

EL PROFESIONAL DE LA INFORMACIÓN

Revista científica y profesional en español
líder en el área de documentación, bibliotecas
y nuevas tecnologías de la información.



Observatorio

La ciencia y nosotros

Luis-Javier Martínez

Artículos

Las universidades públicas españolas en
Google Scholar: presencia y evolución de
su publicación académica web

Enrique Orduña-Malea, Jorge Serrano-Cobos y Nuria
Llorel-Romero

Google Scholar como herramienta para la
evaluación científica

Daniel Torres-Salinas, Rafael Ruiz-Pérez y Emilio
Delgado-López-Cózar

Utilización de las bases de datos de
patentes como instrumento de vigilancia
tecnológica

Antonio Hidalgo-Nuchera; Santiago Iglesias-Pradas y
Ángel Hernández-García

Análisis

La base de datos *ISOC* como sistema de
información y fuente para el análisis de las
ciencias humanas y sociales en España

Teresa Abejón-Peña, Angeles Maldonado-Martinez,
Luis Rodríguez-Yunta y María-Cruz Rubio-Liniers

Retrieval of very large numbers of items in
the *Web of Science*: an exercise to develop
accurate search strategies

Ricardo Arencibia-Jorge, Loet Leydesdorff, Zaida
Chinchilla-Rodríguez, Ronald Rousseau and Soren
W. Paris

Estrategia para mejorar la difusión de los
resultados de investigación con la Web 2.0

Daniel Torres-Salinas y Emilio Delgado-López-Cózar

Meridià: un portal para la difusión de
la ciencia desde una óptica integral y
cooperativa

Llorenç Arguimbau, Aldara Cervera, Robert Latorre
y Mercè Martí

Servicio público de televisión y patrimonio
audiovisual: el proyecto *VideoActive*

David Fernández-Quijada y Montse Fortino

Entrevista

OSTI-DOE, la fuente de información sobre
energía. Entrevista a Karen Spence

Fernanda Peset y Tomàs Baiget



INFORMACIÓN CIENTÍFICA & TÉCNICA

09:32 AM - SEARCH - ISI THOMSON/SCOPUS



El profesional de la información

Revista bimestral fundada en 1992 por Tomàs Baiget y Francisca García-Sicilia

REDACCIÓN:

El Profesional de la información

Apartado 32.280

08080 Barcelona

epi@elprofesionaldelainformacion.com

PUBLICIDAD:

Tel.: +34-609 352 954

publici@elprofesionaldelainformacion.com

SUSCRIPCIONES:

El profesional de la información

Apartado 32.280

08080 Barcelona, España

suscripciones@elprofesionaldelainformacion.com

http://www.elprofesionaldelainformacion.com/suscripciones.html

Teléfono de atención al suscriptor
+34 609 352 954

SERVICIOS ONLINE:

Maria T. Moreno

mt.moreno@ono.com

DISEÑO:

MASmedios

Director artístico:

Moisés Mañas

MAQUETACIÓN:

Jorge Liras

Romargraf, S.A.

PRODUCCIÓN e IMPRESIÓN:

Romargraf, S.A.

Juventut, 55-57

08904 L'Hospitalet de Ll.

Tel. +34-933 345 466

romargraf@romargraf.es

DISTRIBUCIÓN ONLINE:

MetaPress, Alabama, EUA

http://elprofesionaldelainformacion.com

metapress.com

Depósito legal: B-12303-97

Los trabajos de la sección "Artículos" son aprobados según el sistema tradicional "peer review": al menos dos expertos en el tema, del consejo asesor de la revista y/o externos, deben dar el visto bueno antes de su publicación.

Para conseguir que los trabajos no pierdan su actualidad, la dirección y los evaluadores de esta revista ponen especial esfuerzo en revisar los artículos con gran rapidez, consiguiendo un tiempo medio de aceptación o rechazo de los trabajos de sólo unas pocas semanas.

Dirección editorial:

Tomàs Baiget EPI SCP
http://www.baiget.com

Subdirector:

Javier Guallar Universitat de Barcelona
jguallar@gmail.com

Coordinador editorial:

Carlos Tejada-Artigas Universidad Complutense de Madrid
tejada@ccdoc.ucm.es

Redactores:

Lluís Codina Universitat Pompeu Fabra
http://www.lluiscodina.com

Ricardo Eito Grupo GMV
reito@gmv.es

Elea Giménez-Toledo Instituto de Estudios Documentales sobre Ciencia y Tecnología
elea.gimenez@cchs.csic.es

Javier Leiva-Aguilera Catorze.com
http://www.javierleiva.info

Toon Lowette Grid Electronic Publishing
toon@grid.be

Roser Lozano Biblioteca Pública de Tarragona
rlozano@gencat.net

José-Antonio Millán Libros y bitios
http://jamillan.com

Isabel Olea Universidad de León
isabel.iolea@gmail.com

Fernanda Peset Universidad Politécnica de Valencia
mpesetm@upv.es

Jorge Serrano-Cobos MASmedios
jorgeserrano@gmail.com

Daniel Torres-Salinas Universidad de Navarra
torressalinas@gmail.com

Revisión de lengua inglesa:

Elaine M. Lilly Writer's First Aid
elaine@writersfirstaid.com

CONSEJO ASESOR

Ernest Abadal
Facultat de Biblioteconomia i Documentació. Universitat de Barcelona. Barcelona.

Isidro F. Aguillo
Centro de Ciencias Humanas y Sociales. Consejo Superior de Investigaciones Científicas (Csic). Madrid.

Ramon Alberch
Subdirector General de Archivos Generalitat de Catalunya. Barcelona.

Adela d'Alòs-Moner
Docó. Barcelona.

Ricardo Baeza-Yates
Depto. de Ciencias de la Computación. Univ. de Chile. Santiago. Chile. Yahoo! Research, Barcelona.

Carlos Benito Amat
Servicio de Biblioteca y Documentación Científica. Instituto de Agroquímica y Tecnología de Alimentos, Csic. Burjassot. Valencia.

Jesús Bustamante
Biblioteca, CEDEFOP, Salónica, Grecia.

Carlota Bustelo
Inforárea. Madrid.

Emilio Delgado López-Cózar
Facultad de Biblioteconomía y Documentación. Universidad de Granada. Granada.

Javier Díaz-Noci
Depto. de Comunicación. Universidad Pompeu Fabra, Barcelona

Assumpció Estivill
Facultat de Biblioteconomia i Documentació. Universitat de Barcelona. Barcelona.

Antonia Ferrer-Sapena
Universidad Politécnica de Valencia. Valencia.

Fco. Javier García Marco
Depto. de Ciencias de la Documentación e Historia de la Ciencia. Universidad de Zaragoza. Zaragoza.

Paola Gargiulo
Consorzio per le Applicazioni di Supercalcolo per Università e Ricerca. (Caspur), Roma, Italia.

Johannes Keizer
Food and Agriculture Org. (FAO) United Nations, Roma, Italia

Thomas Krichel
Palmer School of Libr. & Inform. Sci. Long Island Univ., New York, USA

Victoria Manglano
Ovid Technologies, Madrid.

Mari-Carmen Marcos
Depto. de Comunicación. Universidad Pompeu Fabra, Barcelona

Pere Masip
Facultat de Comunicació Blanquerna Universitat Ramon Llull, Barcelona

Charles McCarthieNeville
Opera Software, Oslo, Norway

Joan Roca
Dean of Library Services
Minnesota State University, USA

Robert Seal
Loyola University Chicago
Evanston, Illinois, USA

Ernesto Spinak
Consultor, Montevideo, Uruguay.

Jesús Tramullas
Depto. de Ciencias de la Documentación e Historia de la Ciencia. Universidad de Zaragoza. Zaragoza.

Tema central: Información científica y técnica

OBSERVATORIO

- 485 **La ciencia y nosotros**
Luis-Javier Martínez

ARTÍCULOS

- 493 **Las universidades públicas españolas en Google Scholar: presencia y evolución de su publicación académica web**
Enrique Orduña-Malea, Jorge Serrano-Cobos y Nuria Lloret-Romero
- 501 **Google Scholar como herramienta para la evaluación científica**
Daniel Torres-Salinas, Rafael Ruiz-Pérez y Emilio Delgado-López-Cózar
- 511 **Utilización de las bases de datos de patentes como instrumento de vigilancia tecnológica**
Antonio Hidalgo-Nuchera; Santiago Iglesias-Pradas y Ángel Hernández-García

ANÁLISIS

- 521 **La base de datos ISOC como sistema de información y fuente para el análisis de las ciencias humanas y sociales en España**
Teresa Abejón-Peña, Angeles Maldonado-Martinez, Luis Rodríguez-Yunta y María-Cruz Rubio-Liniers
- 529 **Retrieval of very large numbers of items in the Web of Science: an exercise to develop accurate search strategies**
Ricardo Arencibia-Jorge, Loet Leydesdorff, Zaida Chinchilla-Rodríguez, Ronald Rousseau and Soren W. Paris
- 534 **Estrategia para mejorar la difusión de los resultados de investigación con la Web 2.0**
Daniel Torres-Salinas y Emilio Delgado-López-Cózar
- 540 **Meridià: un portal para la difusión de la ciencia desde una óptica integral y cooperativa**
Llorenç Arguimbau, Aldara Cervera, Robert Latorre y Mercè Martí
- 545 **Servicio público de televisión y patrimonio audiovisual: el proyecto VideoActive**
David Fernández-Quijada y Montse Fortino
- 552 **Evolución de los servicios de redes sociales en internet**
Marcos Ros-Martín

INDICADORES

- 559 **Agregación de índices bibliométricos para evaluar la producción científica de los investigadores**
Francisco Herrera, Enrique Herrera-Viedma, Sergio Alonso y Francisco-Javier Cabrerizo

ENTREVISTA

- 563 **OSTI-DOE, la fuente de información sobre energía. Entrevista a Karen Spence**
Fernanda Peset y Tomàs Baiget

RESEÑAS

- 567 **Manual de indización: teoría y práctica (Isidoro Gil-Leiva)**
Adolfo Alonso-Arroyo
- 569 **Prácticas de las bibliotecas universitarias argentinas (Mercedes Patalano)**
L.-Fernando Ramos-Simón

AGENDA

- 572 **AGENDA**
- 576 **INFORMACIÓN PARA LOS AUTORES**

Bases de datos

- Academic search complete (Ebsco)**
<http://www.ebscohost.com/thisTopic.php?marketID=1&topicID=633>
- Academic search premier (Ebsco)**
<http://www.ebscohost.com/thisTopic.php?marketID=1&topicID=1>
- Bedoc (Inforárea)**
<http://www.inforarea.es/bedoc.htm>
- Compludoc (Universidad Complutense de Madrid)**
<http://europa.sim.ucm.es/compludoc/>
- Dialnet (Universidad de La Rioja)**
http://dialnet.unirioja.es/servlet/revista?&clave_revista=469
- Francis (Inist)**
<http://www.inist.fr/revuesPF.php3?search=profesional>
- ISI Social science citation index, Social SCI, WoS (Thomson Reuters)**
<http://go.isiprproducts.com/> Impact Factor 2008 = 0,400
- Inspec, Information services in physics, electronics and computing (IET, The Institution of Engineering and Technology)**
<http://www.theiet.org/publishing/inspec/>
- ISOC, Índice español de ciencias sociales y humanidades (Iedcyt)**
<http://bddoc.csic.es:8080/ver/ISOC/revi/0721.html>
- ISTA, Information science and technology abstracts (Ebsco)**
<http://www.epnet.com/thisTopic.php?topicID=91&marketID=1>
- Lisa, Library and information science abstracts (CSA)**
<http://www.csa.com/factsheets/lisa-set-c.php>
- Lista, Library, information science & technology abstracts (Ebsco)**
<http://www.libraryresearch.com>
- Pascal (Inist)**
<http://www.inist.fr/revuesPF.php3?search=profesional>
- Scopus (Elsevier)**
<http://www.scopus.com>

Catálogos y servicios bibliotecarios

- Argos-Bolsum (GVA, Generalitat Valenciana)**
http://www1.pre.gva.es/argos/es/contenido_general/recursos/bolsum/
- Cbuc, Consorci de Biblioteques Universitàries de Catalunya**
<http://sumaris.cbuc.es/13866710.htm>
- DoIS, Documents in information science**
<http://wotan.liu.edu/dois/data/julqichq.html>
- Google Scholar**
<http://scholar.google.com>
- In-Recs, Revistas españolas de ciencias sociales (Grupo EC3, Evaluación de la Ciencia y de la Comunicación Científica, Universidad de Granada)**
<http://ec3.ugr.es/in-recs/Biblioteconomia.htm>
- Registros bibliográficos para bibliotecas públicas españolas (Rebeca)**
<http://www.mcu.es/bibliotecas/MC/Rebeca/>
- Universidad de Chile**
<http://www.al-dia.cl/sistema/tablas/listar.asp?r=3199>

Acceso a los textos completos

- MetaPress (2000-)**
<http://elprofesionaldelainformacion.metapress.com/>
- Ebscohost Electronic Journals Service (2000-embargo 1 año)**
<http://ejournals.ebsco.com/direct.asp?JournalID=105302>
- Library, information science & technology abstracts with full text (2000-embargo 1 año)**
<http://www.ebscohost.com/thisTopic.php?marketID=1&topicID=584>
- El profesional de la información (1992-embargo 2 años)**
<http://elprofesionaldelainformacion.com/contenidos.html>
- SwetsWise (2000-)**
<https://www.swetswise.com/>

EPI en Facebook

- <http://www.facebook.com/group.php?gid=36050316757>

EPI en Twitter

- http://twitter.com/revista_EPI

Plataforma de producción OJS

- Recyt, Repositorio español de ciencia y tecnología (Fecyt)**
<http://recyt.fecyt.es/index.php/EPI>



Rapidez, precisión, flexibilidad

En el siglo XXI, más que nunca, la información es poder. Ayudar a empresas e instituciones a gestionar y rentabilizar su caudal informativo, tanto el que poseen como el que genera diariamente su actividad, es el objetivo y la razón de ser de Baratz.

www.baratz.es



91 456 03 60 | informa@baratz.es

Raimundo Fernández Villaverde 28, 1.º 28003 Madrid

La ciencia y nosotros

Por Luis-Javier Martínez

Resumen: La ciencia, el segmento de la cultura más depurado para describir y manejar el entorno, resulta de una algoritmia evolutiva constante y es una ventaja biológica adaptativa de la especie humana. La ciencia evoluciona socialmente sobre registros físicos y medios de información cuyos cambios influyen en la ciencia y en la sociedad. Los documentalistas han venido gestionando estos medios o contenedores del conocimiento. La revolución tecnodigital crea una sociedad de la información industrializada donde la ciencia es una fuerza productiva, tiene carácter neurodigital y forma empresas colectivas. Los contenedores de la ciencia son ahora transitorios, plurales, segmentados, reticulares y se “empresarializan”: constituyen “plataformas informativas específicas”. La posición de los profesionales de la información es débil, pero tienen la posibilidad de convertirse en ingenieros de la información, de intervenir en determinadas plataformas y de competir individualmente gestionando información científica.

