

ARTÍCULO

UN PORTAL DE ACCESO ABIERTO A LA LITERATURA CIENTÍFICA EN IBEROAMÉRICA

Guillermo Chávez Sánchez y Miguel Ángel Ortiz Camilo

Un portal de acceso abierto a la literatura científica en Iberoamérica

La producción de revistas académicas en Iberoamérica

En la década de los 90 del siglo pasado, época en que la tecnología y las telecomunicaciones avanzaban conjuntamente de forma vertiginosa, un grupo de académicos universitarios, en colaboración con otros especialistas de la información de Brasil, Cuba y Venezuela, analizaron e identificaron la pertinencia de contar con un sistema en línea que facilitara a la comunidad iberoamericana integrar esfuerzos en materia de producción, difusión, registro y uso de las revistas académicas producidas en la región, y al mismo tiempo, contribuir a reforzar y elevar la calidad, el impacto y la visibilidad de estas publicaciones.

El sistema referido, fue concebido en el año de 1997 con el nombre de LATINDEX "Sistema Regional de Información para las Revistas Científicas de América Latina, el Caribe, España y Portugal", cuyo objetivo principal fue difundir, hacer accesible y elevar la calidad de las revistas académicas editadas en la región a través del trabajo compartido.

Desde entonces, gracias al esfuerzo compartido de veinte instituciones educativas y de investigación con sede en veinte países iberoamericanos (Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, España, Francia, Guatemala, Honduras, Nicaragua, Panamá, Perú, Portugal, Puerto Rico, República Dominicana, Venezuela y México), hasta el día de hoy, el sistema LATINDEX ofrece tres productos:

- *El Directorio*. Base de datos normalizada con información catalográfica de casi 21,000 revistas (trayectoria, especialización, editorial, responsables, dirección, distribución, precio, entre otros).
- *El Catálogo*. Base de datos con más de 6,300 registros de revistas que cumplen con un umbral de los criterios de calidad editorial propuestos por los miembros representantes de los países socios.
- *Enlace a revistas electrónicas*. Base de datos con las direcciones de sitios web en dónde se publican de forma electrónica más de 4,000 revistas.

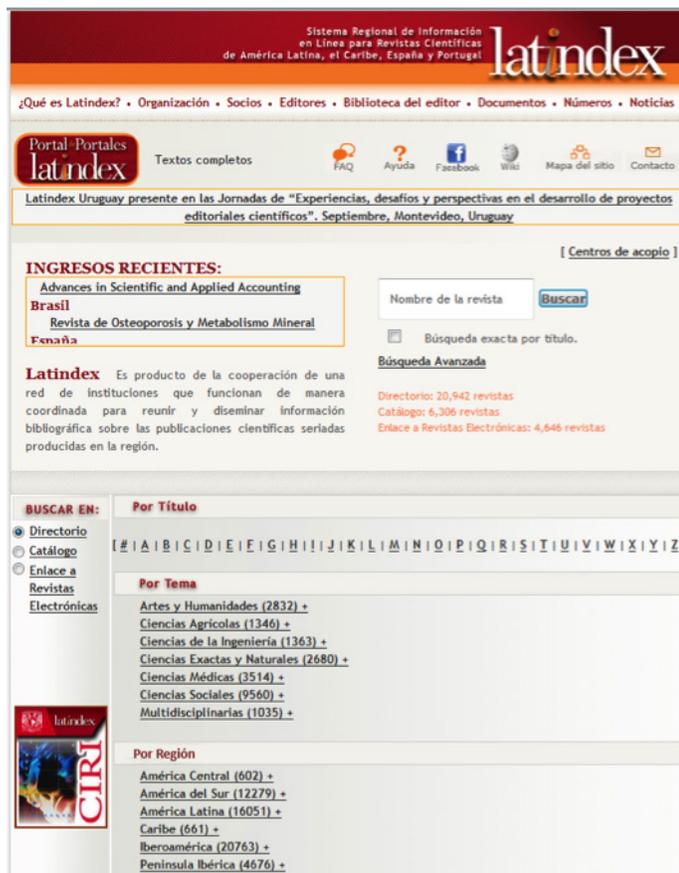


Figura 1. Página principal de LATINDEX (www.latindex.unam.mx)

