

Temaria, un portal de artículos de biblioteconomía y documentación

Ernest ABADAL, Assumpció ESTIVILL, Jorge FRANGANILLO, Jesús GASCÓN y Josep Manuel RODRÍGUEZ GAIRÍN*



Actualmente, la mayoría de revistas científicas se distribuyen a través de portales que recopilan gran cantidad de títulos y facilitan el acceso a ellos desde un único sitio web. Se trata de plataformas tanto de origen comercial (*ScienceDirect*, *Wiley*, *Emerald*, etc.) como de libre acceso (*PLOS*, *BioMedCentral*, etc.). Este sistema de distribución mejora la difusión y la visibilidad de las revistas científicas, ya que sus contenidos

185

pueden llegar a audiencias mayores y aumentar así su impacto. Y facilita el acceso a los usuarios, a quienes les basta con dirigirse a unas pocas sedes web para poder consultar la mayor parte de las revistas de su especialidad.

Los portales de revistas más importantes son extranjeros e incluyen, sobre todo, revistas del ámbito anglosajón o publicaciones editadas en inglés. Sin embargo, recientemente han surgido en España las primeras iniciativas para favorecer la difusión del contenido de las revistas científicas españolas en formato digital. Muestra de ello son *Temaria* y otras propuestas similares como, por ejemplo, *Dialnet* (<http://dialnet.unirioja.es>), *e-Revistas* (www.tecnociencia.es/e-revistas), *RACO* (<http://www.raco.cat/>), *Revicien* (www.revicien.net) y *UCM* (<http://www.ucm.es/BUCM/revistasBUC/portal/>). El propósito de *Temaria* es poner a disposición de bibliotecarios y documentalistas todas las revistas de su área que se publican en formato digital, y hacerlo a partir de un sistema uniforme de descripción e indización.

* Grup de Recerca Organització i Recuperació de Continguts Digitals, Universitat de Barcelona, Departament de Biblioteconomia i Documentació.

Contenido

El portal *Temaria* es una creación del grupo de investigación Organización y Recuperación de Contenidos Digitales. Este proyecto ha sido posible por el respaldo económico procedente de varias ayudas de la Universidad de Barcelona y del Plan Nacional de I+D+I (2004–2007).

Temaria indiza los artículos de revistas científicas españolas de biblioteconomía y documentación accesibles en el entorno web, y facilita la consulta de su contenido por medio del conjunto de metadatos Dublin Core (DC)¹. Actualmente incluye más de 1.800 artículos² publicados en las revistas científicas siguientes: *Anales de documentación*, *Bibliodoc*, *BiD*, *Cuadernos de documentación audiovisual*, *Cuadernos de documentación multimedia*, *Cuadernos EUBD Complutense*, *Documentación de las ciencias de la información*, *Hipertext.net*, *Item*, *Revista general de información y documentación*. Acuerdos firmados con otros editores permitirán en el futuro próximo incorporar al portal el *Boletín de la AAB* y *El profesional de la información*, y ampliar su alcance progresivamente al resto de títulos editados en España.

Uno de los principales objetivos de *Temaria* es contribuir a la difusión de las revistas españolas sobre información y documentación, y con ello aumentar su visibilidad. La indización de los artículos de las publicaciones que incorpora permitirá explotar al máximo su fondo retrospectivo.

Aspectos documentales

186

Los dos principales elementos que sustentan la parte documental del portal son los dos instrumentos con los que se indiza el contenido de los artículos: el sistema de metadatos y el tesoro.

En lo que respecta a la asignación de metadatos, *Temaria* dispone de una *Guía de procedimientos para la aplicación de metadatos* que concreta la aplicación del estándar DC y proporciona directrices detalladas sobre la elaboración del contenido de los elementos. La guía tiene como objetivo asegurar la asignación correcta y uniforme de metadatos³.

Trece de los quince elementos del DC se han definido como obligatorios y repetibles, y se aplican a la descripción del artículo tal como se detalla en la tabla 1. En el proyecto *Temaria*, se ha establecido que el uso del elemento “coverage” sea opcional, y se ha descartado el elemento “source”.

