

BASES DE DATOS DE INFORMACION Y DOCUMENTACION EN INTERNET

Las bases de datos son una herramienta de vital importancia para el desarrollo de la actividad profesional. Existen dos colectivos —que a menudo son el mismo— especialmente beneficiados por las bases de datos. Por una parte, los investigadores y, por otra, los profesionales. No es cuestión de delimitar en este trabajo las tipologías de bases de datos, ni de clasificarlas por sus formas, contenidos, etc., ya que la función de esta sección es seleccionar y reseñar los recursos telemáticos más útiles para los profesionales de la gestión de la información.

Por la necesidad de delimitar conceptos, se podría considerar como base de datos cualquier recopilación organizada de información sobre la que haya habido análisis documental y que disponga de un sistema de búsqueda específico. Partiendo de esta idea, se podrían incluir los catálogos bibliotecarios. Sin embargo, en este artículo se ha optado por emplear como criterio el seleccionar sólo aquellas bases formadas principalmente por descripciones de documentos impresos o electrónicos procedentes de vaciados de revistas de la profesión.

Como segundo criterio se ha decidido no incluir en esta revisión aquellas bases de datos de temática profesional que, a pesar de poder ser consultadas en línea, no son de libre acceso, sino que es necesario contar con una palabra de paso o algún otro sistema de abono. Por lo tanto, sólo se analizan aquí las bases de datos profesionales accesibles de forma gratuita a través de web, compuestas por registros, resúmenes o textos íntegros de artículos de revistas, partes componentes de otras obras, literatura gris y, en menor medida, monografías.

Existen varias páginas web en las que se relacionan las bases de datos de pago. Un directorio de este tipo es el realizado por la biblioteca de la Universidad de Auckland y puede ser consultado en <http://www.auckland.ac.nz/lbr/libsci/lib.htm>. No obstante, se ofrecen a continuación las direcciones de las páginas informativas de las principales bases de datos profesionales de los ámbitos de la archivística, la biblioteconomía y la documentación que no se comentan en este trabajo, ya que su acceso es de pago:

ALISA: <http://www.auckland.ac.nz/lbr/alisa.htm>

Infodata: <http://www.fiz-karlsruhe.de/stn/Databases/infodata.html>

Inspecc: <http://www.iee.org.uk/publish/inspec/>

ISA: <http://www.infoday.com/catalog/infosvcs.htm#info3>

LISA: <http://catalogue.bowker-saur.co.uk/data/198.html>

Library Literature: <http://www.hwwilson.com/liblit.html>

Como ya se dijo, en este artículo sólo se comentan las bases de acceso gratuito (total o parcial) a través de navegadores web. Las reseñas que se ofrecen, se agrupan en bases de datos seleccionadas entre las de autoría española y aquellas que se elaboran fuera de España. Además, se realiza una segunda división en virtud de si las bases de datos recogen exclusivamente documentos de las ciencias de la información, o bien recogen revistas de distintos campos de la ciencia.

1 Bases de datos españolas de documentación

Los profesionales españoles pueden congratularse de contar con un buen número de bases de datos específicas de contenidos relacionados con la archivística, la biblioteconomía y la documentación. Entre unas y otras es posible obtener una información completa sobre cualquier tema relacionado con estos campos. Casi todas ellas están producidas y mantenidas por instituciones relacionadas con la documentación, aunque también se encuentran dos realizadas y mantenidas por empresas privadas del ámbito de la información. Merece la pena destacar la consulta en web de forma gratuita de la base de datos ISOC-DC de amplia tradición en el mundo profesional. El hecho de que esta base sea por fin de acceso libre es una muestra de cómo los profesionales de la documentación están demostrando ser uno de los colectivos más proclives a la cooperación y a la compartición de recursos.

ISOC-DC

CINDOC, CSIC

<http://www.cindoc.csic.es/prod/isocdc.htm>

En el momento de redactar esta reseña la base de documentación del CINDOC se encuentra en fase de adaptación a la nueva realidad que para ella supone el acceso gratuito a través del web. Esto puede implicar cambios en la denominación o en las direcciones de localización en el servidor en el que se encuentra instalada. Por ello, se ha optado por emplear el nombre por el que en estos momentos es conocida la base y la URL desde la que se ofrece información de la misma y a partir de la cual se podrá acceder a su contenido. ISOC-DC vacía unas 35 revistas españolas, además de beneficiarse del vaciado de más de 2.000 revistas de todas las disciplinas de las ciencias humanas y sociales, que ocasionalmente pueden incluir algún trabajo útil para la profesión. También se incluyen actas de congresos, publicaciones colectivas, etc., siempre relacionadas con la información y la documentación en cualquiera de sus vertientes. Su cobertura cronológica se inicia en 1975 y abarca hasta la actualidad. El número de registros supera los 6.000, estimándose en 500 su crecimiento anual. Esta base realiza una exhaustiva descripción de los documentos, ya que, además de los campos formales (autor, título, publicación, etc.), los documentos se sistematizan utilizando una clasificación jerárquica propia, se asignan identificadores y se realizan resúmenes de los trabajos analizados, además de indicar el lugar de trabajo de los autores. Su actualización es diaria y el programa de recuperación Basis Plus. Los documentos incluidos en esta base se encuentran depositados en el CINDOC, centro que dispone de un servicio de acceso al documento para quienes estén interesados en la fuente original.

BEDOC

Inforárea

<http://www.inforarea.es/welcome.asp>

Según se indica en la página de presentación de BEDOC, se trata de una base de datos de información bibliográfica española sobre las áreas de actividad de la em-

presa que la realiza: archivos, bibliotecas, redes, etc. Esta base contiene más de cuatro mil referencias de documentos de autores españoles procedentes de revistas profesionales, actas de congresos y conferencias publicadas desde 1989 hasta la actualidad. Para la recuperación de los documentos se utiliza la herramienta GLAS@cces, pudiendo elegir por consultar de forma sencilla o través de un formulario de búsqueda avanzada. La búsqueda simple se realiza eligiendo qué tipo de listado inicial se desea recuperar: autor, título, materia, etc. Una vez descargado este listado se efectuarán las búsquedas sobre el mismo. La búsqueda avanzada permite combinar términos y campos.

Datathéke

Facultad de Traducción y Documentación, Universidad de Salamanca
<http://milano.usal.es/dtt.htm>

Datathéke tiene el honor de ser la primera base de datos especializada que se puso en línea de forma gratuita en España. Está producida por la Facultad de Traducción y Documentación de la Universidad de Salamanca y sus referencias proceden del vaciado de las revistas profesionales que se reciben en la biblioteca de este centro, estimándose en unos 250 el número de títulos y en alrededor de 12.000 los registros incluidos en la base. Su temática está relacionada con las áreas de conocimiento de la entidad productora: biblioteconomía y documentación, informática, traducción e interpretación; aunque es notablemente más alto el número de documentos que tratan de materias afines a la documentación. Datathéke ha dejado de alimentarse, por lo que en la misma sólo podrán encontrarse documentos publicados entre

1991 y 1996. Su cobertura temática es internacional y sus registros aportan datos formales y de contenido, incluyendo en la mayoría de los casos resúmenes de los artículos y documentos incluidos en la base.