De esta manera, gracias a la participación de los representantes de los países socios, ha sido posible la generación de estos productos, así como la promoción y práctica de actividades académicas para la profesionalización en materia de edición de revistas académicas. Por lo anterior, este sistema se ha posicionado como un referente para la consulta de información sobre la producción editorial en la región y la aplicación de buenas prácticas y estándares editoriales.

Acceso libre a contenidos de revistas académicas iberoamericanas

Hasta finales de 2010, los principales usuarios del sistema LATINDEX eran principalmente especialistas de la información y, hasta entonces, no se había ofrecido un producto de interés para el lector final (por ejemplo: profesores, investigadores o estudiantes), por lo cual, habría que generar y ofrecer nuevos productos.

Con el surgimiento de las iniciativas internacionales de Budapest (2002), Berlín y Bethesda (2003) sobre Acceso Abierto (Open Access¹), durante los últimos diez años, diversas instituciones educativas y de investigación han desarrollado y puesto en línea hemerotecas digitales de acceso abierto que permiten tanto la consulta de artículos a texto completo, como el intercambio de su información catalográfica a través de Internet.

Estas hemerotecas digitales, generalmente son sistemas de gestión documental de código abierto que incorporan el protocolo de interoperabilidad OAI-PMH (por sus siglas en inglés de Open Archives Initiative Protocol for Metadata Harvesting) el cual permite la cosecha y el intercambio de la información catalográfica de los artículos publicados (Metadatos), facilitando la indexación de los mismos por motores de búsqueda especializados como Google Scholar u otros sistemas recolectores de información (Harvesters).

Tomando en consideración lo anterior, se planeó el desarrollo de un sistema complementario al de LATINDEX, con el que se aprovecharan las posibilidades que ofrecen las hemerotecas digitales de acceso abierto, con la finalidad de garantizar un acceso libre, centralizado y organizado a los contenidos de las revistas académicas de la región. De esta manera, se podría contribuir a aumentar la visibilidad y difusión de la investigación científica en Iberoamérica, impulsando la disseminación y la generación de nuevo conocimiento. Con el apoyo financiero de CONACYT, en febrero de 2009 inició el desarrollo del proyecto PORTAL DE PORTALES LATINDEX “Acceso libre a contenidos de revistas académicas iberoamericanas”.



Figura 2. PORTAL DE PORTALES LATINDEX (www.latindex.ppl.unam.mx)

1 De acuerdo a la declaración de Budapest de 2002, por acceso abierto a la literatura científica erudita, entendemos su disponibilidad gratuita en Internet, para que cualquier usuario la pueda leer, descargar, copiar, distribuir o imprimir, con la posibilidad de buscar o enlazar al texto completo del artículo, recorrerlo para una indexación exhaustiva, usarlo como datos para software, o utilizarlo para cualquier otro propósito legal, sin otras barreras financieras, legales o técnicas distintas de la fundamental de acceder a la propia Internet. El único límite a la reproducción y distribución de los artículos publicados, y la única función del copyright en este marco, no puede ser otra que garantizar a los autores el control sobre la integridad de su trabajo y el derecho a ser acreditados y citados (Budapest Open Access Initiative 2002, traducida al español en <http://www.madrimasd.org/informacionidi/e-ciencia/documentos/default.asp>

El principal objetivo de PORTAL DE PORTALES LATINDEX fue ofrecer un portal de acceso organizado a los contenidos publicados en revistas académicas iberoamericanas, disponibles en otros sitios (bases de datos o hemerotecas virtuales), a nivel de texto completo, resúmenes, tablas de contenido o registros bibliográficos.

