

Administración de Repositorios Institucionales Híbridos de Acceso Abierto

Luis E. ROSAS

Facultad de Telemática, Universidad de Colima
Colima, Colima, CP 28040, México

María ANDRADE

Facultad de Telemática, Universidad de Colima
Colima, Colima, CP 28040, México

Nicandro FARIÁS

Facultad de Ingeniería Mecánica y Eléctrica, Universidad de Colima
Coquimatlán, Colima, CP 28400, México

Stalin X. CARAGUAY

Facultad de Telemática, Universidad de Colima
Colima, Colima, CP 28040, México

RESUMEN

Para las instituciones educativas y científicas, es muy importante mantener una buena comunicación con otras instituciones o sistemas, con la finalidad de compartir, generar o renovar conocimientos. En el presente trabajo se propone un esquema para la administración de Repositorios Institucionales híbridos de acceso abierto en ambientes operativos y de ejecución abiertos, a través de la aplicación del protocolo OAI-PMH. La propuesta está sustentada en la utilización de métodos de cosecha y recolección de metadatos para la difusión y transferencia de información, bajo contextos abiertos. Con este sistema se espera facilitar una interacción permanente entre las instituciones educativas con el fin de compartir, generar, renovar o evaluar los conocimientos. De la misma manera es posible ofrecer al usuario una estrategia de acceso a un amplio acervo de información variada, confiable y actualizada que le sirva de soporte en sus tareas de investigación.

Palabras clave: Repositorios Institucionales, metadatos, contextos abiertos, protocolo OAI-PMH.

1. INTRODUCCIÓN

En los últimos años, el rápido avance de la tecnología, ha originado la proliferación de fuentes de información digital. Los revolucionarios sistemas de cómputo y la inesperada velocidad de expansión de las redes de computadoras han facilitado el procesamiento, la distribución y la explotación de este tipo de información. Día a día, nuestras actividades se involucran más con la tecnología, por lo que el intercambio de información se ha convertido en una necesidad primaria en muchos sectores.

La facilidad con que ahora se tiene acceso a las redes de computadoras y el fuerte impulso que le dio la web a Internet, han sido las principales causas de esta enorme avalancha de información digital, utilizada por los millones de usuarios que demandan tanto almacenamiento como recuperación de información.

Esta evolución informática ha provocado un cambio en los conceptos tradicionales que se tenían acerca de los libros, bibliotecas, investigación y aprendizaje. No es difícil imaginarse que en un futuro toda la información será procesada y adquirida a través de medios electrónicos. El libro tradicional será complementado, y posiblemente reemplazado por alguna nueva tecnología que permita al lector un mejor aprendizaje, o simplemente, una lectura más amena. La introducción de las computadoras en el ámbito educativo de todos niveles y el importante auge que ha tenido la educación a distancia -con investigadores, profesores y alumnos distribuidos geográficamente- hace inminente la necesidad de tener acceso a espacios electrónicos de información ordenada, distribuida, común y compartida.

López [1] considera que las bibliotecas que han sido automatizadas están siendo transformadas o complementadas implementando bibliotecas digitales, bajo el concepto general de recopilar, almacenar y organizar información de manera digital. Esto con el propósito de realizar búsquedas, recuperaciones y procesamientos vía las redes de cómputo; todo bajo un ambiente sencillo para el usuario y tomando en cuenta factores fundamentales como la presentación y la

representación de la información, los mecanismos de almacenamiento y recuperación, la interacción humano-computadora, la plataforma tecnológica y el ancho de banda de la red.

Con estos cambios, surge el concepto de metadato, el cual es un elemento que describe el contenido, condiciones y características de un documento, con el fin de definir, identificar, organizar, indizar, filtrar, colocar, preservar, recuperar y administrar ese documento como parte de un conjunto ordenado de recursos de información, ya sean electrónicos o no.

Martínez y Navarra [2] consideran que los metadatos son un conjunto de elementos que poseen una semántica comúnmente aceptada, que surgieron de la necesidad de recuperar la información electrónica dispersa.

Para Voutssás [3] los metadatos pueden crearse para describir los atributos inherentes de un recurso documental cualquiera que éste sea: objeto bibliográfico (libro revista, tesis), registros e inventarios archivísticos, entre los más comunes.