BARC. Bibliografía de archivos

Centro de Información Documental de Archivos (CIDA), M.E.C.

<http://www.mcu.es/bases/spa/barc/BARC.html>

La base de datos BARC centra su contenido en temas archivísticos, aunque incorpora un significativo número de referencias sobre ciencias afines: biblioteconomía, documentación, historia del libro, preservación documental, automatización, etc. Está realizada por el Centro de Información Documental de Archivos (CIDA), que utiliza como fuente el fondo de publicaciones periódicas, monografías y actas de congresos, de su propia biblioteca. Junto a esta base se encuentran otras entre las que destacan las que describen documentos de archivos estatales. La consulta a BARC ofrece muchas posibilidades, ya que se permite la búsqueda de un término en todos los campos y el uso de operadores booleanos y de truncamiento; además, distingue entre la búsqueda por descriptores, a partir de las palabras clave empleadas por el CIDA, y la búsqueda por materias, términos normalizados o bloques temáticos en los que se organiza la información de la base: obras de referencia, ciencias relacionadas con los archivos, archivística, archivos, instrumentos de información, conservación de documentos, reproducción de documentos, edificios e instalaciones, biblioteconomía, documentación, informática, museología, etc.

18:02:37

Consulta por web a los Catálogos en SABINI de las Bibliotecas de la Fundación

Biblioteca Infantil y Juvenil (Salamanca)
 Más de 50 000 registros de libros y otros soportes para niños y jóvenes. Todas las obras, tanto de ficción como de carácter informativo, tienen asignada materia a partir de los Encabezamientos de materia para libros infantiles y juveniles (Madrid, 1994), elaborada por un equipo de trabajo de la Fundación Germán Sánchez Ruzpérez, coordinado por Alberto Villalón.

Biblioteca Pública Municipal de Peñaranda de Bracamonte
 Catálogo multimedia de la Biblioteca Municipal de Peñaranda de Bracamonte. La asignación de materias se realiza a partir de la Lista de encabezamientos de materia para las bibliotecas públicas (Madrid, 1987) publicada por el Ministerio de Cultura.

Centro de Documentación sobre Literatura Infantil y Juvenil (Salamanca)
 Monografías, publicaciones periódicas especializadas, documentos sobre literatura infantil y juvenil, lectura, promoción de la lectura y bibliotecas infantiles y escolares. Todos los documentos tienen asignados descriptores de materia a partir del Tesoro de literatura infantil (Caracas, 1991), elaborado por el Banco del Libro de Venezuela bajo la coordinación de Alberto Villalón, utilizado por todas las instituciones miembros de la Red Latinoamericana de Centros de Documentación de Literatura Infantil y Juvenil.

Búsqueda en todos los catálogos
 Catálogo global de la Fundación Germán Sánchez Ruzpérez
 Unifica en un mismo listado de materias los tres anteriores

LA FUNDACIÓN · CENTRO DE MADRID · CENTRO DE PEÑARANDA · CENTRO DE SALAMANCA
 CATALOGOS Y BASES DATOS · DOCUMENTOS ELECTRONICOS · LIBRERIA · DIRECTORIO · AGENDA

<http://www.fundaciongsr.es/catalogos/frames.htm> 100%

Centro Internacional del Libro Infantil y Juvenil

Fundación Germán Sánchez Ruipérez

<http://www.fundaciongsr.es/catalogos/frames.htm>

La Fundación Germán Sánchez Ruipérez es la entidad no oficial que más esfuerzos ha dedicado en España al ámbito de la lectura, la promoción del libro infantil y juvenil, las bibliotecas escolares y las bibliotecas públicas. Hasta fechas recientes no contaba con un servidor web, pero, desde el momento en que se incorpora al espacio de la información digital, se ha convertido en uno de los webs más empleados por quienes trabajan en campos relacionados con su actividad. De estética esmerada y contenido estructurado, en este servidor se pueden consultar varias bases de datos bibliográficas, entre las que se encuentran la del Centro de Documentación sobre Literatura Infantil y Juvenil. En dicho centro, además de documentos sobre el libro infantil, se vacían publicaciones, monografías y actas de congreso relativas a bibliotecas infantiles, bibliotecas escolares y bibliotecas públicas; en menor medida se incluyen otras tipologías bibliotecarias y documentos sobre temas de campos similares. Los catálogos bibliográficos funcionan con Sabini y la consulta se puede hacer de forma simple o compleja, permitiendo aquí combinar términos de búsqueda y delimitar los años, el idioma de los documentos y la ordenación de los resultados. La visualización de los registros permite una primera vista en la que sólo se ofrecen los campos meramente descriptivos y una posterior ampliación de los documentos en donde también se ofrece un resumen del documento y una serie de descriptores, tomados estos del Tesoro de Literatura Infantil. La actividad del centro productor se inicia en 1988, aunque pueden encontrarse referencias anteriores a esta fecha en su base de datos.

Biblioteconomía y Documentación / Bibliografía de bibliobuses

Biblioteca de Castilla y León

<http://www.bcl.jcyl.es/Catalogos/>

La Biblioteca de Castilla y León ofrece al usuario de Internet la consulta a varias bases de datos bibliográficas, que representan las distintas colecciones que posee esta biblioteca, una de las pioneras en cuanto al uso de Internet como medio de difusión de información. La dirección de Internet que se ofrece más arriba corresponde a la página desde la que se accede a los distintos catálogos. Entre ellos se encuentra el Catálogo de Biblioteconomía y Documentación, que recoge el fondo profesional de la biblioteca, así como la Bibliografía sobre bibliobuses. El primero de estos catálogos no va a ser comentado aquí, ya que, como se dijo en la introducción, se ha tomado como criterio reseñar las bases de datos que vacíen publicaciones periódicas; el mencionado catálogo está compuesto por monografías, por lo que el análisis se centra en la Bibliografía sobre bibliobuses, en la cual sí hay descripción de documentos de revistas, además de monografías, literatura gris y actas de congresos. La consulta a la base, que emplea el programa Absys, se realiza desde un sencillo formulario, pudiendo interrogar por un término en todos los campos simultáneamente o elegir entre los campos de autor, título, materias, editorial y fechas de publicación. Los campos de autor y materia disponen de un índice que puede ser consultado antes de iniciar la búsqueda definitiva.

REID

Sistemas Documentales

<http://www.sisdoc.es/html/Servicios/Recursos/reid.html>

A pesar de no ser gratuita se ha incluido esta base de datos de resúmenes de información y documentación porque permite el acceso libre a una versión de demostración de la misma. Sus fuentes son una veintena de revistas españolas, además de actas de congresos de documentación celebrados en España, vaciados a partir de 1993. Cada registro está compuesto por una exhaustiva serie de datos descriptivos, analíticos y de control. La versión de demostración facilita la consulta a un reducido número de documentos, pero posibilita conocer el contenido y la forma de presentación de los registros de esta base. Junto a esta base la empresa productora ofrece el acceso gratuito a otros recursos de información profesionales.