Actualmente, en el Portal se encuentran disponibles más de 1 millón 260 mil registros catalográficos de artículos almacenados en la base de datos global. Cada uno de ellos cuenta con un URL que posibilita la consulta de los artículos a texto completo en alguna de las 18 hemerotecas integradas.

El desarrollo del proyecto se basó en la iniciativa del Open Access, en la que cualquier usuario debería poder tener libre acceso a artículos académicos para algún propósito legal sin otra barrera que la de una conexión a Internet.

En el sistema del PORTAL DE PORTALES LATINDEX se recolectan los metadatos de los artículos publicados, en hemerotecas virtuales de libre acceso en Iberoamérica, que soportan el protocolo OAI-PMH. De esta manera, es posible conformar una base de datos global con la descripción catalográfica de los artículos con la siguiente información:

- títulos
- nombres de los autores
- palabras clave
- resúmenes
- nombre de la revista de origen
- entidad editora
- fecha de publicación
- tipo de documento
- formato
- dirección URL de consulta
- idioma
- derechos.

Así, a través de la dirección electrónica o URL, se ofrece al usuario la posibilidad de consultar el artículo digital o recurso en la hemeroteca en que se encuentre publicado.

De esta manera, la consulta de la base de datos se puede hacer a través de un motor de búsqueda que permite realizar exploraciones generales de un texto completo en todos los campos almacenados, o en su caso, realizar búsquedas más elaboradas en más de un rubro de la información.

Intercambio e interoperabilidad de datos(Open Harvester Systems)

Para el desarrollo del PORTAL DE PORTALES LATINDEX fue necesario considerar el uso y aplicación de herramientas tecnológicas que permitieran el libre intercambio y la interoperabilidad de los datos depositados en los repositorios.

Estas herramientas como mínimo deberían soportar:

- *Un cosechador de metadatos basado en el protocolo OAI-PMH.* Posibilita la recolección de metadatos de los repositorios que tengan implementado el protocolo OAI-PMH para guardarlos en una base de datos local.
- *La interacción con sistemas proveedores de datos.* Permitirá al sistema de información centralizado hacer accesibles los datos recolectados a otros proveedores de servicios.
- *La integración de un metabuscador.* Será el principal servicio de interacción con el usuario final, ya que permitirá hacer una búsqueda en los metadatos recolectados.

El Open Harvester Systems (OHS), desarrollado y mantenido por PKP (Public Knowledge Project), es un software gratuito de código abierto implementado en PHP y distribuido bajo la licencia GPL (General Public Licence), que ofrece la posibilidad de indexación, recolección y redistribución de metadatos. Es una herramienta adecuada para dar solución a la propuesta.

De forma general, el Open Harvester Systems cumple con las siguientes características (Figura 3):

- Base de datos relacional como medio de almacenamiento de metadatos.
- Ofrece soporte para los manejadores de bases de datos MySQL y Postgres.
- Posibilidad de cosechar metadatos OAI en distintos esquemas (Dublin Core no calificado, PKP Dublin Core extensions, MODS o MARCX).

- Interfaz de usuario sencilla y personalizable mediante CSS o plantillas HTML.
- Motor de búsqueda que permite realizar búsquedas avanzadas sobre los metadatos recolectados.
- Posibilidad de recolección granular (sólo partes específicas del repositorio).
- También es un proveedor de metadatos.