1. ESTIVILL, Assumpció; ABADAL, Ernest; FRANGANILLO, Jorge; GASCÓN, Jesús; RODRÍGUEZ GAIRÍN, Josep Manuel. “Uso de metadatos Dublin Core en la descripción y recuperación de artículos de revistas digitales”. Publicado originalmente en: Dublin Core Annual Conference (2005: Madrid), *DC-2005 Proceedings of the International Conference on Dublin Core and Metadata Applications: vocabularies in practices: September 12–15, Madrid, Spain*, Madrid: Universidad Carlos III de Madrid, 2005, pp. 153-156. Accesible en <<http://temaria.net/estivill2005es.pdf>> [Consulta: 14-09-2006].

2. Estos artículos están escritos, mayoritariamente, en castellano y en catalán, aunque también se pueden encontrar algunos en inglés.

3. *Guía de procediments per a l'assignació de metadades Dublin Core en el marc del projecte Temària, revistes digitals de biblioteconomia i documentació*. <<http://temaria.net/guiaestilo.php>>. [Consulta: 10-09-2006].

Tabla 1. Definición del contenido de los elementos del DC en el proyecto *Temaria*

| Elemento | Contenido |
|-----------------|--|
| title | Título del artículo |
| creator | Autor o autores del artículo |
| subject | Materia del artículo |
| description | Resumen del artículo |
| publisher | Editor de la revista |
| contributor | Colaboradores (se usa para editores literarios y traductores) |
| date | Fecha de publicación del número de la revista que incluye el artículo |
| type | Tipo de recurso (siempre es "text") |
| format | Formato del artículo (por ejemplo, HTML, PDF) |
| identifier | Identificador del artículo (generalmente, su dirección electrónica, pero también podría ser otro identificador, como un DOI o un SICI) Cita bibliográfica: título de la revista, y designación numérica y cronológica del número que contiene el artículo |
| language | Idioma del artículo |
| relation | Relación del artículo con la revista que lo contiene y con otros recursos |
| rights | Información sobre el copyright, la propiedad intelectual y la protección jurídica del artículo |

187

En lo que respecta al acceso por materias, se ha partido del *Tesaurus de Biblioteconomía y Documentación* (TBD), desarrollado por el CINDOC, que dispone de equivalencias en inglés y francés a las cuales se han añadido las equivalencias en catalán. Ahora bien, el TBD cubre únicamente el área de la biblioteconomía y la documentación y, por lo tanto, no puede representar todos los conceptos presentes en los contenidos indizados en la base de datos. Así pues, los artículos que tratan sobre la documentación aplicada a un campo concreto —por ejemplo, documentación médica o documentación musical— solo pueden representarse exhaustivamente si el TBD se complementa con otro tesaurus, de alcance más general, que incluya todas las disciplinas. Se ha elaborado, pues, un segundo tesaurus, desarrollado a partir del tesaurus de la UNESCO, para poder representar otras materias. En el marco del proyecto también se han producido entradas para representar topónimos, personas y entidades cuando estos conceptos son materia de los documentos.

Así pues, el resultado es un tesaurus general estructurado en cinco partes. La primera, referida a la documentación, es básicamente el TBD, con ligeras variaciones —referencias ampliadas, términos nuevos, etc.— que no modifican la estructura básica del tesaurus. La segunda, utilizada para la indización del resto de disciplinas, es un tesaurus general basado en el de la UNES-

co, pero abreviado y adaptado a los contenidos que se van integrando en *Temaria*: es como un microtesauro abierto que puede ir ampliando su alcance en función de las materias que se vayan incorporando en la base de datos al indizar artículos nuevos. La tercera parte, estructurada también jerárquicamente, representa la faceta de lugar y ofrece un tesauro de nombres de lugar que también se basa en el tesauro de la UNESCO. Las dos últimas secciones del tesauro agrupan nombres propios de personas y de entidades. En este último caso, solo se incluyen las formas autorizadas de los nombres en orden alfabético, sin una estructura conceptual.

Interfaz de consulta

El acceso a los contenidos de la base de datos es posible de tres formas: mediante la búsqueda simple, la búsqueda avanzada o la navegación por las diversas secciones del tesauro.

La búsqueda simple permite la consulta global a partir de los principales campos del DC en los que se estructuran los registros de la base de datos. Permite buscar términos en todos los campos del registro, o bien en cualquiera de los campos de título, autor, materia, resumen y dirección electrónica. Los términos pueden introducirse directamente en el formulario de búsqueda o pueden obtenerse de los índices, disponibles en la misma pantalla de consulta. Esta pantalla incluye, además, un filtro que permite seleccionar únicamente los documentos escritos en un idioma concreto: castellano, catalán o inglés.

