Greenstone: A platform for distributed digital library applications. In European Conference of Digital Libraries, p. 137–148, Darmstadt, Germany

Berners-Lee, T. , Hendler, J. , Lassila, O.(2001). The Semantic Web". *Scientific American*, May 2001.Recuperado el 18 de septiembre de 2009 de <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&pageNumber=1&catID=2>

- Berners-Lee, T.(2000). Semantic Web - XML.Recuperado el 18 de septiembre de 2009 de Slides <http://www.w3.org/2000/Talks/1206-xml2k-tbl>
- Berners-Lee, T.(1998). What the Semantic Web can represent.Recuperado el 18 de septiembre de 2009 de <http://www.w3.org/DesignIssues/RDFnot.html>
- Fuhr, N. et al (2007).Evaluation of digital libraries.Int J Digit Libr 8, 21–38
- Hinze1,A. , Buchanan, G. , Bainbridge1, D. , Witten, I. H.(2009).Greenstone: a platform for semantic digital libraries.En. Semantic Digital Libraries Sebastian Ryszard Kruk an Bill McDaniel (eds).Springer, p. 163
- IFLANET(1998) . Functional requirements for bibliographic records .Recuperado el 20 de septiembre de 2009 de <http://www.ifla.org/VII/s13/frbr/frbr.pdf>
- Jacob, E.K.(2003). Ontologies and the Semantic Web.Bulletin of the American Society for Information Science and Technology, 29 (4),19-22.Disponible en línea, recuperado el 22 de septiembre de 2009, de <http://www.asis.org/Bulletin/Apr-03/BulletinAprMay03.pdf>
- Kruk, S.R. , Woroniecki,T. , Gzella, A., Dabrowski,M.(2009) .JeromeDL - The Social Semantic Digital Library .En:Semantic Digital Libraries Sebastian Ryszard Kruk an Bill McDaniel (eds).Springer , pág. 139
- Méndez, E (2009).Introducción a la web semántica: vocabularios, tecnologías y estándares.Curso organizado por el SIU ,29-30 de octubre.
- Méndez, E. (2004).La Web semántica, una web 'más bibliotecaria". Boletín de la SEDIC, nº 41.Recuperado 2 de noviembre de 2009 de http://www.sedic.es/p_boletinclip41_confirma.htm
- W3C (2004b1). RDF Primer. W3C Recommendation 10 February 2004. Editores: F. Manola y E. Miller. Recuperado el 2 de noviembre de 2009, de <http://www.w3.org/TR/rdf-primer/>