2 Bases de datos españolas multidisciplinares con contenidos de documentación

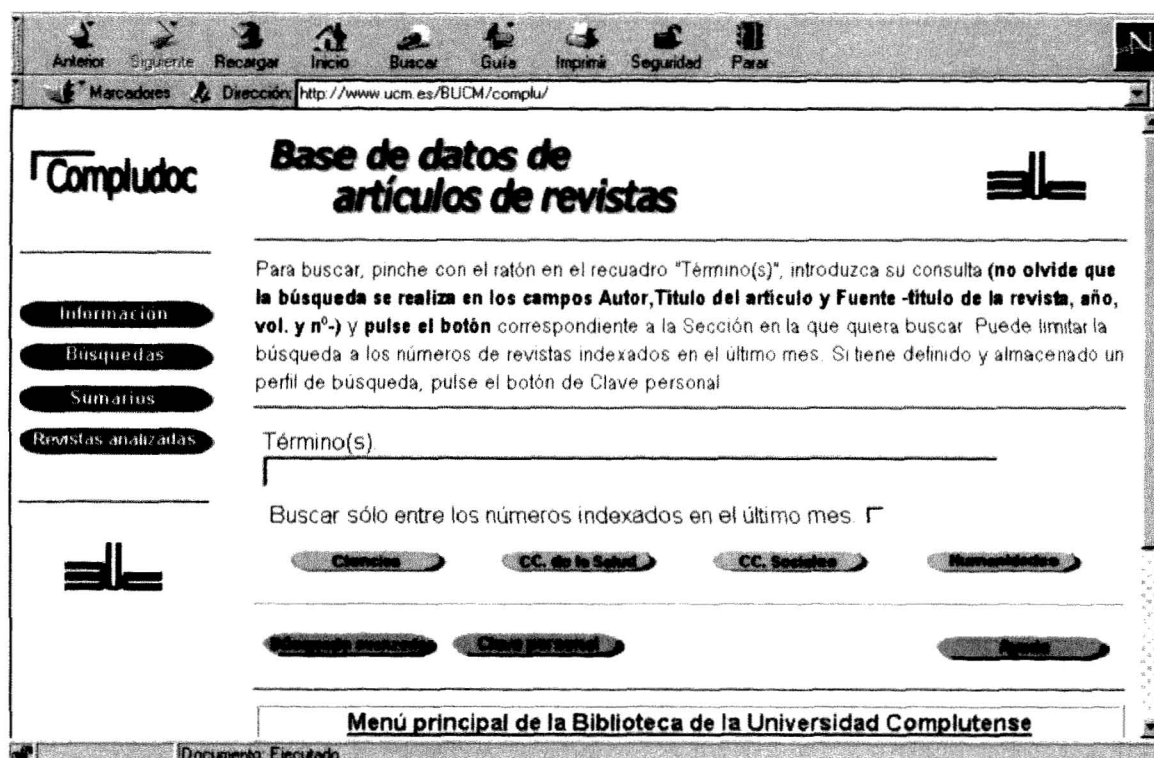
En muchas ocasiones las bases de datos de los campos relacionados con la documentación se presentan como un sector dentro de bases de temática multidisciplinar. Este es el caso de las que se comentan a continuación. Todas ellas contienen una cantidad más o menos importante de registros procedentes de revistas profesionales de documentación. De las cuatro que se comentan aquí, tres proceden del ámbito de la biblioteca universitaria, mientras que una cuarta se inscribe dentro de las iniciativas oficiales de colaboración entre bibliotecas públicas. Es digna de destacar la actividad de las bibliotecas universitarias que vacían sus revistas y ponen en línea el contenido de las mismas; se trata de otro indicador más del nivel de cooperación existente en este tipo de centros.

COMPLUDOC

Biblioteca de la Universidad Complutense de Madrid

<http://www.ucm.es/BUCM/complu/>

Las bibliotecas de la Universidad Complutense de Madrid han creado una base de datos que permite acceder a la descripción bibliográfica de las revistas científicas depositadas en sus centros. COMPLUDOC nace en 1997 y tuvo distintos nombres y direcciones hasta asentarse en su actual URL y denominación. Divide su contenido en cuatro secciones: ciencias, ciencias de la salud, ciencias sociales y humanidades. Los registros de documentación se encuentran en el apartado de Ciencias Sociales, donde hay más de 70.000 registros procedentes de 345 revistas, la mayor parte españolas. Desde estas páginas es posible acceder a un listado completo de las revistas analizadas, así como a los boletines de novedades que las bibliotecas participantes elaboran mensualmente. Compludoc permite realizar búsquedas avanzadas y el empleo de operadores booleanos y de truncamiento. En todo caso, se pueden interrogar por los campos de: autor, título y fuente, ya que simplemente se ofrece una descripción básica de los artículos analizados. Los usuarios



autorizados pueden solicitar copias del documento original, así como otras instituciones, quienes deben emplear sus correspondientes servicios de préstamo interbibliotecario.

REBECA

Ministerio de Educación y Cultura

<http://www.mcu.es/REBECA>

Realizada por el Ministerio de Educación y Cultura, en colaboración con varios gobiernos regionales, el objetivo primordial de REBECA es la formación de una gran base de datos bibliográfica cuyos registros puedan ser importados por la veintena de bibliotecas participantes en el proyecto. Para que la descarga de los registros sea efectiva, todos los documentos de la base están en formato IBERMARC. En estos momentos está compuesta por unos 325.000 registros, correspondientes a varios tipos de documentos, destacando los cerca de 20.000 artículos de revistas, motivo por el que se incluye esta base en el presente artículo. El vaciado de revistas incluye varias revistas españolas de temática biblioteconómica, cuyas catalogaciones analíticas pueden ser descargadas por sistemas compatibles. Además de las colaboraciones de la Subdirección General de Coordinación Bibliotecaria y las que realizan las Bibliotecas Públicas del Estado participantes, REBECA cuenta con otras aportaciones, como la base de datos de la Asociación de Revistas Culturales de España (ARCE).

Sumarios de Revistas Españolas

Universidad Carlos III de Madrid

<http://sauron.uc3m.es/cgi-bin/abweb/L1/T2>

La biblioteca de la Universidad Carlos III de Madrid pone a disposición de sus usuarios, así como de todos los usuarios de Internet, una base de datos bibliográfica procedente del vaciado de alrededor de 300 revistas españolas e hispanoamericanas de todos los campos de la ciencia. Entre ellas se encuentran una docena de títulos de biblioteconomía y documentación. La cobertura temática se inicia en febrero de 1997 y abarca hasta la actualidad. La consulta a la base se puede hacer desde la página principal del OPAC de la Universidad Carlos III, como una opción de las disponibles. Entre los distintos catálogos también se encuentra el del fondo de monografías del Departamento de Biblioteconomía y Documentación. El número de registros de la base se acerca a los 15.000 documentos de todas las ramas del saber. La búsqueda está limitada a los campos de autor personal, título de la revista y título del artículo, aunque puede realizarse de forma simultánea en todas ellas. Los resultados se presentan, como es habitual en Absys, con una pantalla intermedia en las que sólo aparecen los autores y los títulos de los artículos. Los registros se presentan en forma de catalogación analítica, es decir, que únicamente se ofrecen los datos del autor, el título del artículo y los datos de la fuente: revista, volumen, fechas, páginas, etc.