Para generar metadatos es necesario apearse a algún estándar de metadatos bien definido para lograr una homogeneidad y estructura organizada en la información que se transmite [4]. Dublin Core es uno de los estándares de metadatos de mayor aplicación y que ha dado lugar incluso a otros formatos. El modelo de metadatos Dublin Core es un esfuerzo internacional e interdisciplinar abocado a definir el conjunto de elementos básicos para describir los recursos electrónicos y facilitar su recuperación [5]. Este estándar consta de quince campos subdivididos en tres grupos: Contenido, Propiedad intelectual y Aplicación, los cuales se encuentran resumidos en la Tabla 1.

Tabla 1. Estructura de los metadatos Dublin Core.

CONTENIDO	PROPIEDAD INTELECTUAL	APLICACIÓN
<i>Título (Title).</i>	<i>Creador (Creator).</i>	<i>Fecha (Date).</i>
<i>Materia (Subject).</i>	<i>Editor (Publisher).</i>	<i>Formato (Format).</i>
<i>Descripción (Description).</i>	<i>Colaborador (Contributor).</i>	<i>Identificador (Identifier).</i>
<i>Fuente (Source).</i>	<i>Derechos (Rights).</i>	<i>Tipo (Type).</i>
<i>Relación (Relation).</i>		<i>Lenguaje (Language).</i>
<i>Cobertura (Coverage).</i>		

El protocolo para la transmisión de contenidos en Internet OAI-PMH consiste en una interfaz sencilla que se utiliza para acceder a la información bibliográfica disponible en un archivo o repositorio. Se creó con la misión de desarrollar y promover estándares de interoperabilidad para facilitar la difusión eficiente de contenidos en Internet [6]. Utiliza transacciones HTTP para emitir preguntas y obtener respuestas entre un servidor o archivo y un cliente o servicio recolector de metadatos. El cliente puede pedir al servidor que le envíe metadatos según determinados criterios, como por ejemplo la fecha de creación de los datos. En respuesta, el servidor devuelve un conjunto de registros en formato XML, incluyendo identificadores (URL's por ejemplo) de los objetos descritos en cada registro.

2. TRABAJOS RELACIONADOS

En su investigación [7] Santillan demuestra la importancia y utilidad de los Repositorios Institucionales en el proceso de divulgación de información, especialmente a través de bibliotecas digitales, basándose como caso de estudio en el proyecto E-LIS (E-prints in Librarianship and Information Science). El autor afirma que las universidades, por su carácter de productoras y consumidoras de información, son las más interesadas en el desarrollo y florecimiento del acceso abierto y prueba de ello es el reciente aumento en el desarrollo de Repositorios Institucionales por parte de las mismas, como se refleja en el Registry of Open Access Repositories (ROAR) y cuyo comportamiento se muestra en la figura 1.

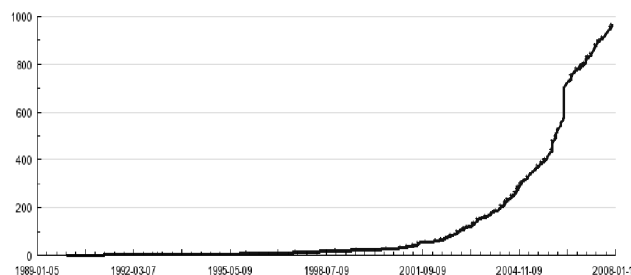


Figura 1. Incremento de Repositorios Institucionales [8].

Santillán menciona también que estos repositorios son la prueba fehaciente del impacto positivo del OA (Open Access) como alternativa al sistema tradicional de comunicación científica y como vía para la integración de los países en desarrollo al circuito mundial de la información y el conocimiento

Chawner [9] utiliza un modelo para la clasificación del tipo de software utilizado en la implementación de Repositorios Institucionales y bibliotecas digitales. Obtuvo como resultado que la mayoría del software libre enfocado a la operación de este tipo de servicios está siendo implementados a muy pequeña escala en comparación con otros proyectos de desarrollo de software llevados a cabo por bibliotecas privadas más

grandes que a menudo reciben apoyos de organismos de investigación. Las estadísticas obtenidas por la autora, evidencian la necesidad de desarrollar aplicaciones para Repositorios Institucionales, pero sin embargo no propone ningún método o proyecto para llevarlo a cabo.