188

Como en el caso de la búsqueda simple, la búsqueda avanzada permite consultar la base de datos a través de los campos de título, autor, materia, resumen y dirección electrónica, facilita la entrada de términos a través de la consulta en los índices de estos campos y ofrece la posibilidad de filtrar los documentos según su idioma.

Como prestaciones adicionales, esta búsqueda alternativa permite combinar varios términos de un mismo campo o términos de campos diferentes, y ofrece filtros adicionales para limitar el resultado a una de las revistas indizadas o a un intervalo de fechas cualquiera.

Finalmente, también es posible examinar el tesauro en los tres idiomas —catalán, español e inglés—, navegando por su estructura jerárquica, una disposición que presenta el campo temático de forma intuitiva y esclarecedora.

Como la información está organizada mediante un vocabulario controlado que señala las relaciones semánticas entre sus términos, es posible expandir los descriptores: los tres mecanismos de búsqueda son capaces de ampliar las búsquedas por materia a todos los conceptos más específicos, cuando los hay. Así, al introducir como argumento de búsqueda un descriptor del tesauro, el sistema de recuperación busca documentos que contengan ese término o cualquiera de los que representen conceptos más específicos.

Con independencia del procedimiento seguido para la consulta de la base de datos, los resultados se presentan en forma de cita bibliográfica normalizada, una característica posible gracias al uso de los elementos DC. Actualmente, las citas se presentan de acuerdo con el estilo MLA (Modern Language Association), y está previsto incorporar más pautas de estilo para que sea posible escoger el modelo de citación de entre diversas opciones. Cada cita incluye el

nombre del autor o de los autores del artículo, la fecha de publicación, el título del artículo, el título de la revista, la designación numérica y cronológica del número que incluye el artículo, la paginación (si se aplica), y la dirección electrónica del artículo. Opcionalmente, se puede visualizar también el resumen del artículo y los descriptores de materia que se le han asignado. Los descriptores aparecen en forma de hipertexto, de modo que es posible navegar con agilidad hacia otros documentos de temática similar.

Las referencias resultantes de una búsqueda pueden ordenarse alfabéticamente por autor o por orden cronológico, en cualquier momento. El criterio de ordenación puede ser ascendente o descendente.

La interfaz de consulta de *Temaria* es multilingüe: las pantallas de consulta —los formularios de búsqueda y el tesoro— están disponibles en catalán, castellano e inglés. El acceso multilingüe por materias es una prestación de vital importancia, ya que permite al usuario realizar cualquier tipo de consulta utilizando los descriptores de materia (campo DC.Subject) en su idioma⁴.

Suscripción a los contenidos

Por medio de la tecnología de sindicación RSS 2 (Really Simple Syndication) es posible suscribirse a los contenidos indizados en el portal y estar al día de las novedades. RSS es un formato de descripción de datos basado en XML que facilita la distribución automática de información mediante unos ficheros llamados fuentes o suministros. Gracias a este mecanismo de difusión, convertido ya en el principal medio para divulgar información digital en forma de titulares, es posible tener noticia de nuevos artículos publicados en una revista concreta, escritos por un autor determinado o relacionados con algún tema⁵. Cualquiera que sea la necesidad, basta con introducir la dirección del suministro RSS en un agregador de contenido, una aplicación destinada a leer fuentes de sindicación, para estar al día de las novedades respectivas. Estas aplicaciones aparecieron en forma de programas de escritorio y también como servicios basados en el entorno web, pero hoy día ya van integrados en las últimas versiones de los navegadores más extendidos.

189

4. ABADAL, Ernest; ESTIVILL, Assumpció; FRANGANILLO, Jorge; GASCÓN, Jesús; RODRÍGUEZ GAIRÍN, Josep Manuel. "L'accés multilingüe per matèries a articles de revista". ISKO. Capítulo Español. Congreso (7è: 2005: Barcelona). *La dimensió humana de l'organització del coneixement*. Barcelona: Universitat de Barcelona, Departament de Biblioteconomia i Documentació, 2005, pp. 33-50. Accesible en <<http://bd.ub.es/isko2005/abadal.pdf>> [Consulta: 14-09-2006].

GASCÓN, Jesús; SERVER, Adán; ABADAL, Ernest; ESTIVILL, Assumpció; FRANGANILLO, Jorge; RODRÍGUEZ GAIRÍN, Josep Manuel. "Aplicación de un tesoro multilingüe: problemas planteados y objetivos futuros". Presentación en el seminario "Uso de lenguajes documentales en la web semántica", Madrid, 12 de diciembre de 2005. Accesible en <<http://temaria.net/gascon2005.pps>> [Consulta: 18-09-2006].