Base de datos de sumarios

CBUC

<http://sumaris.cbuc.es>

La base de datos de sumarios del Consorci de Biblioteques Universitàries de Catalunya está formada por más de 650.000 descripciones de artículos de revistas multidisciplinares. Para su elaboración se parte de los sumarios proporcionados por la empresa Ebsco de las más de cinco mil revistas a las que están suscritas las bibliotecas que forman parte de esta entidad cooperativa. Al tratarse de una base de sumarios, los datos que se ofrecen son los fundamentales para localizar un documento: autor, título del artículo, fuente e ISSN. Las incorporaciones se hacen de forma semanal, por lo que la base está siempre muy actualizada. En su gran mayoría, las revistas vaciadas son extranjeras, sobre todo estadounidenses, lo que no resta interés ya que es evidente su utilidad para la investigación. Los artículos descritos pueden ser solicitados mediante préstamo interbibliotecario. Se echa en falta más información sobre el contenido temático de la base, ya que no existe ningún campo de descriptores o clasificación en la base.

3 Bases de datos internacionales de documentación

Se comentan en este tercer apartado una serie de bases de datos producidas fuera de España y especializadas en documentación en su sentido más amplio. En estas bases se pueden encontrar artículos de revistas profesionales de archivística, biblioteconomía y documentación y, además, otras ciencias afines y relacionadas. Al tratarse de bases ela-

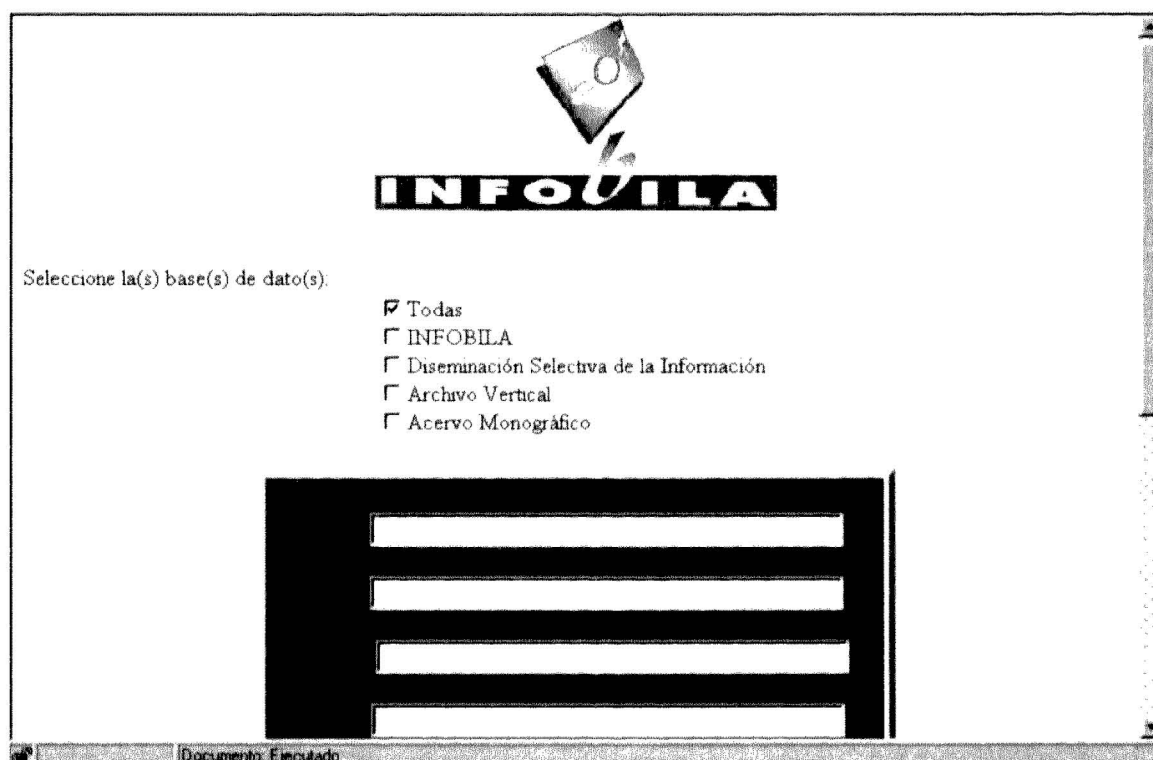
boradas fuera del estado español, los documentos descritos fueron publicados en revistas de las áreas geográficas en las que desarrollan su función los organismos productores. En estas bases internacionales apenas se recogen documentos editados en España.

INFOBILA

Centro Universitario de Investigaciones Bibliotecológicas

<http://cuib.unam.mx/infobila/indice.html>

Esta base de datos tenía ya bastante fama dentro del panorama bibliotecológico latinoamericano antes de ofrecerse de forma gratuita a través del web. En las páginas de presentación de la misma se dice que la base de datos Información y Bibliotecología Latinoamericana surgió en la Biblioteca del Centro Universitario de Investigaciones Bibliotecológicas de México, en el año de 1985 con la intención de controlar la producción latina de documentos relacionados con las bibliotecas, los archivos y la documentación en general. En un principio, la base se nutría de los fondos que se reciben en la biblioteca del centro productor, pero en la actualidad recibe contribuciones de bibliotecas de varios países americanos. La base está realizada en el programa Micro CDS/ISIS bajo formato MARC y posee cerca de 12.000 registros de documentos de autores latinoamericanos. La información puede recuperarse por varios campos: autor, título, país, fechas, descriptores, etc. Los registros ofrecen una descripción de los documentos, además de incorporar un resumen de los mismos y unos descriptores extraídos del Tesoro Latinoamericano en Bibliotecología y Ciencias de la Información.



Seleccione la(s) base(s) de dato(s):

- Todas
- INFOBILA
- Diseminación Selectiva de la Información
- Archivo Vertical
- Acervo Monográfico

Documento Ejecutado

BUBL Journals- Library and Information Science

BUBL Information Service, Andersonian Library, Strathclyde University

<http://bubl.ac.uk/journals/lis/>

Base de datos de artículos procedentes de alrededor de 130 revistas internacionales, electrónicas e impresas, especializadas en documentación. Entre las publicaciones que se vacían sólo hay un título español. La URL que se ofrece en este trabajo recoge únicamente publicaciones profesionales de información y documentación, pudiéndose consultar de manera global (todas las revistas simultáneamente) o bien a partir de los sumarios de un título en concreto. Las revistas han sido agrupadas en una serie de campos temáticos: generales, adquisiciones, catalogación e indización, tecnología, etc. La relación de publicaciones vaciadas informa asimismo de si en la base BUBL Journals sólo está el sumario de cada revista o también el resumen de cada artículo. La búsqueda puede realizarse de forma simple o mediante operadores de exactitud, truncamiento y booleanos. Los resultados de la consulta se ofrecen en una primer listado que remite al sumario del número en el que se ha encontrado la información. En algunas revistas remite a la dirección en la que se encuentra el texto completo del artículo o bien lo reproduce directamente. Debido a que integra artículos procedentes de revistas impresas y electrónicas esta base es una de las más recomendables para los profesionales de la información.