Palabras clave: Ciencia, Evolución, Profesionales de la información.

Title: Science and us

Abstract: Science is the most refined cultural element to describe and manage the environment. It is the result of a permanent evolutionary algorithm and represents an adaptive biological advantage for the human species. Science evolves socially through physical records and information media, whose changes influence both science and society. Information professionals have been managing these physical records or knowledge containers. The technodigital revolution grows an industrialized information society where science is a productive workforce, with a neurodigital nature, and develops as a collective enterprise. Science containers are now very changeable, plural, segmented, networking, and take on a business look: they become “specific information platforms”. Information professionals’ position is rather weak, but they have the possibility to turn into information engineers, to participate in some of those platforms and to compete individually as information managers in the science arena.

Keywords: Science, Evolution, Information professionals.

Martínez, Luis-Javier. “La ciencia y nosotros”. *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 485-491.

DOI: 10.3145/epi.2009.sep.01



Luis-Javier Martínez es licenciado en filosofía (lógica y teoría de la ciencia) por la Universidad de Salamanca y cursó estudios en la Escuela de Documentalistas (Madrid). Ha sido subdirector y director de la Biblioteca de la Universidad de Cantabria. Actualmente es jefe del área de ciencia y tecnología en dicho organismo y redacta el blog Observatorio de Martinej.

Ciencia como algoritmia

MUCHAS FUENTES, NO SÓLO WIKIPEDIA, vinculan la palabra “algoritmo” a una obra del científico musulmán al-Jwarizmi, que vivió en Persia entre los siglos VIII y IX. La Real Academia Española da una etimología diferente, pero su definición de “algoritmo” como “un conjunto ordenado y finito de operaciones que permite hallar la solución de un problema” es de amplia aceptación. En origen las operaciones se referían a números arábigos, al cálculo aritmético. Pero la noción de algoritmo se ha extendido al tratamiento de símbolos algebraicos o signos de cualquier clase. Remite ahora, en general, a procesos de computación controlada de la información para ejecutar tareas o programas, por ejemplo mediante ordenadores. Hay

algoritmos de muchas clases, incluso probabilísticos y recursivos. Son muy interesantes los algoritmos evolutivos o genéticos que, en inteligencia artificial, manipulan información codificada digitalmente de manera similar a como evoluciona la información biológica registrada en forma de DNA; es decir, mediante replicación, variación y selección.

Podría verse la ciencia y los procedimientos científicos a lo largo de la trayectoria del homo sapiens como una algoritmia evolutiva de la información transmitida culturalmente (no genética): una algoritmia dirigida a representar y manipular la realidad de manera cada vez más eficiente. Se trata de una concepción no alejada de la de Hull (1988), Thagard (1993) y otros. La información científica aparece así como un segmento cre-

cientemente cualificado de la cultura, el segmento más perfeccionado y fiable como descripción del mundo.

“Los medios y tecnologías informativas son el sustrato en el que tiene lugar la evolución de la ciencia”

Las representaciones humanas del entorno han evolucionado mediante un cálculo científico que las ha manejado con lenguajes y técnicas comunicativas: signos orales, textos escritos a mano y textos impresos, hasta el nacimiento de la computación artificial. De ahí la importancia de los registros materiales de la información en la historia de la ciencia: son caldo de cultivo de la algoritmia del conocimiento en cuanto extensiones de la capacidad cerebral de la especie. Y se manifiestan plurales y cambiantes a lo largo del tiempo.

Todo esto evoca los ideales de **Leibniz** (1646-1716), filósofo, matemático y bibliotecario. Este autor se interesó por la lógica, el lenguaje o la semiótica, las dinámicas de los signos y su relación con el mundo. Entre otras cosas, propuso una *característica universalis* como lenguaje lógico susceptible de ser computado mediante máquinas de razonar para producir conocimiento fiable. En cierto sentido fue un teórico de la información *avant la lettre*, y tampoco estuvo ajeno a una visión evolucionista de la realidad.

Naturaleza y sociedad en la ciencia

Mediante la algoritmia cognitiva, y sobre la base de un cerebro y un lenguaje muy potentes, la información sobre el entorno se ha depurado entre los homo sapiens constantemente, llevando al progreso del conocimiento científico. Ello ha sido la clave para la adaptabilidad biológica de la especie ante cualquier ambiente, para su (literalmente) arrasador éxito evolutivo. La ciencia es, por tanto, un fenómeno natural y algo inexorable para los homo sapiens, predestinados a “saber más”. Hasta la osadía y el castigo prometeicos incluso, en clave mitológica.

El perfeccionamiento de la información científica ha tenido lugar gracias a la comunicación. Es fruto de la participación de muchos cerebros en interacción, a lo largo del tiempo y en cada instante. Por eso la ciencia es un fenómeno social, porque se necesita la contribución de mucha inteligencia distribuida para acometer el cálculo y asegurar la fiabilidad de los resultados. La ciencia no sólo es parte de la cultura de la sociedad, sino que su generación es estructuralmente comunicativa y pública (**Maltrás**, 2001).

La ciencia resulta, pues, a la vez, natural y social: de forma inevitable, crece, se reproduce y diversifica, y lo hace mediante instituciones y protocolos de interacción. Los medios o tecnologías informativas son el sustrato en el que tiene lugar ese progreso del conocimiento. La complejidad creciente de la investigación y la institucionalización de la ciencia han provocado que surjan, con los medios, mediadores, especialistas en facilitar la construcción y difusión del saber académico manejando sus registros físicos. Todo ha pasado por muchas etapas y variadas formulaciones. Ha habido múltiples formas de registro, publicación o intercambio entre los investigadores, y también depósitos de información científica, bibliotecas, capaces de canalizar la comunicación.

Las (r)evoluciones científicas

El espíritu científico no da saltos, aunque haya habido épocas con mayor o menor éxito en el constante proceso de mejora de la información sobre la realidad. Las discontinuidades han tenido que ver muchas veces con los medios tecnológicos y documentales gracias a los que el conocimiento se produce y reproduce; los cambios en dichos medios han propiciado fases de expansión.

“La ciencia es a la vez un fenómeno natural y social”

La ciencia, aún escasamente teórica, adquirió cierta solidez al aparecer la escritura en los antiguos estados hidráulicos de Oriente Medio. El florecimiento del saber en la Grecia clásica fue posible gracias a un eficiente sistema de escritura fonética que dinamizó la enseñanza y la erudición, más que a un “genio griego”: si aprender a escribir tomaba poco tiempo y redactar o leer escritos era asequible a cualquier persona ilustrada y despierta, se comprende que la ciencia progresara (**Solís; Sellés**, 2007, pp. 22-23).

La acumulación de conocimiento accesible con facilidad en las bibliotecas de Alejandría, Antioquía, Éfeso, Esmirna, Siracusa, Rodas, etc., creó la masa crítica conveniente para que la algoritmia de la información produjera las espléndidas manifestaciones de la ciencia helenística. El florecimiento de la ciencia en la Escuela de Chartres y la época gótica tienen como trasfondo la intensa recuperación de textos clásicos griegos desde la cultura musulmana, que revitalizó las dinámicas informativas en algunos puntos de Occidente.

Pero el ejemplo por antonomasia de discontinuidad en el avance científico es el de la ciencia moderna, en

cuyo origen está la imprenta (Eisenstein, 1979). Ésta introdujo técnicas de trabajo intelectual incomparablemente más potentes que las de la época de los manuscritos, que movilizaron de manera espectacular la producción de conocimiento, lo que fue cristalizando en concepciones novedosas y métodos fecundos: un saber más fiable, experimental, compartido, contrastado y progresivo. Disponer de abundantes copias del saber clásico o contemporáneo estimuló la comparación, crítica, contrastación y perfeccionamiento de las ideas y los datos. La nueva industria de la comunicación, y no un “espíritu de la modernidad”, provocó la efervescencia intelectual que revolucionó los saberes hasta desembocar en la ciencia del siglo XVII.

Desde entonces han ido surgiendo técnicas de comunicación cada vez más eficaces (de la imprenta manual a la mecánica, etc.) y formas de publicación especializadas (revistas científicas, etc.) que facilitaron sucesivas o continuas explosiones informativas, con el aumento exponencial en la actividad científica. En ese contexto nació la documentación moderna, un sistema de mediación y apoyo a la comunicación de los contenidos científicos. Los profesionales de la información surgieron como parte de la división del trabajo inherente a un período de gran crecimiento de la producción intelectual en el medio impreso. Sin ser expertos en los propios contenidos, la mayoría de las veces, se han dedicado a gestionar los “contenedores” y han sido útiles mediando entre científicos creadores y consumidores.

“En el medio tecnodigital crece una sociedad de la información industrializada”

La evolución sociocultural y su ritmo son resultado de la velocidad en las transferencias de información entre los homo sapiens, como sucede con las mutaciones o recombinaciones genéticas en la evolución biológica. Un aspecto particular de esa evolución ha sido la “especiación” y propagación de la ciencia como segmento diferenciado del resto de la cultura, que no provee a las poblaciones humanas, por ejemplo, de información reguladora de las conductas (a diferencia de las antiguas cosmovisiones globales). La singularización y autonomía del saber científico en la modernidad se manifiestan asimismo en instituciones propias, como las academias, universidades y sociedades científicas, y en reglas epistémicas y sociales que constituyen el *ethos* de la ciencia académica (Maltrás, 2001; Ziman, 2003).

Sociedad de la información industrializada

Si puede evocarse a **Leibniz** como referente especulativo de la ciencia en cuanto algoritmo informativo, fue en el siglo XX, en el círculo de intereses de la II Guerra Mundial, cuando la explotación de la lógica booleana mediante dispositivos electromecánicos condujo a la efectiva automatización del tratamiento y la comunicación de datos (**Turing, Von Neumann, Shannon**, etc.) En especial, **Claude E. Shannon** (1916-2001), con su teoría de la información, puso las bases conceptuales de la información industrializada e impulsó la digitalización, la difusión de los códigos, circuitos y máquinas digitales. Su trabajo hizo posible definir la información en términos matemática y operacionalmente precisos, así como medir su cantidad (bits) y mejorar de forma drástica las técnicas para su transmisión. Convirtió la información en una entidad concreta y general, en una mercancía universal tratable de manera industrial.

La consiguiente aparición y extensión de las conocidas TIC origina grandes transformaciones, al propiciar que la información se replique y prolifere de manera intensa, acelerada y masiva, en una escala sin parangón con la época anterior. La economía de los servicios y de los sectores más basados en la información se industrializa (además de globalizarse): sustentada en la infotecnología, se abre a la producción y consumo de masas, bajo esquemas de organización industrial. Crece así una sociedad de la información industrializada, y no tanto una sociedad del conocimiento contrapuesta estrictamente o superior a la sociedad industrial que la precedería.

“La ciencia al modo industrial supone una ruptura o desafío respecto a la ciencia académica”

El medio tecnodigital constituye una extensión de la capacidad cerebral de los individuos que, frente a los registros impresos, resulta casi ilimitada e instantánea (bits vs. átomos). Aporta capacidad de memoria, pero también de cálculo o computación y de comunicación o relación (y hasta de actuación sobre el ambiente): es como una protointeligencia o sistema neural, que trasciende y conecta a los individuos. Las últimas innovaciones en la industria TIC asocian a todos los homo sapiens individuales entre sí y con el propio sustrato digital en un nuevo entorno cultural, tal vez un tercer entorno, como diría **Echeverría** (1999). Parece estar surgiendo una inteligencia de enjambre: colectiva, cooperativa,

reticular, externalizada, neurodigital, *ciborg*. Una inteligencia basada no tanto en la excelencia intelectual de los individuos como en las cualidades emergentes del conjunto. No es tan insólito si se piensa que el cerebro es una red de neuronas y que la cultura oral o escrita (y sus correspondientes sociedades) también se basan en redes comunicativas, aunque más rudimentarias.

En este marco, dimensiones o subsectores informativos concretos, como el periodismo, la educación o la ciencia están experimentando cambios que, por supuesto, afectan a los profesionales de la información que trabajan en ellos.

La ciencia al modo industrial

El impacto de los nuevos medios en la inteligencia científica de la especie está siendo y será revolucionario, tanto o más que lo fue con la imprenta. En el caldo de cultivo tecnodigital donde ha surgido la sociedad de la información industrializada, de manera natural e inexorable, la ciencia crece y se diversifica, pero masiva e industrialmente. La información cualificada y fiable se multiplica mediante una algoritmia del conocimiento que ahora se ejecuta sobre una mezcla variable de inteligencia humana y artificial.

Pero, ¿cómo es la ciencia al modo industrial? Echeverría (2003) habla de tecnociencia o revolución tecnocientífica, Ziman (2003) de ciencia post-académica. Con ayuda de ambos podemos presentar una caracterización de sus rasgos, que en efecto suponen una ruptura con la ciencia de la “academia”.

La ciencia es una fuerza productiva estratégica

La funcionalidad biológica de la ciencia para los homo sapiens deviene sistema económico, modo de producción, motor de desarrollo. El conocimiento avanzado y experto es un factor económico esencial, una fuente de riqueza y de poder, un activo crucial de los estados, las sociedades civiles y los emprendedores, que sirve para la supremacía empresarial, política y militar. Las fronteras entre ciencia y tecnología, universidad e industria, etc., se difuminan. La investigación se organiza y el conocimiento se gestiona como una cadena productiva orientada a la rentabilidad. Se implantan mecanismos de *taylorización* y *management* para sostener la investigación a gran escala, que consume muchos recursos. Aunque una pluralidad de valores, a veces en conflicto, inspiran la actividad científica, ésta adquiere primordialmente un carácter instrumental y dirigido: es un medio para la acción, para la realización de objetivos o intereses, más que una pura búsqueda de la verdad. Fines corporativos, creación de mercados, demandas sociales prácticas o políticas públicas (cambio climático, energía, epidemias, etc.) son con frecuencia la meta de la investigación orientada. Pero

incluso la ciencia básica se legitima como base para la germinación de creatividad y resultados a medio o largo plazo.

