

Modernización de la difusión de los Congresos de Ingeniería de Organización

Modernization of the spread of the Congresses of Industrial Engineering

Aparicio Ruiz P¹, Guadix Martín J¹, Onieva Giménez L¹, Arango Pastrana CA¹

Abstract In recent years, the Congress of Industrial Engineering has promoted approaches to analysis and design features of this engineering, to develop sciences and the related methods to make an excellent promotion of this discipline. The growth and success of previous conferences has not been accompanied by the necessary dissemination on the Internet. The inescapable need for indexing in the current search systems and compliance with the OAI-PMH repository for papers that are found on the website of the Association for the Development of Industrial Engineering (ADINGOR) has led to a modernization of the current systems.

Resumen En los últimos años el Congreso de Ingeniería de Organización ha promovido los enfoques de análisis y diseño característicos de esta Ingeniería, desarrollando las ciencias y los métodos relacionados y promoviendo de manera excelente esta disciplina. El crecimiento y éxito de los últimos congresos no ha ido acompañado de la difusión necesaria en internet. La ineludible necesidad de indexación en los sistemas de búsqueda actuales y el cumplimiento del protocolo OAI-PMH para el repositorio de artículos que se encuentra en la web de la Asociación para el Desarrollo de la Ingeniería de Organización (ADINGOR) ha generado una modernización de los sistemas actuales.

Keywords: OAI-PMH, indexing, congress

Palabras clave: OAI-PMH, indexación, congresos

¹ Pablo Aparicio Ruiz(✉), José Guadix Martín, Luis Onieva Giménez, Carlos Alberto Arango Pastrana

Grupo de Ingeniería de Organización. Escuela Superior de Ingeniería, Universidad de Sevilla, C/Camino de los Descubrimientos,S/N 41092 Sevilla, Spain
e-mail: pabloaparicio@us.es, guadix@esi.us.es, onieva@us.es, cap@esi.us.es

1 Introducción

En todas las disciplinas de las ciencias o las humanidades hay publicaciones y personas que contribuyen a dar forma a esa disciplina, los congresos de ingeniería de organización son sin duda un lugar de referencia en el desarrollo de esta disciplina.

Las publicaciones son reconocidas por su significativa contribución y sus autores son muy apreciados por nuestra comunidad científica, sin embargo, es necesario que los miembros de otras disciplinas sigan nuestras publicaciones, generando un mayor impacto de esta importante fuente de conocimiento, al hacer referencia a ellas en el desarrollo de sus propias contribuciones.

En las colecciones de publicaciones que existen en línea, es importante alcanzar un mayor impacto debido a la importancia que este tiene para las carreras de nuestros profesionales y de la propia disciplina, por esta razón es necesario obtener la mayor difusión de este congreso, y la modernización de su sistema de difusión.

En el área de las publicaciones académicas se presenta un movimiento tecnológico conocido como acceso abierto (open access), este genera las condiciones necesarias para la interoperabilidad de las publicaciones en la red, a la hora de descargar, leer, buscar o distribuir textos académicos.

La acción de compartir comunicaciones no solo puede aumentar la productividad e impacto de los investigadores sino también puede ser la llave para establecer colaboraciones con otros grupos de investigación, aumentando las posibilidades de generar nuevas publicaciones, por lo que se ha desarrollado un sistema que permite la indexación en el sistema de mayor difusión académica que no cumple los estándares de calidad, Google Académico, y tres bases de datos que tratan la información con mayor rigor, como son Scopus y Web of Science [WoS] (Ortega, 2007), o Dialnet como versión en castellano.