5. RODRÍGUEZ GAIRÍN, Josep Manuel; FRANGANILLO, Jorge; ABADAL, Ernest; ESTIVILL, Assumpció; GASCÓN, Jesús. "Sindicación de contenidos en un portal de revistas: *Temaria*". *El profesional de la información*, vol. 15, núm. 3, mayo-junio de 2006, pp. 214-221. Accesible en <<http://temaria.net/gairin2006.pdf>> [Consulta: 14-09-2006].

La suscripción a artículos de revista mediante un agregador aporta varios beneficios. Las personas en busca de información reciente se benefician de la sindicación porque no se ven obligadas a consultar la base de datos ni las revistas indizadas en ella, que tal vez no ofrezcan ninguna novedad. Cuando el agregador detecta la aparición de contenidos nuevos que responden al perfil del suscriptor, se lo notifican y le presentan entonces una lista de referencias, en forma de enlaces de hipertexto, con los resúmenes correspondientes. De este modo, con solo activar un enlace, se puede leer el documento original en el navegador.

El engranaje de la redifusión de contenidos ofrece además otras ventajas: los agregadores confeccionan una especie de noticiero a medida, libran del correo basura y garantizan el anonimato al usuario, que no tendrá más noticias de un editor una vez que haya cancelado la suscripción. La sindicación se presenta, pues, como una reformulación de los tradicionales servicios de alerta bibliográfica, que se han proveído, hasta hace poco, únicamente a través del correo electrónico⁶. Y también abre la puerta a otras maneras de absorber información: dado que los suministros RSS son documentos XML estandarizados, la información contenida en ellos se puede reutilizar, por ejemplo, integrándola en un sitio web cualquiera⁷. Así, con el apoyo de la sindicación, los editores de las revistas indizadas en Temaria pueden atraer visitas y aumentar la visibilidad de sus publicaciones, más allá del tráfico que les dirige el propio portal.

Aspectos tecnológicos

190

La aplicación informática se ha desarrollado con el lenguaje de programación PHP y el gestor de bases de datos MySQL. Se escogió esta combinación por varios motivos:

—La implementación queda libre de costos porque PHP es un lenguaje gratuito y MySQL es un sistema disponible bajo licencia GPL (licencia pública general, gratuita para aplicaciones sin ánimo de lucro).

—Los dos componentes son modelos actuales, de modo que es amplio el repertorio de recursos disponibles para el desarrollo de aplicaciones: documentación de dominio público, *scripts* libres, etc.

—PHP y MySQL son modelos concebidos el uno para el otro y, por lo tanto, interactúan con fluidez.

En el diseño de interacción del portal se han considerado principios fundamentales sobre diseño gráfico y usabilidad con el ánimo de ofrecer una interfaz amistosa, ordenada y funcional. También se ha procurado distribuir la información de modo consistente y predecible, en modelos de pantalla concretos y uniformes, para que todo lo necesario —controles e información— se pueda localizar con rapidez y con independencia del contexto.

6. FRANGANILLO, Jorge; CATALÁN, Marcos A. "Bitácoras y sindicación de contenidos: dos herramientas para difundir información". En: *BiD: textos universitaris de biblioteconomia i documentació*, 2005, diciembre, núm. 15.

7. MOFFAT, Malcolm. "RSS: a primer for publishers and content providers". En: *The new review of information networking*, 2003, pp. 123-144. Accesible en <http://www.techxtra.ac.uk/rss_primer/> [Consulta: 18-09-2006].

Líneas de trabajo futuras

Entre los propósitos para el desarrollo de *Temaria* durante los próximos meses, se han planteado un conjunto de ampliaciones que han de sumar valor al portal: se ha proyectado indizar nuevos títulos de revista y se prevé extender las conexiones semánticas de los términos del tesauruso para incluir relaciones de sinonimia.

Otras actuaciones previstas son la incorporación de nuevas prestaciones (directorio, nuevos modelos de estilo de citación, etc.), la visualización de los resultados de las consultas agrupados por nuevos criterios y el desarrollo de un sistema informático que ha de permitir búsquedas en el texto completo de los artículos. En esta última acción, interesará analizar la validez y la efectividad de los metadatos respecto al texto completo en el proceso de recuperación de la información.