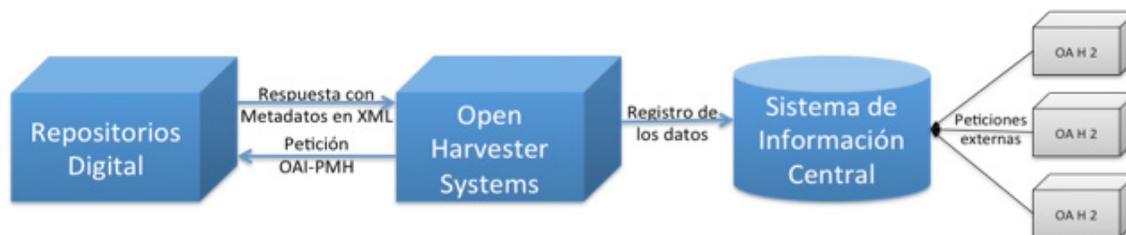


Figura 3. Diagrama general del OHS.

Arquitectura del Open Harvester Systems

La arquitectura del **Open Harvester Systems** se basa en un patrón Modelo Vista Controlador (MVC), en el que el controlador recibe las peticiones del usuario. La vista muestra al usuario los contenidos de una forma legible y ergonómica y el modelo almacena los contenidos en una base de datos. De esta manera, el usuario puede consultar los contenidos y los administradores pueden insertar, actualizar y eliminar datos (Figura 4).

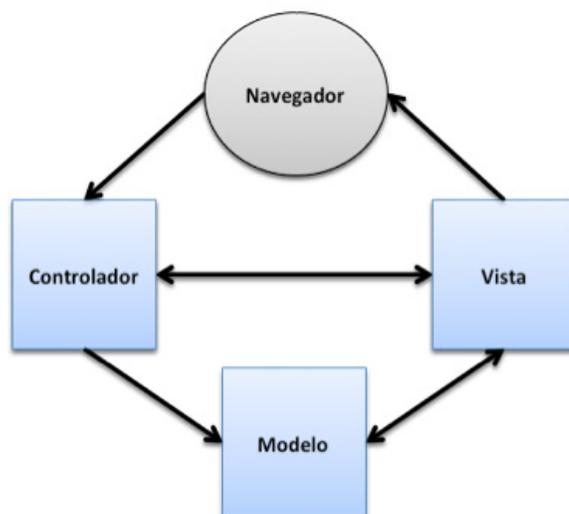


Figura 4. Modelo Vista Controlador del OHS.

La arquitectura de la plataforma tecnológica está definida por un modelo de capas donde se integran los siguientes sistemas y servicios:

- *Sistema Operativo Linux* (en su distribución CentOS en su versión 6.x. o superior). Este sistema operativo es la base para la puesta en producción del Open Archives Harvester 2, que permite garantizar una adecuada respuesta a la demanda de los recursos que en él operen.
- *Base de Datos MySQL* (versión 5.x. o superior). Es el medio con el que se almacenarán, gestionarán, editarán o eliminarán los datos recolectados y las referencias a los objetos digitales.
- *Servicio HTTP* (Apache HTTP versión 2.2.x. o superior). Servicio que permitirá ofrecer una interfaz gráfica de usuario basada en el uso de hipertexto y hojas de de estilo en cascada.
- *Intérprete PHP* (versión 5.x. o superior). Instalado en paralelo al servicio HTTP, este servicio se encarga de la ejecución de métodos y funciones para interpretar las peticiones del usuario.
- *Contenedor de servlets* (Apache Tomcat versión 6.0. o superior). Servicio necesarios para la instalación y uso de Lucene Solr.
- *Servicio de indexación* (Lucene Solr versión 3.x. o superior). Servicio para la indexación de los registros recolectados y depositados en la base de datos, simplificando y agilizando las búsquedas y recuperación de los datos.
- *Aplicación de correo electrónico* (Sendmail versión 8.x. o superior). Servicio para la comunicación de los administradores a través de correo electrónico.



Figura 5. La arquitectura en capas del Open Harvester Systems.

Conclusiones

EL PORTAL DE PORTALES, al igual que sistema LATINDEX, es el resultado de la colaboración y participación de un grupo de trabajo multidisciplinario en la UNAM, con la participación de personal de la Dirección General de Bibliotecas (DGB), del Instituto de Física, y de la Dirección General de Cómputo y Tecnologías de Información y Comunicación.