Esta comunicación tiene como objetivo divulgar entre los profesionales un nuevo protocolo para la transmisión de contenidos en Internet denominado OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting). Desde sus inicios en 1999, existe un creciente interés por la divulgación y uso de este protocolo. Su uso permite adherirse a múltiples sistemas de información existentes mediante el registro en sus repositorios de los metadatos necesarios, el acceso a éstos es posible a través del protocolo OAI-PMH. Alrededor del 70% de los repositorios utilizan el protocolo OAI-PMH y el restante 30% protocolos de otro uso. Se estima que el 82% de todos los documentos están en Inglés. El 18% restante se publican en otros idiomas, especialmente francés, alemán y español. Los metadatos no incluyen el texto completo del documento, pero puede contener su referencia. Fuera de todos los repositorios OAI-PMH, alrededor del 30% contienen una URL de referencia del documento, permitiendo su descargar mediante una solicitud de Web estándar. Para Barrueco y Subirats (2003) la “OAI-PMH está llamado a ser a las bibliotecas digitales lo que HTTP es hoy al web”.

Así mismo, existe un estándar para la información y documentación de un conjunto de metadatos Dublin Core. Este estándar permite que la estructura de la web se elabore pensando en el usuario final, permitiendo que la navegación sea ágil. El documento presenta la tecnología aplicada, las posibles alternativas, el desarrollo del proyecto, y finalmente, una visión del comportamiento y estado actual del sistema.

2 OAI-PMH

El protocolo OAI-PMH se creó para la extracción normalizada de metadatos de los archivos electrónicos, asegurando con ello la infraestructura para la preservación y la interoperabilidad o compatibilidad entre archivos, independientemente de las organizaciones a las que pertenezcan o de donde se encuentren.

Open Archives Initiative (OAI), hace referencia a la iniciativa que permite desarrollar sistemas para el depósito de documentos científicos completos, mediante una vista abierta (gratuita o no) de la arquitectura del sistema. Este marco de desarrollo permite el intercambio de múltiples formatos bibliográficos entre distintas máquinas utilizando un protocolo común, mediante la comunicación de metadatos sobre cualquier material almacenado en soporte electrónico.

2.1 Alternativas

(Crow, 2004) presento “A Guide to Institutional Repository Software”, una guía de software para repositorios institucionales, elaborada por el Open Society Institute para ayudar, a aquellas organizaciones que deseen utilizar algunas de las herramientas disponibles según se ajuste a las necesidades de su proyecto de repositorio. En ella se analizan un total de nueve sistemas que satisfacen tres criterios básicos: cumplir la última versión del protocolo OAI-PMH (v2.0), ser de código abierto, y estar disponibles de forma pública y gratuita para su descarga e implementación. En (Teresa Silió, 2005) podemos encontrar algunos ejemplos de proveedores de datos OAI. El conjunto de herramientas más importante lo podemos encontrar en el espacio web de la Open Archive Initiative (<http://www.openarchives.org>) que actualmente presentan 31 herramientas que cumplen OAI-PMH y una breve descripción.

Debido a los requisitos técnicos de los servidores de Adingor, solo eran adaptables las aplicaciones que utilizan el lenguaje PHP y la base de datos mysql. Estas aplicaciones son: CDSware (actualmente denominado CDS Invenio o Indico), OAIbiblio, Open Journal Systems (OJS) que pertenece al Public Knowledge Project y CWIS.

Sin embargo, no eran necesarias todas sus herramientas, y existen dudas sobre su mantenimiento. Además, en el proyecto fue determinante el modo en que se realizaría la ardua tarea de inserción de los artículos de los últimos diez años. Por estas razones se decidió desarrollar una herramienta de software propia que cumpliera los requisitos técnicos y la estructura de descripción de los recursos de información corresponde a la norma ISO 15836 Dublin Core.

Aunque en nuestro caso hemos desarrollado una plataforma propia que aplica el protocolo según las propiedades necesarias del sistema, en la red existen repositorio o paquetes de software dirigidos a la implementación integral de archivos o repositorios que cumplen con el protocolo OAI-PMH, y cuyo software está diseñado para las tecnologías más habituales Java, PERL, APS, JSP, PHP, etc.