Library Reference Center

EBSCO

<http://www.epnet.com/lrc.html>

A pesar de que se requiera introducir datos de identificación, esta base de datos es de acceso gratuito. La empresa Ebsco es su responsable y en ella se describen e indizan artículos procedentes de una treintena de publicaciones relacionadas con las bibliotecas, de las cuales ninguna es española. No obstante, se encuentran en la base títulos muy importantes dentro del panorama internacional de las revistas profesionales. Las búsquedas pueden realizarse a través de un formulario general para recuperar un término en cualquier campo o bien empleando un formulario de búsqueda avanzada. Los primeros resultados se muestran en una tabla y ya contienen los datos básicos de identificación de los artículos. Posteriormente se ofrece una descripción más amplia, que incluye un resumen de cada documento. Los registros pueden ser enviados a través de correo electrónico.

Current Cites

University of Berkeley

<http://sunsite.berkeley.edu/CurrentCites/>

Base de artículos electrónicos y publicaciones accesibles en línea. Los documentos recogidos en Current Cites proceden del análisis de 225 webs. El período de cobertura es desde enero de 1995 hasta la actualidad. La búsqueda de un término se realiza sobre las descripciones y el texto completo de los artículos, permitiéndose

combinar términos a través de los operadores booleanos. La base utiliza el programa Harvest para cargar la información y su posterior recuperación; para ello emplea el denominado Summary Object Interchange Format (SOIF). Los resultados de las búsquedas ofrecen una relación de documentos sobre la que el usuario puede optar por acceder al artículo desde su localización original o bien consultarlo, junto a la correspondiente descripción, en la ficha SOIF de Current Contents. Esa base realiza un análisis documental muy básico de los recursos electrónicos contenidos en la misma. En realidad se limita a reproducir los datos de localización: autor, título y URL del artículo, y a copiar el resumen y el texto del documento en cuestión. A pesar de todo, Current Cites es una buena herramienta para localizar documentos profesionales que estén accesibles en línea.

Index Morganagus

University of Berkeley

<http://sunsite.berkeley.edu/~emorgan/morganagus/index.html>

Base de datos de texto completo de artículos y documentos de publicaciones electrónicas. Su filosofía y tecnología es muy similar a la base de datos anterior, ya que también ofrece el texto completo de los artículos vaciados y, además, emplea Harvest para la gestión de los datos de la base. Index Morganagus utiliza unas cincuenta publicaciones electrónicas para alimentar su base. Entre ellas se encuentran las principales revistas profesionales de biblioteconomía y documentación; la mayoría son anglosajonas, con algún caso aislado de publicaciones alemanas o francesas, pero sin representación de publicaciones elaboradas en España. La base puede interrogarse de forma global, empleando un único formulario para las consultas simples y las booleanas. La consulta también puede limitarse a revistas electrónicas concretas, de las cuales, además de su sumario, se ofrece una amplia información sobre su política editorial, secciones y otros datos de interés. Como ocurría en la base anterior, los resultados se muestran de tal forma que el usuario pueda optar por conectar directamente con la fuente original o bien consultar el archivo SOIF, que genera Harvest, en el cual se ofrecen datos básicos de identificación del artículo y el texto completo del mismo.

4 Base de datos internacionales multidisciplinares con contenidos de documentación

Al igual que ocurría en el caso español, existen varias base de datos multidisciplinares que son de gran utilidad para la localización de recursos bibliográficos y electrónicos de temas profesionales de cualquiera de los ámbitos de la documentación. Algunas de las bases que se comentan aquí, como es el caso de ERIC, tienen una orientación temática que no se corresponde en un principio con la biblioteconomía o la documentación, pero, dado que poseen un elevado número de registros sobre estos temas, se ha estimado conveniente incluirlas en esta selección de bases de datos. Existen más bases multidisciplinares de las que aquí se comentan, pero las que a continuación se exponen son muy útiles para el usuario español, bien por la calidad de sus informaciones o bien por el tipo de fuentes que utilizan para la elaboración de sus bases.

ERIC

Educational Research Information Center.

<http://ericir.syr.edu/Eric>

ERIC es una de las bases de datos más afamadas de todo el mundo. Está patrocinada por el Department of Education Office of Educational Research and Improvement (OERI) y recoge artículos de revistas y otros tipos de documentos similares relacionados con la educación, entre los que se encuentran las bibliotecas, entendidas como complemento del proceso educativo. La elaboración de la base se inicia en 1966, aunque su acceso gratuito, a través de Internet, es muy reciente. ERIC vacía alrededor de 750 revistas profesionales. Además de artículos de revistas, en ERIC se pueden encontrar referencias y resúmenes de materiales audiovisuales, bibliografías, monografías, actas de congresos, documentos oficiales, informes, descripciones de software y tesis. ERIC cuenta con dos sub-bases: Resources in Education y Current Index to Journals in Education, siendo la segunda la que recoge el vaciado de artículos de revistas que ha motivado que ERIC se encuentre en esta selección. Además, el centro encargado de esta base posibilita la reproducción de los artículos contenidos en la misma.

Article@INIST

Institut de l'Information Scientifique et Technique (INIST), CNRS

<http://services.inist.fr/public/re/cons1t.htm>

El INIST es la institución francesa de mayor nivel en materia de investigación, lo cual implica que posee un fondo documental ingente y de gran calidad. Este centro pone a

disposición del usuario sus catálogos de monografías, publicaciones periódicas y artículos de revistas. Su temática es enciclopédica, pero posee un porcentaje muy elevado de documentos relacionados con la biblioteconomía y la documentación, que son los contenidos en la base de datos conocida como Pascal. La base de artículos sobrepasa los cinco millones de documentos, en su mayor parte posteriores a 1992, aunque hay una cantidad significativa de artículos anteriores a esa fecha. La consulta puede hacerse de forma sencilla o experta, los resultados muestran una descripción completa de cada artículo, pero sin adjudicar materias. Para conseguir resultados exhaustivos y pertinentes, el sistema permite emplear distintos operadores y modificadores. Muchos artículos pueden ser consultados directamente en formato pdf si se cuenta con la autorización correspondiente; en todo caso, el INIST posee un servicio de suministro de documentos para servir los artículos originales de todas las referencias que se recogen en su base.