La ciencia es neurodigital, *ciborg*

La ciencia aparece como un complejo sistema sociotécnico (Cronin, 2005) integrado por individuos y máquinas. Más allá de la considerada tradicionalmente como *big science* (astronomía, física de partículas, biología molecular, etc.), la computación artificial se convierte en esencial para la producción de ciencia, mucho más que una mera herramienta. La algoritmia basada en la comunicación entre sistemas cognitivos naturales deja paso a enormes sistemas informáticos de cálculo y tratamiento de información en los que participan equipos de investigadores a veces de todo el mundo (e-ciencia, comunidades virtuales). Las tecnologías de computación distribuida *grid* y otras infraestructuras llevan a cabo la captación de observaciones mediante sensores remotos e instrumentos de medición, el registro de datos experimentales, su análisis numérico, la simulación y modelización matemáticas, las tareas avanzadas de visualización y representación y el archivo masivo de información numérica o textual elaborada, así como la minería y explotación secundarias de los datos archivados, en posteriores proyectos de investigación.

La ciencia es una empresa colectiva

A lo largo del siglo XX la ciencia ha sido cada vez más cooperativa, hecho constatable viendo la cantidad de artículos que se publican firmados por más de cien investigadores. La ciencia al modo industrial depende de amplios equipos multidisciplinares de científicos y de ordenadores. Además, se articula en complejas redes y proyectos de investigación en los que intervienen agentes muy diversos: ingenieros y técnicos, políticos, militares, analistas de mercados, empresarios, gestores, grupos de interés, etc. La investigación se basa en extensa colaboración, grandes proyectos, grandes iniciativas, verdaderas empresas. La aparición de tecnologías cooperativas, la Web 2.0, las redes sociales científicas y las ciberinfraestructuras, aunque generan resistencias y contradicciones en el sistema institucional académico preexistente, promueven también formas de colectivización que conducen más allá de la mera cooperación, hacia el *crowdsourcing*. La inteligencia científica también se colectiviza, lo cual no significa que todos los resultados se vuelvan públicos, pues en realidad una parte será conocimiento-propiedad, sujeta su divulgación a los intereses de explotación.

Naturalmente, junto a la ciencia al modo industrial convive una ciencia artesanal, o una ciencia más clásicamente académica, inspiradas en métodos y valores en parte diferentes (reglas clásicas “mertonianas”, independencia intelectual, libertad de cátedra, ética del hac-

“¿Cuál es el lugar de los profesionales de la información en la tecnociencia industrial, colectiva y digital?”

ker, individualismo, fines sociales a largo plazo, etc.), lo que conduce a inevitables tensiones. Pero incluso estas facetas de la ciencia contemporánea, aunque no se disuelvan en el complejo científico-industrial, se basan también en sistemas digitales que conducen a una algoritmia cognitiva colectiva, reticular, neurodigital, etc.

La información científica al modo industrial

Como en otras épocas, las técnicas de comunicación, en este caso las tecnologías digitales (TIC), tienen una influencia decisiva en las formas que adopta la comunicación científica y en el modo de practicarse la investigación, que abordamos en el apartado anterior. Por ello, se están produciendo alteraciones muy importantes, que continuarán en el futuro, en los sistemas de información científico-técnica, a pesar de que no pocas veces desafían las normas sociales y epistémicas de la ciencia académica (evaluación, arbitraje, reconocimiento, consenso, estabilidad de los registros, etc.). Estamos ante un panorama complejo y cambiante, en el que los contenedores o canales comunicativos se caracterizan por los siguientes rasgos generales:

Transitoriedad

Los medios de comunicación científica mutan y evolucionan con gran facilidad y rapidez, a causa del cambio constante en las TIC característico de nuestra sociedad infoindustrial. Algunas soluciones se consolidan y duran más tiempo y otras resultan evanescentes o no llegan a sus públicos objetivo. La experimentación e innovación crea a veces una sensación de tránsito hacia un paradigma estable, pero en realidad lo único permanente es el dinamismo, la sucesión de novedades.

Polimorfismo

La comunicación científica adopta formas muy diferentes. Los contenedores se diversifican según las tecnologías que les sirven de base, conviviendo también modelos económicos y regímenes de propiedad intelectual diferentes. La efervescencia de sistemas tecnológicos favorece que coexistan diferentes canales alternativos, incluso para los mismos propósitos, llevando a la redundancia. Se suaviza la diferencia entre comunicación formal e informal, y se acrecienta por otro lado el volumen y relevancia de la información no textual: numérica, gráfica o visual. No es de esperar una simplificación significativa en los formatos del

intercambio científico, sino adaptaciones a ambientes concretos.

Especiación

Con la proliferación general, las ramas del árbol divergen, tiene lugar la formación de diferentes especies de comunicación o canales científicos, adaptados cada cual a ambientes determinados, para propósitos peculiares. La segmentación por comunidades disciplinares, temáticas, es muy relevante, pero también se genera diversidad en relación con clientelas específicas, proyectos y macroproyectos de investigación, políticas públicas, entornos geográficos, estrategias cooperativas o distintas formas de acceso a los contenidos. En definitiva, soluciones variadas para practicar la ciencia.

Reticularidad

El esquema clásico de la comunicación científica era un circuito desde el autor al lector donde todos los tipos de agentes operaban en una posición estable, con unas funciones claras y unas relaciones exclusivas con los nodos inmediatos, como en un ecosistema típico. Ahora todos los agentes pueden hacer casi de todo y entrar en contacto con todos, formándose una red densa y potente, de “mundo cercano”, que cataliza, aunque puede inestabilizar, los flujos de información. En términos económicos se produce una liberalización, se crea un mercado global y abierto a la competencia y cooperación entre los agentes informativos y entre las informaciones, según su distinta relevancia.

“Empresarialidad”

Si se me permite la fea expresión, los contenedores de información científico-técnica asumen ahora un carácter empresarial. Responden a objetivos, propósitos o proyectos deliberados, con o sin fines lucrativos, pero desmarcándose del *status quo*, innovando. Son iniciativas singulares, muchas veces empeños cooperativos o internacionales, en ocasiones individuales, que se ofrecen a los agentes de la investigación asumiendo algunos riesgos. Tienen algún sentido de orientación a resultados, bien sea de tipo económico, bien de naturaleza intrínsecamente científica, o social (promover el libre acceso al saber, por ejemplo). Como la propia investigación, la comunicación científica se empresarializa.

Como consecuencia de todo esto, en la era de la información industrializada no parece que vaya a existir un modelo único y estable de comunicación científica; a la proliferación industrial de los contenidos corresponde la de los contenedores, como vemos en otros sectores de la creación y consumo cultural. Crece la cantidad y crece la diversidad.

Sin embargo, la algoritmia científica necesita solidez, algún sustrato en el que evolucionar; si no, colap-

saría. El propio carácter de fuerza productiva central que asume la ciencia requiere que la complejidad y volatilidad no traspasen el umbral de lo gestionable. Y, por supuesto, la implicación de sujetos humanos en un entorno más que líquido, gaseoso, sería impracticable. Por todo ello parece que lo definitorio de la época son lo que podríamos llamar, para resumir en un concepto, “plataformas informativas específicas”: múltiples, singulares, con distintos alcances, coberturas, métodos y técnicas, orientadas cada cual a un propósito (forma de publicación, comunidad temática, proyecto, red de científicos, línea de investigación, objetivo de difusión o acceso, etc.), y por tanto consistentes al menos consigo mismas y adaptadas a su nicho ecológico.

“A la proliferación industrial de los contenidos científicos corresponde la de los contenedores”

Qué intermediación, o qué hay de nosotros los bibliotecarios-documentalistas

Este caleidoscópico panorama es muy diferente del mundo de la ciencia impresa, donde el manejo referencial y físico de los contenedores materiales, incluso sin pericia en los contenidos, otorgaba una función evidente a los profesionales de la información en la algoritmia del conocimiento, en el circuito comunicativo. ¿Cuál es su lugar en la tecnociencia industrial, colectiva y digital, basada en el flujo de bits entre múltiples máquinas y cerebros mediante plataformas informativas específicas y heterogéneas?

La producción industrial de conocimiento se asienta en la preponderancia de la mediación computacional, como la parsimoniosa ciencia medieval en la minuciosidad del copista de manuscritos. Las redes virtuales, la inteligencia científica colectiva y reticular, parecen facilitar que los investigadores se basten como nodos, sin necesidad de nodos intermedios. Además, los flujos y procesos de información siguen, según una dinámica natural, las rutas más fáciles y con menor coste, usando el *middleware* más sencillo. ¿En qué medida son los profesionales de la información ese *middleware*, capaz de responsabilizarse en las ciberinfraestructuras tecnocientíficas de la capa de contenidos, la *content layer* de Borgman (2007)?

Sin embargo, si la ciencia es una empresa colectiva en la que intervienen equipos muy amplios y diversos grupos de profesionales, ¿es posible que los especialistas en documentación tengan un papel significativo, que puedan agregar valor en la organización del cono-

cimiento reduciendo costes y tiempos en la producción o el consumo? La necesaria multidisciplinariedad de las redes de investigación ¿no abriría oportunidades? Quizá se perciba como imprescindible una conservación y transmisión del conocimiento en el tiempo que exijan plataformas informativas específicas muy estables, donde sea necesario el trabajo formal y sistemático de bibliotecarios y documentalistas. O la propia complejidad del panorama de la comunicación científica puede hacer aconsejable contar con expertos en la materia, capaces de (típicamente, nuevamente) asesorar y orientar en la selva del conocimiento.

Algunos análisis sobre el particular son en el fondo pesimistas (Nicholas; Rowlands, 2008). Aunque no se trata de un pronóstico, por mi parte creo que la posición corporativa de los profesionales de la información dentro de la ciencia industrializada es más bien débil o incierta (quizá al contrario que en la educación). Quienes actúan en los equipos científicos sin ser decisores o gestores, o son expertos en disciplinas científicas relevantes al caso o en la infraestructura tecnológica. Los especialistas en documentación, como hemos repetido, siempre han manejado, sobre todo, los contenedores, pero ahora éstos son exclusivamente tecnodigitales. Abrir un nicho diferenciado y significativo (no una madriguera) entre los contenidos (las ciencias) y los contenedores (infotecnología) no es sencillo. Dicho de otra forma, ¿cuánto de sustancial, sistemático, exclusivo y no efímero puede aportar la profesión si se dejan fuera las TIC en sí mismas y la especialización en las diversas ramas de la ciencia, en un contexto como el que describimos en el anterior apartado?

La respuesta es triple, creo que tres son las direcciones en que hay perspectivas. La primera es en realidad la propia y plena reconversión de los profesionales en tecnólogos de la información, sólidamente competentes en los actuales (y futuros) contenedores de la ciencia, llevando a cabo una actividad que estaría mejor encuadrada entre las ingenierías e informáticas. Coincidiendo con Pérez-Agüera (2008) los profesionales de la información, si no son alguna clase de ingenieros de la información, ¿qué son, en la sociedad infoindustrial?

El segundo camino es más fiel a la tradición. Se trata de liderar o participar en plataformas informativas específicas donde el tratamiento de los contenidos sea factor relevante y la cualificación del documentalista represente un valor, como repositorios y colecciones digitales de textos u otra documentación. Ahora, con la proliferación de archivos y redes de datos primarios de la investigación, en la e-ciencia, la figura del *data curator* es otra oportunidad (Martínez-Uribe; Macdonald, 2008). Intervenir en proyectos de e-ciencia entraña, no obstante, tener buenos conocimientos en materia de gestión y experiencia temática para el entendimiento

con el investigador (**Garritano; Carlsson, 2009**). Y requiere una buena dosis de competencias tecnológicas, lo que nos remite al camino anterior.

En tercer lugar, sí puede haber nodos especiales de mediación en el seno de la tecnociencia reticular. Sólo que su lugar no está ya reservado de antemano, tienen que ganarse los vínculos, obtener su sitio en el mercado libre de la comunicación científica. **Freire (2008)** habla en un sentido más amplio, de comisarios digitales, que no son necesariamente personas. Pero individuos creativos, competentes, ambiciosos y adaptables, capaces de añadir valor personal a los equipos científicos, pueden ser aliados útiles como gestores de información en las actividades de investigación, quizá al precio de desdibujar sus atribuciones tradicionales. Aunque el dilema colectivo gravita sobre los responsables de formar titulados, uno a uno los profesionales de la información pueden hacer aportaciones a la algoritmia colectiva de la ciencia industrializada. Aportaciones más amenas, esperemos, que apretar las tuercas al estilo de *Tiempos modernos*.