Nikolov y Stoehr [10] exponen algunas estadísticas que evidencian la “muerte” de millones de artículos científicos en internet, debido a que muchos autores no almacenan sus trabajos en Repositorios Institucionales, sino que lo hacen en páginas personales o servidores públicos donde corren el riesgo de perderse debido a la constante inestabilidad de las tecnologías utilizadas. Una de sus estadísticas más relevante, expone que cerca del 11% de las URL’s de los documentos PDF que contienen referencias a las publicaciones de ciencias de la vida, no eran accesibles dentro de los 5 meses después de haber sido cosechados. Al mismo tiempo proponen la cosecha de metadatos como una forma de abordar este problema y como factor crucial para preservar la literatura científica en línea.

Sin embargo, los autores mencionados únicamente se enfocan en la problemática y las estadísticas y no proponen algún método o esquema para implementar las cosechas que exponen. Situación que en este trabajo se plantea a través de la metodología y el esquema propuestos más adelante.

3. DESARROLLO DEL TEMA

Hoy en día cada institución educativa y científica requiere mantenerse en comunicación con otras instituciones o dependencias, con la finalidad de compartir, generar o renovar conocimientos.

Los investigadores, por ejemplo, al momento de desarrollar proyectos se ven obligados a consultar información de diferentes fuentes (libros, revistas, tesis, memorias, etc.); este proceso de consulta resulta fallido en algunos casos ya que el acervo inverosímil que transita en Internet expone, en su mayoría, suposiciones no comprobadas como datos importantes.

En contraste con lo mencionado anteriormente, el hecho de apoyarse en material científico respaldado por investigadores reconocidos garantiza que las ideas ahí expuestas han sido probadas y sus resultados y conclusiones son reales.

Dado a lo expuesto antes, resulta de gran importancia una colaboración constante entre instituciones y dependencias para compartir fuentes de información que incrementen el acervo bibliográfico y así dar un soporte significativo a los investigadores que buscan referencias bibliográficas para dar fundamento a sus proyectos.

El objetivo general de este trabajo es desarrollar un esquema de administración de Repositorios Institucionales híbridos que sea capaz de ofrecer,

intercambiar y difundir material bibliográfico en Internet con la participación de diferentes instituciones educativas y de investigación de forma libre y abierta, logrando con ello una interacción y retroalimentación permanente entre dichas instituciones así como una mayor disponibilidad de la información y un fácil acceso a la misma. El procedimiento para lograr el propósito es:

- Pactar convenios con dependencias o centros de investigación que estén interesados en participar en la iniciativa de archivos abiertos.
- Implementar un módulo para la conectividad entre Repositorios Institucionales y bibliotecas digitales en la Web.
- Representar la información del material bibliográfico a transmitir en forma de metadatos Dublin Core a través de XML.
- Implantar un módulo en PHP para cosechar y exportar metadatos entre los Repositorios Institucionales y bibliotecas digitales implicadas.
- Desarrollar un buscador para localizar documentos bibliográficos de manera rápida y eficiente.

Para lograr los objetivos buscados se utilizan métodos de intercambio de información a través de internet enfocándose específicamente en cosecha y recolección de metadatos estructurados en XML bajo el estándar de generación de metadatos Dublin Core y apegándose en todo momento a la aplicación del protocolo OAI-PMH. En la figura 2 se muestra el diagrama conceptual de este proyecto y su descripción se hace en los párrafos siguientes.



Figura2. Diagrama conceptual del esquema propuesto.

Interfaz

Uno de los factores más importantes a tomar en cuenta en el desarrollo de un sistema de información son las interfaces, las cuales deben ser amigables y atractivas hacia el usuario para que éste se interese en el sistema y realice sus actividades con interés y de forma intuitiva. Si una interfaz es relativamente compleja, el usuario puede rehusarse a usarla por temor a equivocarse o simplemente provocará errores debido a la confusión que la interfaz le cause. A lo largo de la historia de la informática han ido apareciendo diferentes tipos y estilos de interfaz de usuario dependiendo, sobre todo, de la capacidad gráfica de los equipos computacionales [11].

La interfaz de usuario es en la mayoría de los casos el componente más crítico del sistema. Los usuarios generalmente no entienden sobre el mundo interno de las computadoras, que se compone de bits, bytes, ficheros, circuitos, entre otros. Es más, conocen el sistema por medio de su interfaz, el texto, las imágenes o los sonidos que aparecen en los dispositivos de salida del sistema en cuestión (pantallas, altavoces, etc.). Los usuarios sólo son capaces de explotar las posibilidades que la tecnología ofrece si sus interfaces transmiten dichas posibilidades [12].