UNCOVER

The UnCover Company

<http://uncweb.carl.org/>

UnCover es una de las bases de datos más amplias de todo el mundo. Vacía cerca de 18.000 revistas de todos los campos del conocimiento, lo que ha dado lugar a que su número de registros se acerque a los nueve millones. Un dato importante es que todas las revistas contenidas en esta base están en inglés, otro dato es que la base responde a fines comerciales, ya que la entidad productora es un servicio de acceso al documento. Para consultar la base se solicita un perfil de usuario registrado, pero de todas formas puede realizarse una consulta sin identificarse durante un tiempo limitado. Para ello debe escogerse la primera opción Search Uncover Now en la página de búsquedas en la base. El formulario de consulta permite escoger el campo en el que se recuperará la información, así como delimitar los años de la búsqueda. En la información que se ofrece de cada registro se incluye un resumen del artículo y el precio y las condiciones de suministro del mismo. A pesar de ser un servicio comercial, esta base es de gran utilidad para la investigación debido a la gran cantidad de documentos que contiene.

UNESBIB / UNESDOC

UNESCO

<http://unesdoc.unesco.org/ulis/unesbib.html>

<http://unesdoc.unesco.org/ulis/ged.html>

La Unesco dispone de un gran número de bases de datos, de las cuales algunas están accesibles en línea de forma gratuita. Las que aquí se comentan recogen documentos de temáticas dispares, entre las que se encuentran la biblioteconomía y documentación. Debido a las funciones de esta institución y a la actividad que desarrolla en América Latina, estas bases contienen documentos de interés para el profesional de la información español y latinoamericano. UNESBIB equivale a la base de datos de la biblioteca de la Unesco, en la cual se encuentran vaciados artículos de publicaciones periódicas profesionales, además de informes, actas de congresos, monografías, etc. UNESDOC es la base de los documentos de texto completo, entre los que

se encuentran informes, estudios, etc., muchos de ellos relacionados con las bibliotecas. La primera cuenta con una cifra cercana a los cien mil documentos de todos los campos del saber, especialmente los relacionados con la actividad de la Unesco, entre los que se encuentra el desarrollo de las bibliotecas. Ambas se interrogan mediante un formulario que permite concretar al máximo la búsqueda. Es especialmente relevante el número de documentos profesionales producidos en América Latina, lo que hace de esta base una herramienta útil para el investigador y el profesional español.

Angela Sorli Rojo (CINDOC, CSIC)

José Antonio Merlo Vega (Universidad de Salamanca)

SISTEMAS DE RECUPERACION DE INFORMACION DISTRIBUIDA EN INTERNET. UNA REVISION DE SU EVOLUCION, SUS CARACTERISTICAS Y SUS PERSPECTIVAS. TERCERA PARTE

8 Sistemas distribuidos con indización asistida

Desde el punto de vista arquitectónico, directorios y bases de datos de recopilación automática plantean el problema de la multiplicidad de accesos simultáneos a servidores centralizados. Desde el punto de vista del análisis y la recuperación del contenido, representan los extremos opuestos de un continuum que separa la representación indiscriminada de contenidos con falta de atención a las estructuras de los documentos y, por otra parte, los sistemas que estructuran en registros *sui generis* el contenido de los recursos distribuidos. Al énfasis en la exhaustividad, la cobertura y la rapidez de actualización de los sistemas de recopilación automática se contraponen la selectividad, la búsqueda de precisión y la «burocratización» de los directorios y esquemas clasificatorios. La entrega a procedimientos automáticos frente al conservadurismo del análisis humano.

La búsqueda del equilibrio entre ambos extremos parece estar representada por un tercer modelo que combina, por un lado, arquitecturas distribuidas y, por otra parte, aprovecha descripciones normalizadas de los recursos producidos no por el propio sistema, sino por los mismos distribuidores de los documentos.

Los sistemas Archie y VERONICA, asociados inicialmente a los protocolos FTP y Gopher, y Harvest y ALIWEB (Archie Like Indexing of the WEB), basados en los primeros y en el sistema WAIS, fueron sólo el preludio de la tendencia creciente hacia el control de contenidos de los recursos distribuidos a través de Internet, actualmente potenciada gracias a la Iniciativa sobre Metadatos.

Las descripciones de Harvest (48) y ALIWEB (49) parten criticando la innecesaria sobrecarga en servidores y conexiones producida por los sistemas de recopilación automática. También el excesivo tráfico originado por la propia popularidad de estos sistemas, así como su dificultad de tratar con formatos informativos heterogéneos. Su alternativa para la recuperación de recursos se basa en la indización en los servido-

res de origen, la existencia de descripciones normalizadas de cada recurso, una arquitectura distribuida de las bases de datos recopiladas y el empleo de programas cliente de consulta y recuperación. Estas características están directamente inspiradas por el concepto de WAIS (Wide Area Information Server), desarrollado a finales de los años 80 por Brewster Kahle.

WAIS era un sistema en que múltiples bases de datos especializadas se distribuían en servidores dispersos controlados por un directorio, y cuyos contenidos eran accesibles y recuperables mediante el empleo de programas cliente. Los usuarios obtenían una lista de las bases de datos y, en respuesta a una expresión de búsqueda dirigida a una base de datos seleccionada, se accedía a los servidores que la contenían. Como resultado, se obtenía una descripción de los textos y la posibilidad de obtener completos los documentos.

Aunque el propio Kahle se refería a WAIS como una «herramienta de Internet para la búsqueda de información» (50), declaraba que consideraba Internet como un medio de distribución de información y se refería a los organismos públicos, los editores, las bibliotecas y las empresas dispersas como sus principales mercados. Por otra parte, ninguna de las características que en su opinión diferenciaban WAIS de los sistemas de recuperación tradicionales se han mantenido: la interfaz amigable que acepta expresiones de búsqueda en lenguaje natural, las posibilidades de búsqueda booleana y limitación por campos destinadas al usuario avanzado y el feedback de relevancia se han ido incorporando en mayor o menor medida a los sistemas de recopilación automática y aún a los directorios.

Archie, que siguió cronológicamente al sistema WAIS a principios de 1992, permitía la recuperación por palabras clave de ficheros informáticos, de los cuales controlaba 2 millones en 1994. En noviembre de 1992, se anunció VERONICA, desarrollado en la Universidad de Reno, y que, en respuesta a una búsqueda sobre menús Gopher, ofrecía otro menú de resultados. En enero de 1995, VERONICA indizaba 5.057 servidores y dos meses antes su índice incluía unos 15 millones de ítems (51). En ambos casos existía un esquema para la descripción de los recursos, aunque a veces se tratara de una simple línea (52).