Bibliografía

- Borgman, Christine L.** *Scholarship in the digital age: information, infrastructure and the internet*. Cambridge, Massachusetts: MIT Press, 2007.
- Cronin, Blaise.** *The hand of science: academic writing and its rewards*. Lanham, Md: Scarecrow Press, 2005.
- Echeverría, Javier.** *Los señores del aire: Telépolis y el tercer entorno*. Barcelona: Destino, 1999.
- Echeverría, Javier.** *La revolución tecnocientífica*. Madrid: Fondo de Cultura Económica, 2003.
- Eisenstein, Elizabeth L.** *The printing press as an agent of change: communications and cultural transformations in early modern Europe*. Cambridge: Cambridge University Press, 1979.
- Freire, Juan.** "Conocimiento y usuarios en la cultura digital". *Nómada: reflexiones personales e información sobre la sociedad y el conocimiento abiertos* [blog], 11 de noviembre de 2008. <http://nomada.blogs.com/jfreire/2008/11/conocimiento-y-usuarios-en-la-cultura-digital-revista-frc.html>
- Garritano, Jeremy R.; Carlsson, Jake R.** "A subject librarian's guide to collaborating on e-science projects". *Issues in science and technology librarianship*, 2009, Spring. <http://www.isrl.org/09-spring/refereed2.html>
- Hull, David L.** *Science as a process: an evolutionary account of the social and conceptual development of science*. Chicago: University of Chicago Press, 1988.
- Maltrás, Bruno.** "Generación y comunicación del conocimiento científico". En: Lancaster, Frederick W.; Pinto-Molina, María (coords.) *Procesamiento de la información científica*. Madrid: Arco/Libros, 2001, pp. 19-40.
- Martínez-Urbe, Luis; Macdonald, Stuart.** "Un nuevo cometido para los bibliotecarios académicos: data curation". *El profesional de la información*, 2008, mayo-junio, v. 17, n. 3, pp. 273-280.
- Nicholas, David; Rowlands, Ian.** (eds.) *Digital consumers: reshaping the information professions*. London: Facet Publishing, 2008.
- Pérez-Aguiera, José R.** "Ingeniería documental frente a artesanía documental. ¿Cuál es el modelo a seguir?". *El profesional de la información*, 2008, mayo-junio, v. 17, n. 3, pp. 257-260.
- Thagard, Paul.** *Computational philosophy of science*. Cambridge, Massachusetts: MIT Press, 1993.
- Solís, Carlos; Sellés, Manuel.** *Historia de la ciencia*. Madrid: Espasa, 2007.
- Ziman, John.** *¿Qué es la ciencia?* Madrid: Cambridge University Press, 2003.
- Luis-Javier Martínez.** *Universidad de Cantabria, Biblioteca.* javier.martinez@gestion.unican.es

Quieres recibir tu copia del Anuario ThinkEPI 2009 en casa*?

ANUARIO
Think
EPI

Haz tu pedido en <http://www.thinkepi.net/>

¡Sólo 41€!

*La tarifa individual particular sólo sufraga el coste marginal o incremental de la imprenta. Está pensada para los particulares que se costean el Anuario de su bolsillo, bien porque quieren leerlo en su casa o bien porque su institución no lo ha comprado.

**BOOK NOW FOR
EARLY BIRD, GROUP AND
ASSOCIATION DISCOUNTS**

online information 2009

1-3 December,
Olympia Conference Centre,
London, UK

ONLINE INFORMATION CONFERENCE 2009

INFORMATION + CONVERSATION =
COLLABORATION + INNOVATION

1-3 December,
Olympia Conference Centre, London, UK

Why you should attend:

- **Hear keynote presentations** from **Dame Wendy Hall** & **Professor Nigel Shadbolt**, University of Southampton, UK; **Blaise Cronin**, Editor-in-Chief, *Journal of the American Society for Information Science and Technology*, USA and **Charlene Li**, Co-Author 'Groundswell: Winning in a World Transformed by Social Technologies', USA
- **Hear from** more than 90 international expert speakers, including representatives from: **Allen & Overy**, **BBC**, **Forrester**, **IBM**, **KPMG**, **Microsoft**, **Outsell**, **Oracle**, **University of Oxford**, **World Wildlife Fund** and **Yahoo!** amongst others
- **Mix and match** sessions from three different topic tracks each day
- **Attend** pre-conference workshops on Monday 30 November covering **'Web Search'** and **'The SharePoint Academy'**
- **Automatic free entry** to the Online Information & IMS 2009 exhibitions which run alongside the conference
- **Learn about:**
 - Advances in the semantic web - what it means for the informational professional
 - Innovations in the social web
 - How information professionals can deliver and demonstrate value

View full programme information and book your place at
www.online-information.co.uk/conference

Twitter Hashtag: #ONLINE09

Official Show Publication:



Platinum Conference Sponsor:



Delegate Drinks Sponsor:



Supporting Associations:



Las universidades públicas españolas en Google Scholar: presencia y evolución de su publicación académica web

Por Enrique Orduña-Malea, Jorge Serrano-Cobos y Nuria Lloret-Romero

Resumen: Se estudia la validez de Google Scholar como servicio que refleja adecuadamente la producción científica de una universidad, comparando su cobertura de las universidades públicas españolas con la mostrada por Scopus. Se analiza la presencia y evolución de la documentación científica de las sedes web de estas universidades desde enero hasta julio (ambos meses incluidos) de 2009. Los resultados muestran que, pese a encontrar cierta interrelación en la productividad de las instituciones en Scopus y Scholar, entre ambos sistemas existen grandes diferencias en los resultados totales que invalidan a este último como un reflejo válido de la producción universitaria. Finalmente, la mayoría de las webs universitarias analizadas presentan un crecimiento positivo aunque moderado en Scholar, durante el período estudiado.

Palabras clave: Sedes web universitarias, Google Scholar, Producción científica, Universidades públicas españolas, Rankings de universidades.

Title: Spanish public universities in Google Scholar: presence, evolution and coverage of their scientific output

Abstract: The validity of Google Scholar as a service that reflects properly the scientific output of a university is analyzed, comparing its coverage of Spanish public universities with that of Scopus. The presence and evolution of scientific documentation on the websites of the Spanish public universities from January to July (both months inclusive) 2009 is also studied. The results show that, despite finding some interrelationship between Scholar and Scopus concerning the productivity of institutions, there are large differences in the total results that override the latter as a valid reflection of university productivity. Finally, Scholar shows positive, albeit modest, growth during the studied period for most university websites analyzed.

Keywords: Academic websites, Google Scholar, Scientific output, Spanish public universities, University rankings.

Orduña-Malea, Enrique; Serrano-Cobos, Jorge; Lloret-Romero, Nuria. "La universidades públicas españolas en Google Scholar: presencia y evolución de su publicación académica web". *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 493-500.

DOI: 10.3145/epi.2009.sep.02



Enrique Orduña-Malea es ingeniero técnico de telecomunicaciones, esp. sonido e imagen, y técnico especialista en imagen fílmica, licenciado en documentación y máster en contenidos y aspectos legales en la sociedad de la información (especialidad en contenidos multicanal), por la Universidad Politécnica de Valencia (UPV). Investigador predoctoral en la división de investigación en tecnología e información (Calsi) del Instituto de Diseño y Fabricación (IDF) de la UPV. Redactor jefe del Anuario ThinkEPI, webmaster y miembro del think-tank sobre estudios de estrategia y prospectiva de la información, y miembro del equipo de trabajo del gestor social de noticias DocuMenea.



Jorge Serrano-Cobos, licenciado en documentación por la Univ. de Granada, ha trabajado como arquitecto de información y consultor en diseño de sistemas de información e intranets y especialista en marketing online para compañías como Planeta DeAgostini, Serikat Consultoría Informática y Google Inc. Es director del Departamento de Contenidos de MASmedios.com. Sus intereses profesionales van desde la arquitectura de información, el diseño de interacción (incluyendo opacs, juegos multimedia, etc.) al estudio de information seeking behaviour, web mining, search analytics, e-commerce y e-marketing. Miembro del think-tank sobre estrategia y prospectiva de la información ThinkEPI.



Nuria Lloret-Romero es doctora por la UPV, master en desarrollo de sist. de información online, soportes magnéticos cd-rom (Masdoc) de la UB, y master en dirección financiera de empresas (Fund. Estudios Bursátiles de Valencia). Prof. titular y directora del Dpto. de Comunicación Audiovisual, Documentación e Historia del Arte (Dcadha) de la UPV. Subdirectora del Inst. de Investigación Univ. IDF y directora del grupo de investigación Calsi, de dicho Instituto. Evaluadora de la UE (programa IST). Socia fundadora de MASmedios, vicepres. de la Fed. Valenciana de empresas de Servicios Avanzados y asesora del Comité de Tecnologías de la Información de la Cámara de Comercio de Valencia.

1. Introducción

La cibermetría como disciplina da un salto cuantitativo en 2004 con la puesta en marcha del *Ranking web of world universities*¹, primer ranking de universidades de cobertura mundial basado en criterios meramente cibernéticos (Aguillo; Granadino; Ortega; Prieto, 2006; Aguillo; Ortega; Fernández, 2008).

Pese a la existencia de otros rankings^{2,3} y estudios acerca de la presencia web de las universidades mediante la información proporcionada por los buscadores (Bar-Ilan, 2004; Payne; Thelwall, 2004), esta iniciativa marca un hito debido a sus objetivos, cobertura, impacto mediático y orientación documental.

Tal como los responsables del proyecto expresan¹, “la posición en el ranking resume el rendimiento global de la universidad, aporta información para estudiantes o profesores y refleja el compromiso académico con la difusión del conocimiento científico”.

En este contexto, el concepto de “sede web académica” debe entenderse como un conjunto de webs ligadas jerárquicamente a una página principal, identificable por una url, y que forma una “unidad documental” reconocible e independiente de otras por su representatividad institucional (Aguillo, 1998).

Por tanto, asumiendo la web como un canal de comunicación entre la universidad y sus usuarios, la calidad y cantidad de documentación accesible desde su dominio web (que podrá alojar a una o más sedes web) serán indicadores de su grado de comunicación; siendo éste, a su vez, indicador parcial de su calidad como institución.

No obstante, la documentación publicada dentro de un dominio web académico es heterogénea, pudiendo ser clasificada fundamentalmente en administrativa, bibliográfica, científica, docente e informativa.

“Google Scholar destaca por la inclusión de citas y por la mayor estabilidad de los resultados frente a los motores de búsqueda generalistas”

La cantidad y calidad de cada uno de estos tipos de documentación sirven potencialmente para valorar, entre otros criterios, las diferentes misiones de la universidad (docencia, investigación y transferencia), o al menos la sub-misión de su puesta online.

Sin embargo, los motores de búsqueda no discriminan apropiadamente estas clases de documentación automáticamente ni proporcionan comandos adecua-

dos para filtrarla de forma conveniente a nuestros propósitos (es decir, los ítems se encuentran etiquetados, pero no es posible filtrar por ellos), siendo este hecho delicado desde el punto de vista cuantitativo.

En 2006 se lanza *Google Scholar* con la intención de resolver algunos de estos problemas, así como de indizar la llamada internet invisible. Sus resultados pasan supuestamente un filtro por el que el sistema considera que son científicos, o se tienen al menos ciertas garantías. Esto se logra mediante acuerdos comerciales con proveedores, lo que explica en parte su cobertura.

De ser cierto significa que *Scholar*, con los comandos de búsqueda apropiados, podría ser usado como instrumento para medir la cantidad de documentación científica que las universidades publican en la Web y/o dentro su dominio web.

Esta última posibilidad permitiría medir parte del compromiso de comunicación entre la universidad y sus usuarios comentado anteriormente, dejando la documentación docente y de transferencia aparte. Este aspecto llevó a los responsables del ranking web a adoptar *Google Scholar* como uno de sus indicadores, a partir de su edición de 2006.

Google Scholar destaca sobre todo por la inclusión de citas (es la única base de datos multidisciplinar que las suministra, junto a *Scopus* y *WoS*) y por la mayor estabilidad de los resultados frente a los motores de búsqueda generalistas.

Aun así, se detecta una serie de inconvenientes en el uso de *Scholar* como herramienta bibliométrica (Torres-Salinas; Ruiz-Pérez; Delgado-López-Cózar, 2010):

- No todos los documentos indizados se pueden considerar académicos.
- No proporciona ninguna información sobre las fuentes que cubre.
- Cobertura disciplinar desigual en ciencias sociales y humanidades.
- Carencia de comandos básicos de consulta para realizar estudios bibliométricos.
- Ausencia absoluta de normalización en los datos (autores y fuentes).

A esto se unen otros problemas en la realización de análisis de citas (Jacsó, 2008a; Harzing; Van der Wal, 2008).

Este trabajo, por su parte, plantea otra serie de cuestiones relacionadas con la cobertura, visibilidad y evolución de los contenidos alojados en los dominios web de las universidades españolas:

Universidades	Código	url	Universidades	Código	url
Universitat Autònoma de Barcelona	UAB	http://www.uab.es/	Universidad de Oviedo	UO	http://www.uniovi.es/
Universidad Autónoma de Madrid	UAM	http://www.uam.es/	Universidad de Salamanca	USAL	http://www.usal.es
Universidad Carlos III de Madrid	UC3M	http://www.uc3m.es	Universidad de Santiago de Compostela	USC	http://www.usc.es/
Universidad Complutense de Madrid	UCM	http://www.ucm.es/	Universidad de Sevilla	US	http://www.us.es
Universidad de Alcalá	UAH	http://www.uah.es/	Universitat de València	UV	http://www.uv.es
Universitat d'Alacant	UA	http://www.ua.es/	Universidad de Valladolid	UVA	http://www.uva.es/
Universidad de Almería	UAL	http://www.ual.es/	Universidad de Vigo	UVI	http://www.uvigo.es/
Universitat de Barcelona	UB	http://www.ub.edu	Universidad de Zaragoza	UZ	http://www.unizar.es/
Universidad de Burgos	UBU	http://www.ubu.es	Universidad del País Vasco	EHU	http://www.ehu.es
Universidad de Cádiz	UCA	http://www.uca.es/	Universidad Internacional de Andalucía	UIA	http://www.unia.es/
Universidad de Cantabria	UC	http://www.uncan.es	Universitat Jaume I	UJI	http://www.uji.es/
Universidad de Castilla-La Mancha	UCLM	http://www.uclm.es/	Universidad Miguel Hernández de Elche	UMH	http://www.umh.es/
Universidad de Córdoba	UCO	http://www.uco.es/	Univ. Nacional de Educación a Distancia	UNED	http://www.uned.es
Universidad de Extremadura	UNEX	http://www.unex.es/	Universidad Pablo de Olavide	UPO	http://www.upo.es/
Universitat de Girona	UDG	http://www.udg.edu/	Universidad Politécnica de Cartagena	UPCT	http://www.upct.es/
Universidad de Granada	UGR	http://www.ugr.es	Universitat Politècnica de Catalunya	UPC	http://www.upc.es/
Universidad de Huelva	UHU	http://www.uhu.es/	Universidad Politécnica de Madrid	UPM	http://www2.upm.es
Universidad de Jaén	UJA	http://www.ujaen.es/	Universitat Politècnica de València	UPV	http://www.upv.es
Universidad de La Coruña	UDC	http://www.udc.es	Universitat Pompeu Fabra	UPF	http://www.upf.edu/
Universidad de La Laguna	ULL	http://www.ull.es	Universidad Pública de Navarra	UPNA	http://www.unavarra.es/
Universidad de La Rioja	UR	http://www.unirioja.es/	Universidad Rey Juan Carlos	URJC	http://www.urjc.es/
Universitat de les Illes Balears	UIB	http://www.uib.es	Universitat Rovira i Virgili	URV	http://www.urv.es/
Universidad de Las Palmas de Gran Canaria	ULPGC	http://www.ulpgc.es/			
Universidad de León	UL	http://www.unileon.es/			
Universitat de Lleida	UDL	http://www.udl.es/			
Universidad de Málaga	UMA	http://www.uma.es/			
Universidad de Murcia	UM	http://www.um.es/			

Tabla 1. Universidades públicas españolas (fuente: Crue)

– ¿Hay alguna correlación entre la presencia web de las universidades españolas en *Google Scholar* y su producción académica?