Estos sistemas, una vez instalados y configurados en un servidor, ofrecen una serie de funcionalidades básicas para su mantenimiento y gestión. En primer lugar, permiten la gestión, soporte o flujo documental, la edición, revisión y evaluación de documentos o sus metadatos y la transformación de formatos de los ficheros. En general, también ofrecen interfaces de usuario, de autor y de administrador, y permiten la creación de diversos grupos de usuario y niveles de acceso. Estas funcionalidades no eran necesarias en el sistema. En cuanto al contenido, facilitan la creación de colecciones de materiales y generalmente aceptan múltiples formatos de archivo. Las funcionalidades relativas al uso del archivo se limitan a diversas opciones de búsqueda y navegación entre los registros del repositorio (palabras clave, búsqueda por campos de metadatos), así como funciones para su visualización y descarga. En nuestro sistema existe un pequeño buscador, pero actualmente los autores utilizan grandes sistemas de indexación, por esta razón el objetivo fundamental del proyecto es la indexación del sistema a otros como Scopus, Dialnet o Google Académico.

3 Buscadores e indexación

El uso de la meta información, permite mejorar el posicionamiento de la información en los buscadores. El posicionamiento web es de vital importancia, pues los usuarios apenas consultan más allá de los treinta primeros resultados, y una gran masa social sólo los 10 primeros (Codina, 2005). La difusión de los artículos, mejora el posicionamiento en los buscadores, debido a que se aumenta el número de web que enlazan el sistema o sus comunicaciones.

Los buscadores más relevantes son:

SCOPUS (Elsevier) es el mayor navegador científico de la web (Codina, 2005) que permite la búsqueda, acceso y consulta de referencias bibliográficas completas. También aporta otras herramientas de interés.

Web of Science (Thomson Reuters) es un servicio en línea de información científica, que facilita el acceso a un conjunto de bases de datos en las que aparecen citas de artículos de revistas científicas y otros tipos de material impreso.

Dialnet es un proyecto que desde sus inicios trató de apoyar el acceso a los textos completos de la literatura científica en cualquiera de sus manifestaciones. Mediante enlaces directos, en el caso de contenidos a texto completo que figuran en otros servidores, o alojando los documentos a texto completo en el servidor de Dialnet. Desde 2004, Dialnet se convierte en un repositorio que cumple con los protocolos OAI-PMH, lo que garantiza la recuperación y el acceso a estos artículos alojados desde diversos buscadores académicos de todo el mundo. Hoy en día Dialnet es el portal hispano que más metadatos está agregando a los proveedores de servicio.

Google Académico (o Google Scholar) es un buscador de publicaciones científicas de amplia cobertura documental, que permite: buscar el texto completo de un trabajo, buscar la producción bibliográfica de un autor, de una revista o sobre un tema y buscar las citas que recibe un trabajo. En la búsqueda, quedan excluidos documentos no científicos aunque no puede asegurar ningún control de calidad ni existe una normalización en los datos de los resultados, generándose duplicidad de los resultados, lo que crea confusión. Destaca por la aplicación de un algoritmo que ordena los resultados sobre la misma filosofía que el PageRank, basado en la premisa de que los trabajos más citados son también los más importantes, haciéndola extensible al mundo Web mediante los enlaces. Los resultados se ordenan según el número total de citas recibidas, la disponibilidad del texto completo, el autor y la publicación (Torres-Salinas et al., 2009).

Existen otras alternativas de buscadores, respecto a los buscadores más importantes, cabe destacar que según los estudios de (Kousha y Thelwall, 2007) y de (Meho y Yang, 2006) el 24% del total de citas es común a las BD de Google Scholar, Web of Science y Scopus.

4 Proyecto

El sistema de congresos de Adingor cumple con el estándar OAI-PMH y describe los registros con metadatos Dublin Core. De este modo, se facilita la búsqueda y recuperación de sus contenidos.

El sistema desarrollado, permite que se puedan consultar los datos directamente, mediante la interfaz que proporciona el formato web. Asimismo, el sistema contiene información para Google Académico, mediante el uso de metainformación que permite a los robots de búsqueda catalogar las comunicaciones.

De forma independiente, las comunicaciones también se pueden indexar por los motores de búsqueda en el enlace desarrollado para la presentación de los datos mediante el protocolo OAI, que es la especificación del Dublin Core Metadata Initiative (DCMI) de quince elementos (Unqualified Dublin Core), adoptado de la norma ISO 15836.