8.1 Formatos normalizados y metadatos

El entronque de WAIS con Harvest, ALIWEB y otros sistemas, se basa en el hecho de emplear una representación estructurada de los documentos y de las transacciones. En el caso de WAIS, el esquema correspondía a extensiones de la norma Z39.50 (Information Retrieval Service Definition and Protocol Specification for Library Applications) de NISO (53). Por su parte, Harvest, distribuido a partir del 9 de noviembre de 1994 y cancelado el 20 de agosto de 1998, empleaba el Summary Object Interchange Format (SOIF) (54). El objeto de este formato, a modo de tabla atributo-valor, era tanto unificar la representación de la información recopilada como comprimirla de forma que su transmisión resultara más eficiente. Otro tanto hacía ALIWEB con su empleo del IAFA (Internet Anonymous FTP Archives), definido por el correspondiente grupo de trabajo de la Internet Engineering Task Force. Por su parte, el Proyecto de Biblioteca Digital de la Stanford University emplea también SOIF en su propuesta sobre recuperación. Más concretamente, para corregir los de-

sajustes existentes en los sistemas de búsqueda múltiple (metabuscadores) con el cálculo de relevancia que combina los diferentes algoritmos de ordenación de los sistemas individuales (55).

El esquema de funcionamiento de todos los sistemas de este grupo pasaba por el consenso entre los servicios de recuperación y los proveedores de información acerca del modo de representar el contenido de los documentos. Y es precisamente la existencia de este consenso la que permite delimitar una trayectoria coherente que enlaza WAIS, un servicio no estrictamente originado en Internet, con los iniciales Archie y VERONICA y los sistemas basados en HTTP como Harvest y ALIWEB. Este enfoque ha sido ampliamente revisado por Rachel Heery (56). Se debe tener en cuenta, sin embargo, que todos estos sistemas exigían el acuerdo previo entre proveedores y servicios de recuperación. Algo impensable si lo que se pretende es un control de contenidos de todos los documentos distribuidos a través de Internet. De hecho, las exigencias de mantenimiento y actualización de los sistemas hasta ahora mencionados han provocado su progresivo estancamiento o su desaparición.

La continuación de esta línea está representada por el empleo de metadatos para la descripción coherente de los contenidos de los documentos distribuidos. Pero atribuidos por los creadores de los propios documentos.

«Dado que Internet contiene más información de la que catalogadores, indizadores y elaboradores de resúmenes pueden gestionar con los medios y sistemas disponibles, se llegó al acuerdo de que una alternativa razonable para obtener metadatos de los recursos electrónicos sería proporcionar a los autores y proveedores de información medios para que describieran los recursos por sí mismos» (57).

Esta afirmación de Stuart Weibel figura en el documento resultante del primer encuentro sobre el control de contenidos de recursos distribuidos en Internet mediante metadatos, celebrado en Dublin (Ohio) del 3 al 5 de marzo de 1995. Desde entonces, la Dublin Core Metadata Initiative se ha ido configurando como la plataforma y una de las tendencias más prometedoras para una recuperación eficiente de información distribuida en Internet.

En el mismo documento, Weibel emplea la metáfora del continuum y propone *«como solución alternativa que equilibre ambos extremos, la creación de registros que resulten más informativos que los índices pero sean menos completos que un registro catalográfico formal. Si sólo se requiere un pequeño esfuerzo para crear tales registros, se podrían describir más objetos, especialmente si los autores de los recursos se animan a crear la descripción. Si, por otra parte, dicha descripción sigue una norma establecida, sólo la creación del registro requeriría intervención humana, mientras que el descubrimiento y la recopilación de recursos se podrían efectuar mediante herramientas automatizadas».*

Aunque proliferan las definiciones del concepto, cabe describir los metadatos como valores que se presentan asociados a su carga semántica, expresada por la unión entre un elemento estructural (autor, título, fecha...) y las correspondientes variables. El conjunto inicial se limitaba a identificar el significado de un grupo de elementos descriptivos con objeto de mejorar la detección de recursos en el espacio Web que se pudieran considerar objetos similares a documentos (Document-Like-Objects). El resultado del segundo seminario fue la adopción del Esquema de Warwick (Warwick Framework), un modelo conceptual de una arquitectura de contenido para paquetes de metadatos de diversos tipos. En el tercer seminario se exten-

dió el esquema para la descripción de imágenes y, poco después, el conjunto inicial de 13 elementos se extendió a 15, de los que existe traducción española (58).

Internet no cuenta, como ya se ha visto, con un universo de proveedores de información cualificados y expertos en campos temáticos concretos. Así que, si sólo se produjera una normalización de elementos y ciertas instrucciones sintácticas, la situación sería tan caótica como la que se derivaría de sistemas de indización tradicionales que se basaran únicamente en las palabras clave asignadas por autores noveles o remedaría la problemática de los sistemas de recopilación automática. Afortunadamente, el seminario celebrado en Camberra (3 a 5 de marzo de 1997) vino a incorporar al Dublin Core el calificador esquema y el subelemento tipo (59). Mediante su aplicación es posible despejar ciertas ambigüedades que se podrían plantear a los autores de los documentos y que, de hecho, son habituales en la catalogación e indización por profesionales.

De hecho, los sucesivos desarrollos del esquema inicial suponen un grado de normalización progresivamente mayor. Así, a la lista consensuada de elementos de descripción viene a añadirse la normalización de valores, que aprovechan esquemas preestablecidos. Además, la ambigüedad queda también reducida por el empleo de subelementos tipo.

8.2 Los requisitos del modelo de indización asistida

El énfasis en la necesidad de consenso de este modelo, reiterado en los párrafos anteriores, no es casual. Para garantizar un funcionamiento eficiente de este modelo es necesario, en primer lugar, el acuerdo de los grandes proveedores de información, por ejemplo los grandes editores del sector electrónico. En segundo lugar, es preciso que se produzca de forma generalizada la utilización del conjunto de datos, sus calificadores y su sintaxis. En tercer lugar, se requiere que las sucesivas versiones de los lenguajes de marcas acojan el esquema vigente y sus desarrollos. Por último, los programas de recopilación automática y los sistemas de descripción de los directorios y esquemas deben de reconocer y aceptar el valor de los metadatos.

El primer requisito no sólo se cumple, sino que los grandes proveedores se cuentan entre los primeros impulsores de la idea de utilizar un formato normalizado como mecanismo simple para representar los metadatos: así, el comunicado del 8 de septiembre de 1997 que anunciaba el apoyo de NetScape a la iniciativa Resource Description Framework, mencionaba entre los impulsores a nombres de peso en el sector de la edición electrónica. Entre ellos, CNN, CBS, Time Inc o Knight-Ridder eran sólo algunos nombres (60). Por otra parte, el 3 de octubre siguiente se produjo el anuncio del primer borrador público de esta iniciativa, con el apoyo de nombres no menos relevantes (61), que fue presentado oficialmente en la quinta reunión de la Dublin Core Initiative una semana más tarde.

La adopción generalizada del esquema por parte de los productores iniciales de información depende de que los redactores de páginas en lenguaje de marcas o los usuarios de programas para su confección cuenten con facilidades para la inclusión de metainformación en ellas. Por otra parte, no se puede descartar el empleo de procedimientos que, tras el análisis automático del código de cada documento, puedan

generar una lista normalizada de valores (62). Ya existen, por añadidura, proyectos y sistemas automatizados que transforman los diversos formatos normalizados de descripción y los valores de catalogación al esquema de metadatos (63, 64). Cabe destacar en esta línea, por su elegancia y simplicidad, el procedimiento propuesto por Massimo Marchiori, basado en la «propagación» de los valores de los metadatos de unos a otros documentos en función del grado de interconexión entre ellos, que sirve de base para un cálculo borroso (fuzzy) de su similitud de contenido (65).