– ¿Qué volumen de documentación presentan las universidades españolas a partir de sus dominios web en *Google Scholar*?

De esta forma, los objetivos principales son los siguientes:

– comparar la cobertura que proporciona *Google Scholar* de las universidades españolas con la mostrada por *Scopus*, y

– analizar la presencia y evolución de la documentación alojada en los dominios web de las universidades españolas.

Este trabajo se centra únicamente en las universidades públicas españolas, excluyendo a las privadas

con el propósito de tratarlas de forma exclusiva en el futuro.

2. Metodología

2.1. Correlación entre *Scholar* y *Scopus*

La productividad académica de las universidades se extrae directamente de *Scopus* mediante los datos del ranking *Scimago*⁴, relativos a la producción total de las universidades públicas españolas (tabla 1), desde 2003 hasta 2007.

Para la publicación web, se consultan todos los dominios web de las universidades públicas españolas⁵ desde *Scholar* para el mismo período de tiempo. Este proceso se lleva a cabo en junio de 2008.

La consulta se realiza mediante el comando “site” al dominio principal de cada universidad. En el caso

de la *UB* y la *UDG*, se combinan los resultados de los dominios .es y .edu.

Finalmente, los datos de ambas series (*Scopus* y *Scholar*) se exportan a una hoja de cálculo, donde se obtiene la correlación entre ambas.

Se debe indicar que no se recuperan los artículos publicados por autores de una determinada universidad recogidos en *Scholar*, sino los artículos publicados en la parte pública del dominio web.

Esto significa que no se evalúa la producción científica de la universidad, sino más bien su compromiso de facilitar documentación científica a su comunidad mediante la Web y en concreto, de *Scholar*.

2.2. Presencia y evolución

El marco de trabajo se sitúa en el análisis de los dominios web de las universidades públicas españolas, desde enero hasta julio (ambos meses incluidos) de 2009.

Se recogen datos de todas las universidades públicas españolas, a excepción de la *UPF*, *URV* y *UIA*, de las que no se dispone de información debido a problemas técnicos⁶, por lo que no se consideran.

Los datos se obtienen de nuevo mediante la consulta directa a *Scholar*, pero sin ninguna restricción temporal. Posteriormente se exportan, durante la primera semana de julio, a una hoja de cálculo para ser tratados.

La toma de datos es mensual; se realizan varias catas por mes y se obtiene la media de dichas muestras. Esta operación se repite para cada dominio web de cada universidad, obteniendo un valor en bruto por mes. De éstos se obtiene la media mensual, denominada *Scholar(m)*, el recorrido y un valor llamado de representatividad relativa en *Scholar* (*RScholar*).

El recorrido se obtiene como la resta del valor último (junio) menos el primero (enero) e indica la ganancia o pérdida en los meses analizados.

RScholar se calcula normalizando de 0 a 100 los valores en bruto de cada dominio web para cada mes. Se otorga el valor 100 al dominio web de la universidad con valor máximo, y el resto se normalizan de forma proporcional. Finalmente, la media de los 6 valores normalizados (uno por mes) de cada dominio web proporciona el valor de *RScholar*.

Se recogen asimismo los datos relativos a la resta entre los valores totales en *Google Scholar* y los valores delimitados temporalmente a partir del año 2000 (en adelante *Scholar recent*, *Gr*), para cada mes y dominio web.

De esta forma se premia la publicación web reciente de las universidades y no el simple volcado retrospectivo

“Pese a que las posiciones no coinciden, 7 de las 10 primeras universidades en *Scopus* se encuentran igualmente entre las 10 primeras en *Scholar*”

vo de material académico, que favorece a las universidades más antiguas al disponer de más documentación acumulada.

A partir de *Google recent* se obtiene la media de los 6 valores mensuales obtenidos, *Gr(m)*, así como los correspondientes a su representatividad relativa (*RGr*).

Finalmente, se indica que los resultados obtenidos no deben ser entendidos como aproximaciones a la calidad de las universidades o clasificaciones ordenadas donde un primer puesto signifique ser “mejor”, sino simplemente como reflejo de lo que *Google Scholar* está mostrando de los dominios web académicos, síntoma de la buena o mala comunicación vía web que las universidades públicas españolas tienen con sus usuarios.

3. Resultados

3.1. *Scopus* vs *Scholar*

La tabla 2 muestra las diez primeras universidades públicas españolas en productividad en *Scopus* y *Scholar* para 2003-2007.

Pese a que las posiciones no coinciden, 7 de las 10 primeras universidades en *Scopus* se encuentran igualmente entre las 10 primeras en *Scholar*. Aun así, se observan ciertos comportamientos divergentes: mientras la *UAM*, *USC* y *US* pierden representatividad en *Scholar*, la *UDG* y *UM* la mejoran. La *UPCT*, *UHU* y *UCO* mejoran igualmente, aunque quedan fuera de las 10 primeras.

Los casos de la *UR* y *UCM* son especialmente reseñables pues la cantidad de publicación web recuperada en sus dominios web es mayor que la producción académica recogida por *Scopus* en el mismo período. Es más, la *UR* aparece en el último puesto en productividad científica en *Scopus* mientras que se encuentra en primera posición en *Scholar*. Estos resultados se atribuyen, entre otros factores, a la indización de la base de datos *Dialnet* de la *UR* por parte de *Scholar*.

El caso de la *UCM* es diferente pues aun siendo la tercera universidad en producción científica, sus resultados en *Scholar* son desproporcionados debido a la indización de su servicio de sumarios *Compludoc*.

Por su parte, la figura 1 muestra la correlación entre los datos de *Scholar* y *Scopus* (dejando fuera de los cálculos a la *UR* y *UCM* por sus elevados valores en *Scholar*).

Universidades	Scopus	Universidades	Scholar
UB	11.243	UR	30.700
UAB	9.364	UCM	16.300
UCM	8.911	UM	1.550
UAM	7.851	EHU	1.490
UV	7.621	UB	1.445
UPC	6.884	UGR	1.280
UGR	5.774	UAB	1.070
USC	5.097	UPC	778
US	4.927	UDG	671
EHU	4.675	UV	661

Tabla 2. Productividad (Scopus) y publicación web (Scholar) de las 10 primeras universidades públicas españolas (2003-2007)

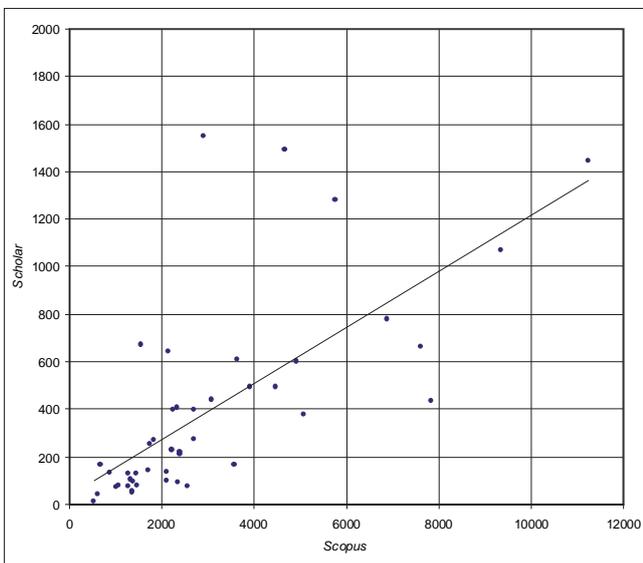


Figura 1. Correlación entre Scholar y Scopus (2003-2007)

Se obtiene una curva de regresión positiva, es decir, la mayoría de las universidades más productivas en Scopus lo son igualmente en Scholar, con un coeficiente de correlación de Pearson de 0,70 (si se tienen en cuenta UR y UCM resulta un coeficiente de sólo 0,08).

3.2. Producción y publicación web

La tabla 3 muestra los 46 dominios web estudiados, ordenados de forma descendente según su valor de RScholar. Así mismo se muestran los valores de representatividad extrayendo del cálculo a la UR y UCM (RScholar*) y la media de los valores brutos mensuales. Completan la tabla los valores de recorrido, RGr y Gr(m).

La UR y UCM aparecen las primeras en el ranking de RScholar, a una distancia considerable del resto. El dominio de la primera, de hecho, es abrumador incluso respecto a la UCM. Tras estas dos universidades, el resto aparece con valores más similares.

Respecto a RGr, los resultados muestran que 31 universidades (de 46) mejoran su representatividad al tener en cuenta únicamente la publicación web reciente, entre ellas las 17 primeras del ranking RScholar. Destaca la UCM (que incrementa cerca del 35% de representatividad total), así como la UM, EHU y UDG.

Si observamos el recorrido, vemos caídas de la UB (5.540 documentos menos que en enero), UM y UB. En cuanto a dominios web de universidades con recorrido positivo, destacan UGR y UJI. Así mismo, merecen atención los comportamientos de la UVA y UL (sin apenas cambios), y de la UDC (crecimiento 0).

Es igualmente de interés la relativamente buena posición de las universidades politécnicas, con dos entre las 10 primeras (UPC y UPM), y la UPV cerca, en la decimosegunda posición.

En la figura 2 se muestra la distribución –exponencial, o de larga cola-, de la media de registros por universidad.

3.3. Evolución

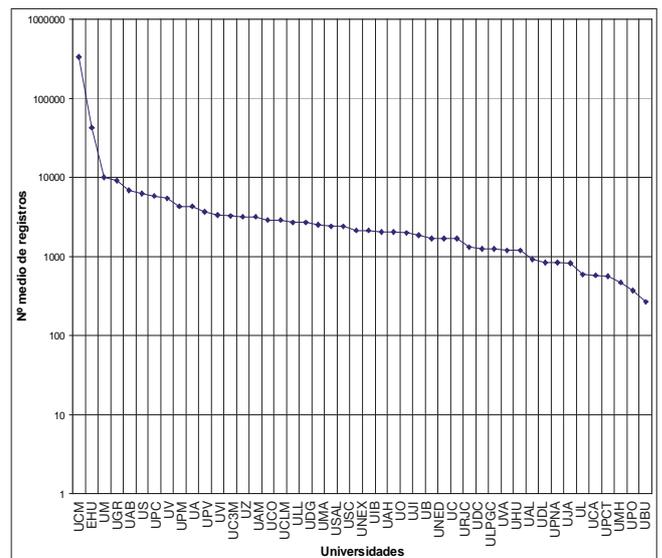


Figura 2. Distribución del número medio de registros por universidad (enero-junio 2009)

En la figura 3 se detalla la evolución del tamaño de las 10 primeras universidades, a excepción de la UR y UCM.

En términos generales, UM, EHU y UGR sobresa- len respecto al resto. Todos los dominios web, no obstante, presentan unas tendencias estables y ligeramente positivas durante todos los meses estudiados a excepción de marzo, donde se producen cambios importantes en prácticamente todos los dominios web, que determinan en buena medida los valores de recorrido obtenidos. Destaca el crecimiento de la UGR, US, UV y UA, y las caídas en la UM, UAB y UPC.

Universidades	RScholar	RScholar*	Scholar(m)	Recorrido	RGr	Gr(m)
UR	100,00	–	334.000	55.000	100,00	48.600,00
UCM	12,82	–	42.450	-2.300	47,78	19.450,00
EHU	3,03	98,58	10.087	380	6,00	2.561,70
UM	2,74	88,35	9.048	-1.900	9,37	4.000,00
UGR	2,07	67,68	6.922	1.140	4,02	1.795,00
UAB	1,88	60,75	6.218	-730	3,31	1.426,70
US	1,75	56,97	5.828	600	1,87	828,50
UPC	1,65	53,59	5.485	-370	2,47	1.050,00
UV	1,31	42,63	4.360	740	1,99	893,50
UPM	1,29	42,14	4.312	430	1,57	764,50
UA	1,11	36,17	3.698	830	1,15	528,83
UPV	0,99	32,26	3.300	830	1,64	720,83
UVI	0,97	31,77	3.250	410	1,27	536,00
UC3M	0,95	31,04	3.175	50	1,44	641,33
UZ	0,95	30,86	3.157	390	1,40	623,17
UAM	0,88	28,57	2.922	450	1,26	573,17
UCO	0,87	28,48	2.913	480	2,22	947,50
UCLM	0,81	26,44	2.705	210	0,66	299,33
ULL	0,80	26,22	2.682	480	0,78	336,83
UDG	0,75	24,51	2.505	480	2,66	1.208,20
UMA	0,73	23,60	2.415	-80	1,25	550,83
USAL	0,71	23,30	2.383	280	0,87	399,00
USC	0,64	21,00	2.148	250	1,32	580,00
UNEX	0,64	20,94	2.142	410	0,46	197,67
UIB	0,61	19,86	2.032	250	0,85	382,00
UAH	0,61	19,74	2.020	130	0,61	279,17
UO	0,60	19,66	2.012	200	0,60	267,50
UJI	0,55	18,22	1.863	1.200	0,98	519,17
UB	0,53	16,34	1.678	-5.540	0,98	316,50
UNED	0,51	16,69	1.707	180	1,01	467,67
UC	0,51	16,62	1.700	100	0,27	122,33
URJC	0,39	12,68	1.297	320	0,34	154,50
UDC	0,37	12,07	1.235	0	0,40	174,50
ULPGC	0,37	12,16	1.243	361	0,24	106,50
UVA	0,36	11,64	1.192	10	0,22	99,83
UHU	0,36	11,67	1.193	270	0,31	150,67
UAL	0,28	8,97	919	-190	0,23	111,00
UDL	0,26	8,25	845	-132	0,30	121,00
UPNA	0,25	8,17	836	65	0,16	67,50
UJA	0,25	7,99	817	42	0,33	146,67
UL	0,18	5,83	597	-13	0,21	95,33
UCA	0,17	5,64	577	98	0,10	50,00
UPCT	0,17	5,55	567	479	0,32	170,33
UMH	0,14	4,53	463	19	0,34	138,83
UPO	0,11	3,60	368	88	0,09	41,33
UBU	0,08	2,61	268	27	0,05	20,83

Tabla 3. Valores medios brutos y de representatividad de la producción (Scopus) y publicación web (Scholar) por universidad (enero-junio 2009). Algunas variables se han redondeado para mayor claridad.