En esta propuesta, la interfaz de usuario pretende ser relativamente simple pero sin descuidar con ello el potencial que el sistema puede tener. El usuario se encontrará ante una interfaz amigable capaz de ofrecerle el material bibliográfico que busca de manera intuitiva y simple. Así mismo, la interfaz también le ofrece al usuario la posibilidad de realizar búsquedas avanzadas con la finalidad de obtener resultados más específicos en cuanto al tipo, enfoque, área de conocimiento, etc. del material buscado.

Almacenamiento

La cuestión del almacenamiento es crucial para el éxito del proyecto. Es en este bloque donde se almacenan los metadatos de todo el material bibliográfico disponible en sus respectivos repositorios. Un repositorio no es más que el lugar donde reside una determinada información utilizada para fines operativos. Dicha información puede ser de cualquier tipo, por ejemplo de clientes, productos o como en este caso, de material bibliográfico [13]. La función de los repositorios en el presente proyecto reside en almacenar los metadatos de todo el material disponible para consulta manteniéndolos siempre ordenados y clasificados. Los repositorios de metadatos se mantendrán siempre actualizados apoyándonos en los siguientes bloques de Interoperabilidad y Comunicación para hacer frecuentemente cosechas de nuevos metadatos de material bibliográfico en repositorios externos de la Web y actualizándolos en los repositorios locales.

Interoperabilidad

La interoperabilidad semántica o interoperabilidad de los metadatos está destinada a la descripción de los recursos

de información para facilitar el intercambio de información y la recuperación óptima por parte de los usuarios. En este ámbito se suelen utilizar un conjunto de herramientas para la representación del conocimiento contenido en los recursos, generalmente provenientes del ámbito de la documentación y la gestión del conocimiento, como son: los vocabularios controlados, los metadatos, las ontologías y los topic maps.

Los metadatos son un conjunto de elementos que poseen una semántica comúnmente aceptada, ya que surgieron por la necesidad de recuperar la información electrónica dispersa [14]. XML (Lenguaje de anotación extensible) ofrece un formato para la descripción de datos estructurados. Es decir, XML es un metalenguaje, dado que con él podemos definir nuestro propio lenguaje de presentación y, a diferencia del HTML, que se centra en la representación de la información, XML se centra en la información en sí misma. La particularidad más importante del XML es que no posee etiquetas prefijadas con anterioridad, ya que es el propio diseñador el que las crea a su criterio, dependiendo del contenido del documento [15]. En la figura 3 se muestra un ejemplo de un metadato con etiquetas DublinCore estructurado en XML.

```
<?xml versión="1.0"?>
<!DOCTYPE MENSAJE SYSTEM "metadato.dtd">
<metadato>
  <contenido>
    <título> Administración de Repositorios
    Institucionales Híbridos de Acceso Abierto </título>
    <descripción> Propone un esquema para la
    administración de Repositorios Institucionales Híbridos
    de Acceso Abierto a través de la aplicación del
    protocolo OAI-PMH </descripción>
  </contenido>
  <propiedad intelectual>
    <creador> Rosas G., Andrade A., Farías
    M.</creador>
    <colaborador>Universidad de Colima
  </colaborador>
  </propiedad intelectual>
  <aplicación>
    <fecha> Enero 24, 2011 </fecha>
    <lenguaje> Es </lenguaje>
  </aplicación>
</metadato>
```

Figura 3. Ejemplo de un metadato estructurado en XML.

Los principales objetivos que se buscan con la utilización de XML como lenguaje de marcado son [16]:

- Representar y distribuir tanto documentos como información textual.
- Intercambio de datos e información estructurada a través de Internet y WWW.

- Integración de datos procedentes de fuentes heterogéneas.
- Eliminar la barrera entre información estructurada e información textual.

Comunicación

La comunicación implica la recíproca relación con mensajes, es decir, con contenidos de comunicación que representan precisamente a la información; en otras palabras, la información es el contenido de la comunicación. A través del bloque de comunicación el sistema se mantendrá en contacto constante con otras instituciones educativas a través de la Web para compartir e intercambiar información bibliográfica confiable de manera abierta y libre.

Al enlazar cada uno de los bloques mencionados queda conformado el esquema propuesto para administrar Repositorios Institucionales híbridos en Internet bajo contextos abiertos, logrando nuestro propósito de ofrecer, intercambiar y difundir material bibliográfico en internet a través de diferentes Repositorios de Instituciones o dependencias educativas y científicas.