También, es importante destacar la colaboración de miembros de otras dependencias como el Public Knowledge Project (PKP) de la Universidad de Stanford, desarrolladores el **Open Harvester Systems**.

Sin lugar a dudas, el sistema LATINDEX y el PORTAL DE PORTALES LATINDEX enriquecen y complementan la amplia gama de productos y servicios tecnológicos en línea que ofrece la Universidad, principalmente a su comunidad académica, pero también al público en general. Ambos, en conjunto con otros sistemas de información como Humanindex (Base de Datos Bibliográfica de Humanidades y Ciencias Sociales), CLASE y PERIÓDICA (Citas Latinoamericanas en Ciencias Sociales y Humanidades e Índice de Revistas Latinoamericanas en Ciencias) o la RAD-UNAM (Red de Acervos Digitales de la UNAM), permite a los estudiantes, profesores o investigadores localizar y acceder a la información académica producida en la región desde diferentes alternativas, enfoques o posibilidades. De igual manera, con estos sistemas también se contribuye a incrementar la visibilidad de la producción editorial en la UNAM relacionada con revistas académicas, ya que en ellos se integran los contenidos del Portal SciELO México y del Portal de Revistas Científicas y Arbitradas de la UNAM.

Todos estos sistemas, representan un ejemplo del esfuerzo y del trabajo constante de la comunidad universitaria para ofrecer productos innovadores con el uso y aprovechamiento de las Tecnología de Información y Comunicación. En su conjunto, conforman una parte esencial para contribuir con la digitalización de las tareas sustantivas de la Universidad (docencia, investigación y difusión de la cultura).

De igual manera, estas propuesta están orientadas a la promoción, adopción e institucionalización de las iniciativas de Acceso Abierto, lo que garantiza el libre acceso a la producción cultural y científica que se produce en la UNAM, y de esta manera se cumple con las declaraciones de Berlín firmadas en 2006 por el entonces Rector Juan Ramón de la Fuente, tal como lo han hecho otras universidades del mundo.

Tanto el sistema LATINDEX, como el PORTAL DE PORTALES LATINDEX son productos que pueden apoyar al programa institucional de Toda la UNAM en Línea, ya que ofrecen elementos adicionales que facilitan la consulta y la recuperación digital de la literatura científica de la Universidad. Esto también representa un mayor impacto de la producción académica, al

incrementar la visibilidad de su producción científica y mejorar su posicionamiento en el mundo a través del uso de tecnologías como el Internet.

Referencias

Apache Solr, “Lucene Solr” [en línea]. [Consultada: 01 de octubre de 2012] Disponible en Internet: [<http://lucene.apache.org/solr/>]

Open Society Institute, “Budapest Open Access Initiative” [en línea]. [Consultada: 01 de octubre de 2012] Disponible en Internet: [<http://www.soros.org/openaccess/read>]

Public Knowledge Project (PKP), “Open Harvester Systems” [en línea]. [Consultada: 01 de octubre de 2012] Disponible en Internet: [<http://pkp.sfu.ca/?q=harvester>]

SOMMERVILLE Ian, “Ingeniería de Software” 9a. Ed. Pearson Educación.

WIKIPEDIA La enciclopedia libre, “Web semántica” [en línea]. [Consultada: 01 de octubre de 2012] Disponible en Internet: [http://es.wikipedia.org/wiki/Web_sem%C3%A1ntica]

WIKIPEDIA La enciclopedia libre, “Minería de datos” [en línea]. [Consultada: 01 de octubre de 2012] Disponible en Internet: [http://es.wikipedia.org/wiki/Miner%C3%ADa_de_datos]

WIKIPEDIA La enciclopedia libre, “Big data” [en línea]. [Consultada: 01 de octubre de 2012] Disponible en Internet: [http://es.wikipedia.org/wiki/Big_data]