Según la última edición del ranking web (enero-2008), la UM y EHU son las universidades con mayor tamaño en Scholar (sin contar a la UR y UCM), logrando las posiciones 71 y 94 a escala mundial respectivamente. Estos datos concuerdan con los resultados obtenidos, teniendo en cuenta la caída de la UM en marzo.

La figura 4 muestra por su parte la evolución de las 10 primeras universidades (excluyendo a la UR y UCM) en Scholar recent, destacando de nuevo el mes

de marzo como momento de cambios, con crecimientos muy notables para la UM, UGR, UDG y UPM.

De nuevo UM (en primera posición), EHU y UGR son las universidades con más registros. Aun así, sorprende la aparición en los primeros puestos de la UDG y UCO así como la pérdida de puestos de la UA y UVI, que abandonan el top ten en publicación web, al que llega la UPV gracias a su importante incremento en mayo.

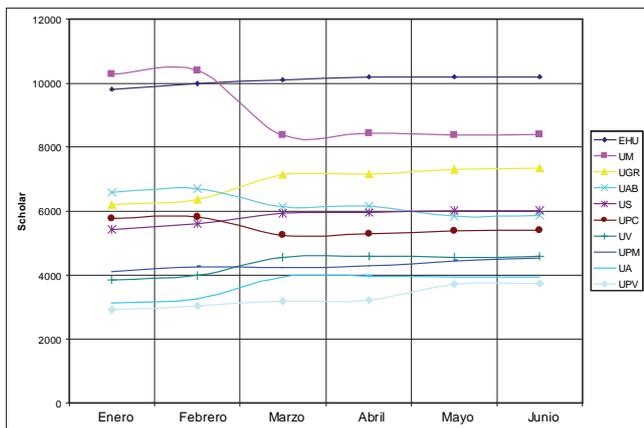


Figura 3. Evolución de los dominios web académicos españoles en Scholar (enero-junio 2009)

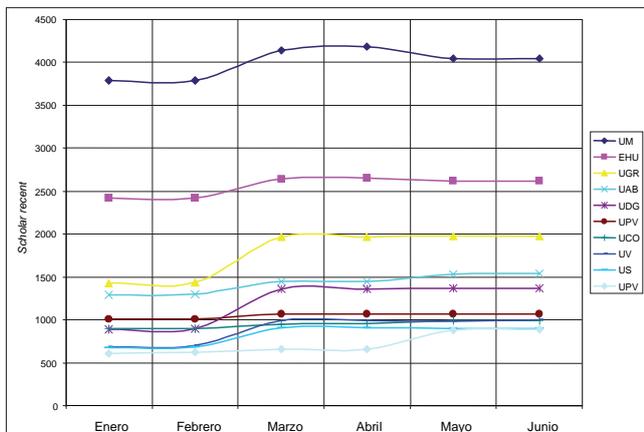


Figura 4. Evolución de los dominios web académicos españoles en Scholar recent (enero-junio 2009)

4. Discusión

Pese a obtener una correlación positiva (*coeficiente de Pearson* de 0,70) entre la productividad en *Scopus* y la publicación web en *Scholar* de las distintas instituciones (exceptuando a la *UR* y *UCM*), se asume la necesidad de una corrección futura por tamaño de universidad, aunque este parámetro no puede ser calculado de forma precisa pues no es proporcionado por todas las universidades, resultando además un término difuso y sin definición clara.

Existen grandes diferencias en los resultados brutos totales, que se asocian a:

- La recuperación tanto de artículos académicos no producidos por la universidad, como documentos no académicos, resúmenes, fichas catalográficas, etc. Muestra de ello es la indización de *Dialnet* o *Compludoc*.

- Diferentes políticas en cada dominio web de universidad en relación con el *Open Access*, que provoca que ciertas universidades queden mal representadas en *Scholar* aun disponiendo de suficiente material académico producido y publicado, como por ejemplo la *USC*⁷,

que pese a ser de las universidades con más producción, todavía no dispone de un repositorio institucional.

“La UR y UCM aparecen las primeras en el ranking de *RScholar*, a una distancia considerable del resto”

- Problemas de *Scholar* en la indización de algunos repositorios institucionales, donde se recopilan metadatos de los proveedores de servicios y no de los repositorios directamente, restando por tanto visibilidad a los dominios web de universidades.

Este es el caso –entre otros– de la *UAM*, que si bien dispone de un repositorio institucional⁸, su contenido no se encuentra disponible dentro del dominio web y, por tanto, no es recogido por la metodología seguida en este trabajo, lo que explica las diferencias entre producción científica y publicación web obtenidas para esta institución.

No es el propósito de este trabajo identificar todas las universidades que se encuentran en esta situación, aunque se asume que la posición de éstas mejoraría si solucionaran este problema, pues se considera una mala práctica⁹ la disposición de documentación académica fuera del dominio web institucional.

5. Conclusiones

El trabajo muestra una comparación entre la publicación web, recuperada por *Scholar* mediante los documentos alojados dentro de los dominios web de cada universidad pública española, y su productividad académica medida desde *Scopus*.

Los resultados demuestran objetivamente que, sin tener en cuenta los dominios web de universidades con grandes servicios de referencia (*Dialnet* y *Compludoc*), la producción académica en *Scopus* y la publicación web en *Scholar* están correlacionadas, aunque son conjuntos documentales de diferente naturaleza.

Pese a la curva positiva de regresión obtenida, la posible interrelación entre productividad en *Scopus* y publicación web en *Scholar* precisa de un análisis más profundo en el futuro, debido a la dependencia de variables externas no controlables como el filtrado de documentos y los métodos de indización de los repositorios institucionales (por parte de *Scholar*) o el tamaño y las políticas de acceso abierto (por parte de las universidades).

Pese a todo, las ventajas de *Scholar* como herramienta de búsqueda de información y como escaparate de atracción de usuarios para las universidades son innegables.

Por ello, tal vez las acciones emprendidas por la UR y UCM disfracen la verdadera actividad en publicación web de dichas universidades en *Scholar* (en todo caso, son carencias en el filtrado de esta herramienta), pero generan una presencia y visibilidad internacional que no logra ninguna otra universidad española.

Por otra parte, sólo 9 de los 46 dominios web universitarios analizados presentan en junio valores más bajos que en enero (recorrido negativo), lo que muestra un crecimiento generalizado, aunque moderado.

Los recorridos negativos (siendo destacable el caso de la UB) pueden deberse a la eliminación de cierta documentación temporal o cambios en las políticas web (como mayor uso de la intranet o de diferentes dominios web, consideradas ambas malas prácticas⁹).

Respecto a *Scholar recent*, los primeros puestos del ranking nacional no presentan muchas variaciones respecto a *Scholar*, aunque sí se observan diferencias en los valores de representatividad relativa. UCM es la que registra un mayor crecimiento y, en menor medida, EHU, UM, UGR, UDG y UCO.

En cuanto a trabajos futuros de investigación, se estima conveniente la comparación de los resultados obtenidos con los correspondientes a las universidades españolas privadas así como con algunas extranjeras, con el fin de contextualizarlos de forma más precisa.

Asimismo, también resultaría de interés el futuro análisis cuantitativo de los diferentes tipos documentales cubiertos en *Scholar* por dominio web de universidad.

6. Agradecimientos

Los autores agradecen enormemente el excelente trabajo realizado por los revisores, que con sus consideraciones y sugerencias han ayudado a mejorar notablemente el texto.

7. Notas

1. *Ranking web de universidades del mundo*
<http://www.webometrics.info/>
2. *Web popularity ranking*
<http://www.4icu.org/top200/>
3. *Ranking universitario de transparencia web*
<http://www.universidad.edu.co>
4. *Ranking de universidades españolas por calidad investigadora*, publicado por *Scimago* a partir de los datos extraídos de Scopus.
<http://www.magisnet.com/pdf/RankingEspanyolUniversidades.pdf>
5. *Conferencia de Rectores de las Universidades Españolas (CRUE)*
<http://www.crue.org/>
6. No se dispone de los datos relativos a enero debido a errores en las *queries* de consulta y en ningún caso hay problemas en dichos dominios para la extracción de información.
7. <http://www.accesoabierto.net/repositorios>
8. <http://digitool-uam.greendata.es/R>
9. http://www.webometrics.info/best_practices.html

8. Referencias

- Aguillo, Isidro F.** "Hacia un concepto documental de sede web". *El profesional de la información*, 1998, v. 7, n. 1-2, pp. 45-46.
- Aguillo, Isidro F.; Ortega, José-Luis; Fernández, Mario.** "Webometric ranking of world universities: introduction, methodology and future developments". *Higher education in Europe*, 2008, v. 33, n. 2/3, pp. 234-244.
- Aguillo, Isidro F.; Granadino, Begoña; Ortega, José-Luis; Prieto, José-Antonio.** "Scientific research activity and communication measured with cybermetric indicators". *Journal of the American Society for Information Science and Technology*, 2006, v. 57, n. 10, pp. 1296-1302.
- Bar-Ilan, Judit.** "A microscopic link analysis of academic institutions within a country – the case of Israel". *Scientometrics*, 2004, v. 59, n. 3, pp. 391-403.
- Harzing, Anne-Wil K.; Van der Wal, Ron.** "Google Scholar as a new source for citation analysis". *Ethics in science and environmental politics*, 2008, v. 8, n. 1, pp. 61-73.
- Jacsó, Péter.** "Google Scholar revisited". *Online information review*, 2008a, v. 32, n. 1, pp. 102-114.
<http://www.jacso.info/PDFs/jacso-GS-revisited-OIR-2008-32-1.pdf>
- Jacsó, Péter.** "The pros and cons of computing the h-index using Google Scholar". *Online information review*, 2008b, v. 32, n. 3, pp. 437-452.
<http://www.jacso.info/PDFs/jacso-pros-and-cons-of-computing-the-h-index.pdf>
- Kousha, Kayvan; Thelwall, Mike.** "Sources of Google Scholar citations outside the science citation index: a comparison between four science disciplines". *Scientometrics*, 2008, v. 74, n. 2, pp. 273-294.
- Mayr, Philipp; Walter, Anne-Kathrin.** "An exploratory study of Google Scholar". *Online information review*, 2007, v. 31, n. 6, pp. 814-830.
<http://www.ib.hu-berlin.de/~mayr/arbeiten/OIR31-6.pdf>
- Neuhaus, Chris; Neuhaus, Ellen; Asher, Alan; Wrede, Clint.** "The depth and breadth of Google Scholar: an empirical study". *Libraries and the academy*, 2006, v. 6, n. 2, pp. 127-141.
- Noruzi, Alireza.** "Google Scholar: the new generation of citation indexes". *Libri*, 2005, v. 55, n. 4, pp. 170-180.
<http://www.librijournal.org/pdf/2005-4pp170-180.pdf>
- Payne, Nigel; Thelwall, Mike.** "A statistical analysis of UK academic web links". *Cybermetrics*, 2004, v. 8, n. 1.
<http://www.cindoc.csic.es/cybermetrics/articles/v8i1p2.html>
- Smith, A.; Thelwall, Mike.** "Web impact factor for Australasian universities". *Scientometrics*, 2002, v. 54, n. 3, pp. 363-380.
- Torres-Salinas, Daniel; Ruiz-Pérez, Rafael; Delgado-López-Cózar, Emilio.** "Google Scholar: ¿una herramienta para la evaluación de la ciencia?". *Anuario ThinkEPI 2010* [en prensa].
<http://www.thinkepi.net/google-scholar-%C2%BFuna-herramienta-para-la-evaluacion-de-la-ciencia>

Enrique Orduña-Malea, Facultad de Informática, Universidad Politécnica de Valencia (UPV). Camino de Vera s/n, 46022 Valencia
enorma@fiv.upv.es

Jorge Serrano-Cobos, MASmedios, Álvaro de Bazán 10, 46010 Valencia
jorgemasmedios.com
jorgeserrano@gmail.com

Nuria Lloret-Romero, Universidad Politécnica de Valencia (UPV), Área de Biblioteca y Documentación Científica, Camino de Ver, s/n, 46022 Valencia
nlloret@upvnet.upv.es

Google Scholar como herramienta para la evaluación científica

Por Daniel Torres-Salinas, Rafael Ruiz-Pérez y Emilio Delgado-López-Cózar

Resumen: Google Scholar es un buscador especializado en recuperar documentos científicos y en identificar las citas que éstos han recibido, convirtiéndose de esta forma en un competidor de otros índices de citas. Diversos estudios han tratado de valorar su capacidad como herramienta bibliométrica. Debido a este interés se hace una introducción a su uso y a sus ventajas e inconvenientes frente a Web of Science y Scopus. Primero se analiza su modo de recopilar información y las propiedades de su interfaz. A continuación se describen los resultados a los que da lugar el buscador. En tercer lugar se analiza la cobertura de fuentes de información y los diferentes tipos documentales que recoge. Se expone cómo esta cobertura provoca un universo de citación diferente al de otros productos. Finalmente se especifican sus problemas de normalización y se expone una serie de precauciones a la hora de usarlo como herramienta de evaluación.



Daniel Torres-Salinas es doctor en documentación científica y trabaja como técnico de gestión de la investigación en la Universidad de Navarra, donde realiza auditorías sobre la calidad y el impacto de la investigación. Asimismo es miembro del grupo EC3 (Evaluación de la Ciencia y de la Comunicación Científica) de la Universidad de Granada donde participa en diferentes proyectos.