Benchwood: Una experiencia de benchmarking para la mejora continua

J. Carlos Prado Prado, José A. Comesaña Benavides, Miguel Calero Ferreiro. 2011. Benchwood: Una experiencia de benchmarking para la mejora continua. *XV Congreso de Ingeniería de Organización*, pag. 114-123. Cartagena.

The screenshot shows a document page with the following sections:

- Resumen**
- Autores**
 - J Carlos Prado Prado
 - José A. Comesaña Benavides
 - Miguel Calero Ferreiro
- Congreso**

(cio2011) XV Congreso de Ingeniería de Organización
- Area**

Administración de empresas. Gestión estratégica
- Palabras Clave**
 - benchmarking
 - colaboración
 - mejores prácticas
- [Ver artículo](#)

A red box highlights the following metadata code:

```
<meta content="Benchwood: Una experiencia de benchmarking para la mejora continua" name="citation_title">
<meta content="J Carlos Prado Prado" name="citation_author">
<meta content="José A. Comesaña Benavides" name="citation_author">
<meta content="Miguel Calero Ferreiro" name="citation_author">
<meta content="2011-09-09" name="citation_publication_date">
<meta content="V international conference..." name="citation_journal_title">
<meta content="114" name="citation_firstpage">
<meta content="123" name="citation_lastpage">
<meta content="http://adigor.es/congresos/web/..." name="citation_pdf_url">
```

Fig. 1 Metainformación y información

Los metadatos proporcionados no son obligatorios y todos son repetibles. Algunos de los metadatos que forman el Dublin Core Metadata Element Set son: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage y Rights. En la ISO 15836:2006 vienen definidos cada uno de ellos.

Google Académico emplea un software automatizado, conocido como "crawlers" ("robots"), en busca de sus archivos para su inclusión en los resultados de búsqueda. Funciona de manera similar a la búsqueda normal de Google. Este necesita que el sitio web este estructurado para permitir que sea rastreado.

Google recomienda que se utilicen las etiquetas: título (`citation_title`) que debe contener el título del trabajo, etiqueta de autor (`citation_author`), la etiqueta de fecha de publicación (`citation_publication_date`), el título de la revista (`citation_journal_title`), el Volumen (`citation_volume`), el issue (`citation_issue`), el intervalo de páginas (`citation_firstpage`, `citation_lastpage`), la url del documento (`citation_pdf_url`) (Fig. 1).

1. Procesamiento de Datos: para la administración y gestión de ficheros (archivos PDF de las comunicaciones) se ha desarrollado un módulo que permite incorporar a la Base de Datos la información referente a las comunicaciones, a partir de los archivos JSON (JavaScript Object Notation, es un formato ligero para el intercambio de datos) editados por los scripts que parsean y que han sido generados de forma individual para cada modelo de artículo de cada congreso. Desarrollados a partir de una combinación de script en shell (scripting) y en php (php scripting) sobre cli (command line interface), para algunos congresos, ha sido necesario el uso de una herramienta de OCR. Las herramientas utilizadas son: tesseract-ocr para el parsing, además de las librerías: pdftk (PDF toolkit) y pdf-tools (gestor de pdf).

2. Almacenamiento: el almacenamiento en una base de datos Mysql, permite la independencia respecto a los cambios respecto del lenguaje de etiquetado, la información no se encuentra almacenada en base a un explícito lenguaje de etiquetado, sino que se posibilita que la información pueda ser presentada en varios formatos de forma independiente.
3. Búsquedas: el sistema tiene dos pequeños buscadores, por autor y por título de la obra, no siendo necesario un diseño más exhaustivo, puesto que los usuarios tienden a usar los buscadores externos en los que se intenta indexar este sistema.
4. Administración: el sistema de administración, está basado en un sistema CRUD (Create, Read, Update or Remove) integrado en el framework php de Symfony.
5. Interfaz de usuario: de estructura sencilla, esta contempla la vista de los congresos, sus áreas temáticas y los artículos. Además cada congreso incorpora información de los comités científicos y de organización, junto con la información del congreso y de las entidades organizadoras y patrocinadoras.