Las reservas sobre el empleo de cualificadores y del subelemento tipo manifestadas en la reciente reunión del grupo de indización de RedIRIS (66) ilustran la necesidad de ajuste entre los lenguajes de marca y la expresión de los metadatos. HTML, XML o cualquier subconjunto del SGML deben posibilitar la inclusión de los conjuntos de metadatos en los documentos y su renacimiento. Afortunadamente, la simbiosis entre RDF y XML permite despejar dudas sobre la capacidad para albergar datos normalizados según el Dublin Core en la redacción de nuevas páginas. Persiste, sin embargo, la duda sobre los millones de documentos ya distribuidos en el espacio Web y su redacción.

Los sistemas de recopilación automática más popularizados (Alta Vista, Excite, HotBot, Infoseek, Lycos, WebCrawler y Northern Light) aceptan etiquetas META en la indización de páginas Web. Sin embargo, Lycos y Northern Light no los emplean en la representación de los resultados de búsqueda y sólo HotBot e Infoseek los emplean en el cálculo de relevancia (67). De entre los sistemas españoles, Olé menciona la posibilidad de buscar entre las palabras clave pero no hace referencia explícita, al contrario que Trovator, al empleo de metadatos en la recopilación, la indización o el cálculo de relevancia.

No obstante todo lo anterior, trabajos recientes ponen el dedo en la llaga del coste de la incorporación de metadatos a las páginas generadas en entornos académicos, comerciales u oficiales. Sus razonamientos se basan en tres hechos: la necesidad de intervención humana (y el consiguiente coste) en la elaboración de las etiquetas, el esfuerzo añadido que supone la incorporación de metadatos a la generación del contenido que, en definitiva, suponen las páginas y el hecho de que, en el sector comercial, el retorno de la inversión realizada, en forma de número de accesos tras la localización de páginas, no parece justificar la inversión requerida (68).

9 Bibliografía

48. *Technical Discussion of the Harvest System*, 9 de septiembre de 1998, <<http://harvest.transarc.com/public/trg/Harvest/technical.html>>.
49. *History of ALIWEB*, 9 de septiembre de 1998, <<http://www.nexor.com/aliweb/doc/history.html>>.
50. UBOIS, J. Wide Area View: An interview with WAIS Inc CEO Brewster Kahle. *Internet World*, 6 (9), September, 1995, <<http://www.internetworld.com/print/monthly/1995/09/feat74.htm>>, 19 de octubre de 1998.
51. MCMURDO, G. How Internet was indexed. *Journal of Information Science Electric Writing*, 21 (6), 3 de julio de 1997, <<http://www.qmced.ac.uk/cis/staff/cimmu/jisew/ewv21n6/default.htm>>.
52. MAÑAS, J. A. Búsqueda y recuperación de información en Internet. *Novática*, 110, 75-81, julio-agosto de 1994.

53. KAHLE, B. *Wide Area Information Server Concepts*, 19 de octubre de 1998, <<http://www.alexia.com/~brewster/essays/wais-concepts.htm>>.
54. *The Harvest Summary Object Interchange Format (SOIF)*, 2 de noviembre de 1998, <<http://harvest.transarc.com/Harvest/brokers/soifhelp.html>>.
55. GRAVANO, L.; CHANG, K.; GARCIA-MOLINA, H.; LAGOZE, C.; PAEPCKE, A. *STARTS, Stanford Protocol Proposal for Internet Retrieval and Search*, 19 de enero de 1997, <<http://www-db.stanford.edu/~gravano/starts.html>>.
56. HEERY, R. Review of Metadata Formats. *Program*, 30 (4), 345-373, October 1996, <<http://www.ukoln.ac.uk/metadata/review.html>>, 4 de noviembre de 1998.
57. WEIBEL, S. Metadata: The Foundations of Resource Description. *D-Lib Magazine*, July 1995, <<http://www.cnri.reston.va.us/home/dlib/July95/07weibel.html>>, 2 de noviembre de 1998.
58. MASSA, J. *Elementos del conjunto de metadatos de Dublin Core: Descripción de Referencia*, creado el 11 de diciembre de 1997, 13 de mayo de 1998, <http://www.rediris.es/metadata/dublin_core_elements.es.html>.
59. WEIBEL, S.; IANNELLA, R.; CATHRO, W. The 4th Dublin Core Metadata Workshop Report. *D-Lib Magazine*, June 1997, <<http://www.dlib.org/dlib/june97/metadata/06weibel.html>>, 2 de noviembre de 1998.
60. NETSCAPE COMMUNICATIONS. *Netscape works with W3C and leading content providers to drive new specification for organizing, describing and navigating information on Internet, intranets and desktops*, 8 de septiembre de 1997, <<http://home.netscape.com/flash1/newsref/pr/newsrelease488.html>>.
61. *World Wide Web Consortium Publishes Public Draft of Resource Description Framework (RDF)*, 3 de octubre de 1997, <<http://www.w3.org/Press/RDF>>.
62. LOPEZ, D. R.; MASSA, J. Dando forma al envase y, con ello, al contenido: Webber. *Boletín de RedIRIS*, 45, 15-21, octubre de 1998, <<http://www.rediris.es/rediris/boletin/45/enfoque1.html>>.
63. SAN SEGUNDO MANUEL, R. Organización del conocimiento en Internet: Metadatos bibliotecarios Dublin Core. *Sextas Jornadas Españolas de Documentación Automatizada*, Valencia, 29 a 31 de octubre de 1998, vol, 2, 805-818.
64. WAUGH, A. Specifying metadata standards for metadata tool configuration. *Seventh International World Wide Web Conference*, Brisbane, 14 a 18 de abril de 1998, <<http://www7.scu.edu.au/programme/fullpapers/1913/com1913.htm>>, 27 de noviembre de 1998.
65. MARCHIORI, M. The limits of Web metadata, and beyond. *Seventh International World Wide Web Conference*, Brisbane, 14 a 18 de abril de 1998, <<http://www7.scu.edu.au/programme/fullpapers/1896/com1896.htm>>, 27 de noviembre de 1998.
66. MASSA, J. *Resumen de la reunión IRIS.SEARCH*, 7 de mayo de 1998, <<http://rediris.es/si/iris-index/coord/gt5/resumen.txt>>.
67. SULLIVAN, D. *Search Engine Watch: Search Engines Features*, 1 de diciembre de 1998, <<http://searchenginewatch.com/webmasters/features.html>>.
68. THOMAS, C. F.; GRIFFIN, L. S. Who will create the metadata for the Internet. *First Monday*, 2(12), December 1998, http://www.firstmonday.dk/issues/issue3_12/thomas/index.html.

C. Benito Amat

Unidad de Documentación.

Radiotelevisión valenciana