Rafael Ruiz-Pérez es profesor de catalogación en la Facultad de Comunicación y Documentación de la Universidad de Granada y miembro del grupo EC3. Sus líneas de investigación y publicación están centradas en la evaluación de revistas científicas y en la mejora de sus aspectos normativos. Es uno de los promotores de In-Recs: Índice de Impacto de las Revistas Españolas de Ciencias Sociales.



Emilio Delgado-López-Cózar es profesor de metodología de la investigación en la Facultad de Comunicación y Documentación de la Universidad de Granada y miembro del grupo EC3. Sus líneas de investigación se centran en la evaluación de revistas científicas y de la ciencia, el estudio de la investigación en ByD, y la evaluación del rendimiento científico. Es uno de los promotores del índice In-Recs.

Palabras clave: Google Scholar, Google Académico, Web of Science, Scopus, Bibliometría, Indicadores bibliométricos, Citas, Publicaciones científicas.

Title: Google Scholar as a tool for research assessment

Abstract: Google Scholar is a search engine that specializes in scientific information and in the identification of the citations that academic papers receive, making it a strong competitor for other citations indexes. For this reason, several studies have attempted to evaluate its capacity as a bibliometric tool. Due to this interest, we present an introduction to its use and the advantages and disadvantages versus Scopus and Web of Science. First, its way of collecting information and features of its interface are analyzed. The following section describes the results that Google Scholar generates. Thirdly, we analyze the coverage of information sources and the different document types to be found, showing how this coverage universe offers different citations versus other products. Finally, we specify the standardization problems of Google Scholar and offer a number of precautions that must be taken into account when using Google Scholar as an evaluation tool.

Keywords: Google Scholar, Google Académico, Web of Science, Scopus, Bibliometrics, Bibliometrics indicators, Citations, Scientific publications.

Torres-Salinas, Daniel; Ruiz-Pérez, Rafael; Delgado-López-Cózar, Emilio. "Google Scholar como herramienta para la evaluación científica". *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 501-510.

DOI: 10.3145/epi.2009.sep.03

1. Introducción

En 1998 internet asistió al nacimiento de uno de sus grandes hitos, el buscador *Google*, creado por **Sergei Brin** y **Larry Page**.

Desde entonces la historia del buscador es conocida ya que desbancó al resto de competidores convirtiéndose en el principal portal de acceso a la información y la verdadera puerta de entrada a internet. Pese a su carácter generalista, se ha convertido en una herramienta in-

Artículo recibido el 11-03-09

Aceptación definitiva: 24-08-09

sustituible en el campo académico ya que gran parte de la comunidad científica lo emplea casi de forma diaria y sistemática. Según **Friend** (2006), cerca del 72% de los profesores lo utiliza para la búsqueda de artículos, lo que evidencia su enorme penetración. *Google Inc.* consciente de su presencia en este sector de usuarios y del enorme volumen de negocio que supone la información científica, lanzó a mediados de noviembre de 2004 *Google Scholar* (en adelante *GS*) o *Google Académico*, con el fin de proporcionar acceso universal y gratuito a las publicaciones científicas.

Es un producto que, a diferencia de las bases de datos bibliográficas tradicionales, no vacía contenidos de revistas sino que rastrea sistemáticamente la Web siguiendo la misma filosofía que *Google* pero haciendo converger en una sola plataforma dos servicios. En primer lugar es un buscador de publicaciones científicas y, en segundo lugar, es un índice de citas que ayuda a conocer el impacto que las publicaciones tienen. Precisamente esta última propiedad es la que más interesa y la que lo convierte en una competencia directa de otros índices de citación como *Web of Science (WoS)*, de *Thomson Reuters*, o *Scopus*, de *Elsevier*. Por estas funciones *GS* se presenta a priori como una aplicación ideal para realizar al menos tres tareas:

- Buscar el texto completo de un trabajo.
- Buscar la producción bibliográfica de un autor, de una revista o sobre un tema.
- Buscar las citas que recibe un trabajo (libro, artículo de revista, tesis, informe...).

“GS es un buscador de obras científicas pero también es un índice de las citas que reciben, convirtiéndose en un competidor de WoS y Scopus”

En esta última función radica el enorme interés que tiene en general para los científicos que desean conocer la visibilidad de sus trabajos, y en particular para que evaluadores de la ciencia y bibliómetras puedan suplir las carencias de *WoS* y *Scopus*.

Otra de las particularidades fundamentales es su gratuidad, marcando una distancia enorme con el resto de proveedores, y más si tenemos en cuenta el elevado precio de las licencias de las bases de datos. Un ejemplo: la licencia nacional de *WoS* que proporciona la *Fundación Española para la Ciencia y la Tecnología (Fecyt)* para las universidades y organismos de investigación nacionales tuvo un coste para el trienio 2005-2008 de 25 millones de euros. En cierta medida *Google*,

mediante *GS*, está fomentando un acceso universal a la información científica y además está viendo favorecida esta política por el incremento de la presencia de publicaciones científicas en acceso abierto, lo que ha hecho que se haya convertido en el aliado perfecto del movimiento *Open Access*.

Ante este panorama *GS* empieza a emerger como una alternativa a las bases de datos que tradicionalmente se han empleado para los estudios cuantitativos de la ciencia. La comunidad bibliométrica le está prestando gran atención tratando de desvelar sus principales funciones. En la mayoría de los análisis realizados, bien a favor (**Harzing; Van-der-Wal**, 2008) o en contra (**Jacsó**, 2005a; 2008b), se intenta calibrar su idoneidad como herramienta de valoración de la actividad científica y concretamente su impacto.

Por ello presentamos una síntesis de las principales propiedades de *GS*, centrándonos en cómo usarlo, en fijar su cobertura real y en analizar las ventajas y limitaciones que posee para su uso bibliométrico.

2. Funcionamiento e interfaz de búsqueda

GS se basa como *Google* en un robot, *Googlebot*, que de forma sistemática rastrea los contenidos de la Web, en este caso la Web académica, recopilando la información colgada de distintos dominios institucionales pertenecientes a universidades, repositorios, páginas de revistas, bases de datos e incluso catálogos de bibliotecas.

Una vez identificadas las referencias o los documentos, éstos son indizados registrando su descripción bibliográfica e incluyendo además las citas bibliográficas cuando se ha localizado el texto completo.

Los formatos que indiza son los habituales en el campo académico como doc o ppt, pero destaca especialmente el pdf seguido del html, aunque también podemos encontrar documentos en postScript. Esta indización a texto completo puede no ser del todo cierta ya que en determinadas ocasiones *Google* solo indiza 101 KB de los sitios web y lo mismo ocurre, aunque con mayor tamaño, con los documentos en pdf; podemos encontrar documentos en pdf de cierto tamaño que no están indizados completamente (**Price**, 2004; **Jacsó**, 2005b). Si efectivamente la información y el contenido relevante, como pueden ser las citas bibliográficas, se sitúan tras el límite de indización, éstas se pierden y no pueden ser recuperadas.

La interfaz de *GS* está basada en la proverbial sencillez de *Google* de manera que no resulta compleja al usuario: la pantalla principal sólo presenta una caja de búsqueda donde podemos introducir los términos que deseamos. Como en *Google*, disponemos de una serie

de operadores que pueden ayudar a mejorar la pertinencia de la búsqueda: el operador “+” permite incluir palabras vacías, “OR” expandir las búsquedas, “filetype:” especificar el formato del documento, “-” eliminar una palabra, o el uso de comillas localizar una frase exacta. Con la opción de búsqueda avanzada, podemos realizar búsquedas por tres campos: autor, título de la revista y año de publicación.

“GS indiza diferentes fuentes de información y variados tipos documentales”

Junto a estos tres campos reseñados, aunque sólo disponible en la versión inglesa, GS presenta un filtro para limitar los resultados por 7 grandes áreas científicas. Podemos localizar otras prestaciones avanzadas en el menú *Preferencias de Google Académico* donde, por ejemplo, desde la opción “idioma de búsqueda” los textos pueden ser limitados a una lengua concreta. En líneas generales, la interfaz apunta al minimalismo y huye de la sofisticación a la que nos tienen acostumbrados otros productos; las opciones de búsqueda, a pesar de tratarse de información científica, son bastante limitadas sobre todo si las comparamos con otras bases de datos bibliográficas. Así, *WoS* cuenta con 12 campos de búsqueda diferentes (tema, título, autor, grupo, publicación, año, dirección, congreso, lengua y tipo documental, ID proyecto y entidad financiadora) y *Scopus* incluye hasta 17 campos diferentes (ISSN, DOI, primer autor, etc.). Dichas bases de datos tienen opciones de filtrado de documentos muy completas que incluyen año de publicación, tipos documentales o revistas pudiéndose además obtener informes bibliométricos de los resultados como ocurre en *WoS* con los *Citation Reports*.

Convendría subrayar como una de las limitaciones principales de la interfaz de búsqueda de GS la ausencia de una opción específica de búsqueda para la localización directa de las citas que ha recibido un trabajo o un autor en un modo similar al que, por ejemplo, encontramos en *WoS* con *Cited Reference Search*.

“Gracias a su exhaustivo rastreo de la literatura científica GS alumbra un corpus documental que de otro modo sería difícilmente recuperable”

3. Presentación e interpretación de los resultados de búsqueda

Una vez lanzada la búsqueda GS devuelve los resultados que considera más pertinentes pero hay que tener en cuenta que sólo podrán ser consultados los 1.000 primeros. El algoritmo que ordena estos resultados se sostiene sobre la misma filosofía que el conocido *PageRank*, basado en una premisa tomada del mundo académico, donde los trabajos más citados son también los más importantes, haciéndola extensible al mundo Web mediante los enlaces.

Sin embargo, se incluye una serie de modificaciones sobre *PageRank* para adaptarlo a propiedades y convenciones propias del mundo científico y académico. Así, a la hora de ordenar los resultados en GS pesan otros factores. Por ejemplo, se considera el número total de citas recibidas, la disponibilidad del texto completo, el autor y la publicación (Maryr; Walter, 2007). Una vez ordenados los resultados, se muestran en un modo similar al de *Google*, aunque tienen una lectura diferente. Resumiendo, podemos encontrar al menos tres tipos de resultados diferentes (tabla 1):

– Enlaces al trabajo a texto completo. En este tipo de resultado obtenemos un enlace directo a la publicación original a texto completo al pinchar sobre el título. Se identifica por una flecha verde y el formato del documento entre corchetes.

1. Resultado que nos dirige al documento original a texto completo
[PDF] ► E-estrategias en la introducción y uso de las TIC en la universidad JM Duart, F Lupiáñez - Revista de Universidad y Sociedad del conocimiento, 2005 - ddd.uab.cat ... 1 1 5 © Josep M. Duart y Francisco Lupiáñez, 2005 © FUOC, 2005 E-estrategias en la introducción y uso de las TIC en la universidad Josep M. Duart ... Citado por 27 - Artículos relacionados - Versión en HTML - Las 15 versiones
2. Resultado que nos devuelve una cita
[CITAS] Aprender sin distancias. JM Duart - Nueva Revista de Política, Cultura y Arte, 2000 Citado por 15 - Artículos relacionados - Las 2 versiones
3. Resultado que nos dirige a la fuente del documento
La motivación como interacción entre el hombre y el ordenador en los procesos de ... JM Duart - dialnet.unirioja.es ... Ayuda. La motivación como interacción entre el hombre y el ordenador en los procesos de formación no presencial. Autores: Josep ... Citado por 3 - Artículos relacionados - En caché

Tabla 1. Tipos de resultados que se pueden obtener de GS

1	2	3		
[PDF]	▶	redalyc.uaemex.mx		
JM Duart, A Sangra, M Josep, A Sangrà - Ciencia, Docencia y Tecnología, 2004 ... 2004 Nora Liliana Dari RESEÑA DE "APRENDER EN LA VIRTUALIDAD" DE JOSEP M. DUART Y ALBERT SANGRÀ Ciencia, Docencia y Tecnología, mayo, año/vol. ...				
Citado por 188	Artículos relacionados	Importar al EndNote	Buscar en Rebiun	Las 4 versiones
4		6	7	5

Figura 1. Elementos destacados de un resultado en GS

– Citas extraídas de documentos indizados. Los resultados vienen marcados con la etiqueta “[CITA]” y no presentan ningún tipo de enlace.

– Enlaces al documento en su fuente original. El resultado remite a alguna de las plataformas (repositorios y otras bases de datos) que GS rastrea. El acceso al documento depende de la plataforma.

“GS incluye indiscriminadamente todas las citas que es capaz de identificar, sin asegurar su calidad”

Todos los resultados, independientemente de su tipo, presentan una estructura similar (figura 1). En la zona superior encontramos una breve descripción bibliográfica del documento (título, autores, revista/fuente, año). Entre corchetes se indica ante qué documento nos encontramos, bien señalando el formato (pdf, html) o el tipo documental (libro, cita) (figura 1, n. 1). Se indica claramente si GS proporciona un acceso directo al documento con una flecha, situada en la zona derecha si el enlace conduce al texto original, o en la izquierda si redirige a otra fuente que proporciona el documento (figura 1, n. 2). También muestra cuál es el sitio web del que GS ha extraído la información (servidor, repositorio, etc.) (figura 1, n. 3); evidentemente esto no está disponible para el caso de “[CITA]”.

En la parte inferior proporciona una serie de enlaces. Destaca en primer lugar “citado por”, donde se muestra el listado de documentos recopilados por GS que citan el trabajo (figura 1, n. 4). Un segundo elemento interesante son las “versiones” (figura 1, n. 5) ya que agrupa bajo un mismo encabezamiento todas las versiones que ha localizado de un mismo trabajo, aunque no siempre realiza esta operación con precisión. Ejemplos de diferentes versiones de un mismo texto son los preprints, documentos de conferencias u otras adaptaciones, dándole a la versión del editor, si se indexa, el carácter de versión principal. La recopilación de las versiones facilita la agrupación de las citas dadas a un trabajo con independencia de su versión. Otras op-

ciones de interés son la capacidad de exportar el registro a un software de gestión bibliográfica (figura 1, n. 6) o la posibilidad, si tenemos configurada esta función, de localizar el documento en una biblioteca gracias a la tecnología *Link Resolver* (figura 1, n. 7) (Hartman; Mullen, 2008).