Actualmente existen varios estándares para la generación de metadatos, sin embargo, se optó por el uso del Dublin Core por ser éste uno de los más flexibles y relativamente fáciles de implementar.

4. CONCLUSIONES

El tema de los Repositorios Institucionales, aunque es un concepto relativamente reciente, está marcando nuevas tendencias y adquiere cada vez mayor fuerza conforme al continuo avance y desarrollo de Internet.

En este artículo se describe una propuesta para diseñar e implementar un esquema de administración de Repositorios Institucionales capaz de ofrecer, intercambiar y difundir material bibliográfico entre diferentes repositorios en Internet utilizando la filosofía de contextos abiertos del protocolo OAI-PMH.

Los resultados preliminares manifiestan la importancia de la colaboración entre instituciones y dependencias para compartir fuentes de información que incrementen el acervo bibliográfico y así dar un soporte significativo a los investigadores que buscan referencias bibliográficas para dar fundamento a sus proyectos.

Para continuar con la investigación sugerimos trasladar la operatividad de nuestro repositorio *Green Access* hacia un repositorio *Golden Acces*, que incluya referencias de editoriales y bibliotecas digitales de reconocido prestigio, tales como del repositorio de REDALYC.

5. REFERENCIAS

[1] López, C. (2000) *Modelo para el Desarrollo de Bibliotecas Digitales Especializadas*. Tesis de

maestría en el Instituto Tecnológico Autónomo de México. Recuperado el 25 de mayo del 2009, de <http://www.bibliodgsca.unam.mx/tesis/tes7c1lg/tes7c1lg.htm>.

- [2] Martínez, J. Á., Navarra, P. L. (2007). *La interoperabilidad de la información*. Barcelona: Editorial UOC.
- [3] Voutssás Márquez, J. (2005). *Bibliotecas y publicaciones digitales*. México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas.
- [4] IEEE. (2010). IEEE draft standard for learning object metadata. Learning Technology Standards Committee. Recuperado el 8 de octubre del 2010, de <http://ieeeltsc.org/>
- [5] DCMI (2010). The Dublin Core Metadata Initiative. Recuperado el 7 de octubre del 2010, de <http://dublincore.org/>
- [6] OAI (2008). Open Archives Initiative, Standards for Web Content Interoperability. Recuperado el 22 de mayo del 2009, de la red mundial <http://www.openarchives.org>
- [7] Santillan, J. (2009). El movimiento del acceso abierto y el mundo bibliotecario desde la experiencia del proyecto E-LIS. Revista AIBDA, Vol. XXX, N° 1-2, Doc. 2, enero-diciembre 2009.
- [8] ROAR (Registry of Open Access Repositories) <http://roar.eprints.org/>
- [9] Chawner, B. (2005). F/OSS in the Library World: An Exploration. International Conference on Software Engineering. Proceedings of the fifth workshop on Open source software engineering, ACM. 1-4.
- [10] Nikolov, N., Stoehr, P. (2008). Harvesting Needed To Maintain Scientific Literature Online. International Conference on Digital Libraries Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries, ACM. 460-460.
- [11] F. Alonso, L. Martínez, F. J. Segovia, Introducción a la Ingeniería del software – Modelos de desarrollo de programas, España: Delta publicaciones, 2005]
- [12] T. G. Saltiveri, J. L. Vidal, J. J. Cañas, Diseño de sistemas interactivos centrados en el usuario, Barcelona: Editorial UOC, 2005.
- [13] H. L. Croxatto, Creando valor en la relación con sus clientes – como desarrollar todo el potencial de una solución CRM y transformarlo en ventajas de negocio, Argentina: Editorial dunken, 2005.
- [14] J. Á. Martínez, P. L. Navarra, La interoperabilidad de la información, Barcelona: Editorial UOC, 2007.
- [15] M.C. Arellano, J. Nogales, B. Martín, La organización hipertextual del ordenamiento jurídico - Posibilidades de XML y estándares relacionados. Revista General de Información y Documentación, 182 – 189. Recuperado de la red mundial el 25 de mayo del 2009: <http://revistas.ucm.es/byd/11321873/articulos/RGID0303220181A.PDF>.
- [16] Metadatos y documentos XML/RDF para Recuperación y organización de la información.

Recuperado de la red mundial <http://metadatos-xml-rdf.awardspace.com/xml.html> el 22 de abril de 2010.