5 *Análisis de la información*

El sistema hace referencia a 16 congresos, de los cuales 10 contienen comunicaciones, se han almacenado 54 patrocinadores, referentes a los últimos congresos. El sistema ha almacenado 1749 comunicaciones y 1724 nombres o alias de autores. Esta última cifra es muy significativa, debido a que muchos autores han utilizado diferentes alias, por diversas razones: indicación del congreso, erratas producidas por el escritor, cambio de alias, uso de guiones entre apellidos o variación de la presentación del nombre de pila. Además, algunos autores no completan las especificaciones de los congresos, por lo que estos autores no tenían asignado correo de contacto, en esos casos, se ha asignado el correo indicado en la comunicación (generalmente el autor principal), por lo que el sistema puede contener errores, sin embargo, la actualización realizada ha permitido recuperar comunicaciones extraviadas o link rotos, corregir erratas en los nombres, en el número de autores o incluso en el título. Incluso se ha enriquecido la información de la base de datos mediante la paginación de aquellos congresos que carecían de esta información.



Fig. 2 Portal web del sistema de congresos.

6 Conclusión

La aplicación desarrollada es una herramienta enfocada por excelencia a la difusión de los Congresos de Ingeniería de Organización. Esta aplicación permite estimular la expansión del conocimiento del congreso mediante protocolos y sistemas que facilitan la indexación en los buscadores. El sistema puede ser mejorado mediante la unificación de alias de los autores. Se espera que en los próximos meses los buscadores empiecen a indexar la información y las comunicaciones presentadas en la web de ADINGOR.

7 Referencias

- Codina L, Marcos MC (2005) Posicionamiento web: conceptos y herramientas. *El profesional de la información*; 14(2)
- Ortega Cuevas S (2007) Las citas bibliográficas del IIMAS en el Web of Science (ISI), Scopus (Elsevier) y Google Académico (Google): ¿diferentes o semejantes?. *Ibersid: revista de sistemas de información y documentación*; 2007:387-392
- Barrueco JM, Subirats Coll I. (2003) Open archives initiative. Protocol for Metadata Harvesting (OAI-PMH): descripción, funciones y aplicación de un protocolo. *El Profesional de la Información*; 12(2):99-106
- UNE-ISO 15836:2011. Información y documentación. Conjunto de elementos de metadatos Dublin Core. (ISO 15836:2009)
- Silió T (2005) Fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archival Information System. *El profesional de la información*; 14(5):365-380
- Torres-Salinas D, Ruiz-Pérez R, Delgado-López-Cózar E (2009) Google Scholar como herramienta para la evaluación científica. *El profesional de la información*; 18(5):501-510
- Kousha K, Thelwall M (2007) Google Scholar citations and Google web/url citations: a multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*; 58(7):1055-1065
- Meho LI, Yang K (2006). Multi-faceted approach to citation-based quality assessment for knowledge management. *World library and information congress: 72nd IFLA General conference and council 2006*.
- Codina L (2005). Scopus: el mayor navegador científico en la web. *El profesional de la información*; 14(1)
- CDSware, <http://cdsware.cern.ch/>
- OPUS 2010, <http://www.opus-repository.org>
- OAIbiblio, beta 2004, <http://www.ibiblio.org>
- Open Journal Systems (OJS), Public Knowledge Project. <http://pkp.sfu.ca>
- CWIS <http://scout.wisc.edu>
- tesseract-ocr <http://code.google.com/p/tesseract-ocr/>
- pdftk, PDF toolkit, <http://www.pdflabs.com/tools/pdftk-the-pdf-toolkit/>
- pdf-tools <http://sheelapps.com/index.php?p=PDFTools.HomePage>
- Symfony <http://www.symfony-project.org/>
- Google Scholar <http://scholar.google.ca/intl/en/scholar/inclusion.html>