“GS parece indexar cualquier revista, independientemente de su calidad”

4. Cobertura documental de GS y su impacto sobre la citación

Una de las propiedades que convierten a GS en un producto único e interesante es su amplia cobertura, que se pone de manifiesto con la indización de diferentes y variados tipos documentales (libros, informes científico-técnicos, *working papers* –informes de trabajo-, comunicaciones y ponencias en congresos, seminarios y jornadas, tesis y tesinas, etc.). Por tanto, no se limita a los trabajos publicados en revistas científicas, como en la mayor parte de las bases de datos. Normalmente el rastreador de GS toma sus registros de sitios donde la información se encuentra en libre acceso o de sitios comerciales que son procesados con el beneplácito de los editores, con los que previamente se ha llegado a algún tipo de acuerdo. Para entender la naturaleza de este producto mostramos algunos de los portales de información científica que cubre GS (Jacsó, 2005a; Meho; Yang, 2007):

- Repositorios: *arXiv.org*, *RePEc*, *E-Lis* o *CiteBase*.
- Portales de revistas: *HighWire Press*, *MetaPress*, *IngentaConnect*, *ACM Digital Library*.
- Bases de datos: *PubMed*.
- Editores comerciales: *Sage*, *Springer*, *Taylor & Francis*, *Nature*, *Blackwell*, *Macmillan*, *Wiley*, *Cambridge University Press*.
- Sociedades Científicas: *American Physical Society*, *American Chemical Society*, *Royal Society of Chemistry*.
- Catálogos online de bibliotecas: *Worldcat*, *Dialnet*, *Institut de l'Information Scientifique et Technique (Inist)*.

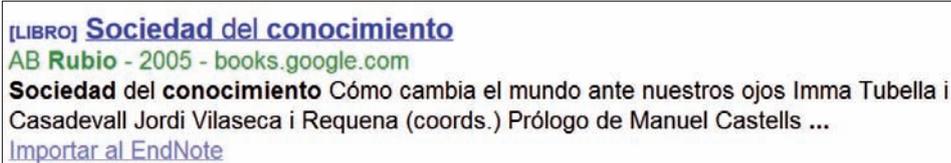


Figura 2. Búsqueda de un libro en GS remitiéndonos el enlace a Google Books

– Institutos y centros de investigación: *National Institutes of Health, NASA, American Institute of Physics.*

Por supuesto a estas fuentes habría que sumar los propios productos de Google como *Google Patents*, y sobre todo *Google Book Project* (figura 2), que ha escaneado ya más de un millón de ejemplares procedentes de 20.000 editoriales y bibliotecas en más de cien idiomas. Hay que señalar que el 10% de los mismos están escritos en español. Esto tiene mucha trascendencia ya que gran parte de los libros escaneados provienen de los fondos de bibliotecas académicas de las universidades del más alto prestigio como Standford, Princeton, Oxford, Harvard o Cornell o incluso de las colecciones de los servicios de publicaciones de las universidades como ocurre en el caso de la *Universidad de Salamanca*. Esta cobertura tan diversa de fuentes de información hace que podamos encontrar en GS una gran gama de tipos documentales:

- Libros
- Artículos en revistas
- Comunicaciones y ponencias a congresos
- Informes científico-técnicos
- Tesis y tesinas o memorias de grado
- Trabajos científicos depositados en repositorios o archivos de preprints
- Sitios web gubernamentales e institucionales
- Cualquier publicación con resumen

Quedan excluidos documentos no científicos como las reseñas de libros y editoriales, libros de texto, periódicos y revistas comerciales.

“Los datos de GS no tienen ninguna normalización, consecuencia de la amplia cobertura, la variedad de fuentes de información y el procesamiento automático de la información”

Una de las ventajas del exhaustivo rastreo de la literatura científica de GS es que alumbra un corpus documental antes casi invisible que de otro modo sería

difícilmente recuperable (**Robinson; Wusteman, 2007**) al menos conjuntamente, y además permite hallar trabajos, sobre todo preprints, mucho antes de que aparezcan publica-

dos en las revistas científicas comerciales (**Schroeder, 2007**). Sin embargo, el rastreo automático e indiscriminado conlleva también una importante limitación: muchos de los documentos indizados distan mucho del concepto de académico (**Noruzi, 2005**). No está claro qué entiende GS por “scholar” por lo que en ocasiones se incluyen entre sus resultados guías de bibliotecas, bibliografías de asignaturas o documentos administrativos. Esto se produce debido a que se suele indizar toda aquella información que cuelga de un dominio académico y el motor es incapaz de distinguir los tipos documentales propiamente científicos o académicos (**Friend, 2006**). Esta cuestión es importante ya que no tiene el mismo significado ser citado por un documento científico (libro, artículo, tesis...) que por otro que no lo es (programa de una asignatura...).

Asimismo, y esto es muy trascendente desde el punto de vista científico, aparecen mezcladas las citas provenientes de revistas arbitradas, es decir, las sometidas a *peer review*, con otras que no emplean ningún sistema de selección y evaluación de los manuscritos que publican. Para **Jacsó (2008b)** esta situación debería tenerse en cuenta a la hora de construir los indicadores bibliométricos ya que éstos tratan de medir el impacto científico a partir de fuentes de acreditada solvencia. Es evidente que GS, al incluir indiscriminadamente todas las citas que es capaz de identificar en cualquier documento, no puede asegurar ningún control de calidad de la información científica que presenta. Esta es la diferencia entre un entorno controlado (bases de datos tradicionales) y uno incontrolado (GS).

Independientemente de los errores que pueda cometer GS en el proceso de indización está claro que su cobertura documental genera un universo de citación diferente al de las otras bases de datos, con una serie de citas que son exclusivamente suyas. Algunos estudios han tratado de valorar el total de citas que puede aportar; por ejemplo **Kousha y Thelwall (2007)** sobre una muestra de 882 trabajos de diferentes áreas muestran como GS rescata 5.589 citas a los mismos, mientras que WoS recuperaba 4.184, con un solapamiento entre ambos de 2.387 referencias bibliográficas (es decir, el 24% del total de citas es común a ambas bases de datos –el 57% de WoS y el 43% de GS). Significa por tanto que GS recupera 3.202 citas únicas, aunque también pierde 1.797 respecto a WoS. Sin embargo, este solapamiento con WoS varía entre las diferentes áreas

científicas: en biología, física e informática gira en torno al 60%, mientras que en química se reduce al 33%. También **Meho** y **Yang** (2006) sobre 1.093 artículos de documentación compararon *GS*, *WoS* y *Scopus* determinando que entre las tres el total de citas únicas era de 5.288 (figura 3). *GS* localizaba un total de 4.184 mientras que las otras dos bases de datos conjuntamente sólo recuperaban 2.733. El solapamiento fue también del 24% (1.629/6.917).

Los datos por tanto parecen apuntar a que *GS* recupera un determinado número de citas únicas dependiendo de las disciplinas. Junto a esta situación, hay indicios además de que diversos tipos documentales podrían verse favorecidos con mayor citación como es el caso de los libros. **Harzing** y **Van-der-Wal** (2008) aportan algunas evidencias hacia una mejor cobertura de las citas recibidas por libros. Así, tomando las diez monografías ganadoras del *Terry Book Award*, *WoS* identificaba un total de 368 citas recibidas por estos libros mientras que *GS* elevaba la cifra hasta 783, lo cual supone un incremento del 128%. En definitiva, sabemos que *GS* no sólo es capaz de recuperar más citas sino que ofrece más citas únicas. Por ello es muy relevante conocer con cierta precisión el origen de las mismas desde una perspectiva documental.

Kousha y **Thelwall** (2008) en un estudio que recopila los trabajos publicados en 39 revistas de acceso abierto intentan revelar precisamente a qué tipos documentales corresponden las citas rescatadas por *GS* y que no son identificadas por otras bases de datos. Según los datos de estos autores, de un total de 5.589 citas, el 35% provenía de revistas científicas, el 25% de congresos/seminarios, un 22% de trabajos depositados

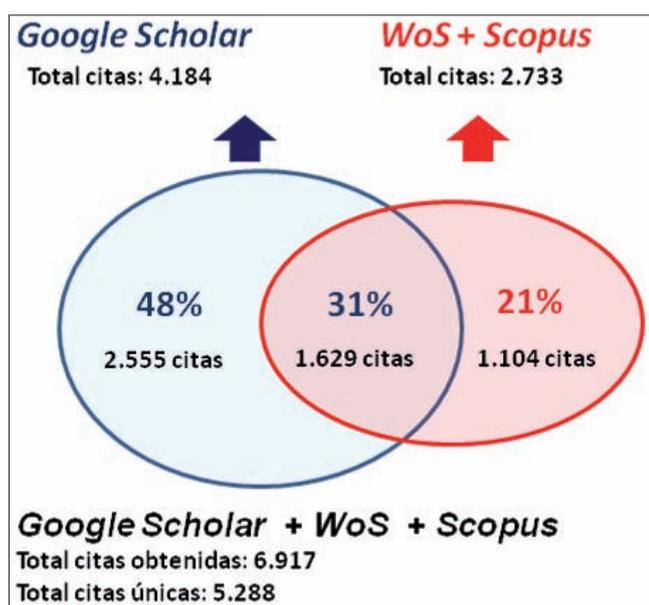


Figura 3. Ejemplo de Meho y Yang (2006) del solapamiento de las citas proporcionadas por *GS*, *Scopus* y *WoS* en el campo de la documentación

“Para la elaboración de un mismo ranking bibliométrico el procesamiento de los datos con *WoS* lleva 10 horas, con *Scopus* 20 y con *GS* 300”

en repositorios y, por último, un 8% de tesis doctorales. Matizan además que en función del área científica estos porcentajes varían. Así por ejemplo en física las citas recibidas de eprints/preprints llegan hasta el 48% y en informática las de congresos/seminarios se elevan al 43%. Estos datos son interesantes porque efectivamente reflejan que *GS* se adapta mejor a las prácticas de las distintas disciplinas, sobre todo aquellas que no utilizan como vía preferente de publicación las revistas científicas (humanidades, ciencias sociales, ingenierías).

En el trabajo de **Meho** y **Yang** (2006), referido a documentación, mientras que el 82% de las citas recuperadas por *WoS/Scopus* correspondían a artículos de revistas y el 18% restante a congresos, en *GS* eran del 43% y 34% respectivamente. Las tesis representaban el 10%, los libros el 6%, los informes el 5%, y otros documentos el 4%. Por tanto, las diferencias son claras: *GS* recupera citas de muy diversas fuentes, siendo las citas de revistas y libros, en porcentajes parecidos, las dominantes.

5. Cobertura de *GS* de revistas científicas

Pese a la capacidad de incorporar otros tipos documentales las revistas científicas siguen siendo el medio fundamental de comunicación por lo que conviene aclarar cuál es la cobertura de *GS* al respecto. Para lograr una perfecta cobertura de una revista, *GS* intenta llegar a acuerdos con editoriales como ocurre con *Science* o *Nature Publishing Group*, lo cual asegura una correcta indización.

No obstante, no siempre es así: *Elsevier*, la principal multinacional de la edición de revistas científicas en el mundo, se ha mostrado reticente ya que esta editorial es la que mantiene *Scopus*, que se puede considerar competencia directa de *GS* (**Meho**; **Yang**, 2007). Pero a pesar de ello, según **Bakkalbasi** et al. (2006) la mayor parte de los contenidos de *Elsevier* son indizados por terceros como el servicio *Ingenta*. Que la indexación de un grupo de revistas asociadas a un determinado editor dependa de un acuerdo es delicado ya que por ejemplo *GS* tampoco incluye las revistas de la *American Chemical Society* (*ACS*).

Esta situación provoca que determinadas disciplinas puedan presentar sesgos de bulto. En el caso de la química, que no esté un editor tan determinante como *ACS* provoca que de las citas recibidas por 373 artí-

culos publicados en estas revistas *GS* sólo sea capaz de rescatar 2.804 de un conjunto total de 8.723 citas recibidas (Bornmann et al., 2009). Un aspecto que conviene aclarar es que cuando un editor deja que *GS* incorpore sus datos no quiere decir que incorpore el texto completo de sus trabajos y la citación que generan los mismos. En la mayor parte de las ocasiones la información se reduce a proporcionar una mera referencia bibliográfica de los contenidos de las revistas.

En cualquier caso es interesante conocer cuál es la cobertura de *GS* respecto a otras bases de datos. Hemos de tener en cuenta que si éstas son empleadas como herramienta de evaluación es esencial conocer el universo de revistas empleado. En este sentido la política de *GS* es oscura ya que no proporciona ninguna información de cuáles son las revistas y qué tipo de indización tiene de cada una de ellas; tampoco sabemos a ciencia cierta cuáles son los editores que han firmado acuerdos con *GS*, información fundamental para conocer la validez de cualquier tipo de material científico (Bauer; Bakalbasi, 2005).

Para solventar este problema las bases de datos bibliográficas presentan los denominados *Master List*, una información muy apreciada por bibliotecas y evaluadores. Ante la desinformación de *GS*, diversos trabajos han tenido como finalidad la comparación de su cobertura con la de otras bases de datos. Así por ejemplo Mayr y Walter (2007) estudiaron cuántas revistas de *WoS* están presentes en *GS*, determinando que del *Science Citation Index (SCI)* tiene el 85% (3.244) y del *Social Science Citation Index (SSCI)* el 88% (1.666).

Uno de los trabajos que mejor refleja la cobertura de las revistas científicas por parte de *GS* es el conducido por Neuhaus et al. (2006), que comprueba las revistas indizadas en 47 bases de datos de diversos campos. Estos autores indican que la cobertura de *GS* de diferentes disciplinas no es homogénea: cubre el 10% de las revistas de humanidades, el 39% de ciencias sociales, el 41% de educación, el 52% de economía y el 76% en ciencia y medicina. Por otra parte, en estos análisis se revela como *GS* tiene casi una cobertura total de las revistas en acceso abierto identificadas por diferentes directorios y bases de datos como el *Directory of Open Access Journals (DOAJ)* o *ACM Digital Library*. Asimismo *GS* incluye todas las revistas de *Pubmed* y *Pubmed Central* (Neuhaus et al., 2006).

Sin embargo, como hemos comentado con anterioridad, el hecho de que una revista esté presente en *GS* no significa que se permita el acceso al original ni que estén indizadas las referencias y citas de dicha revista; de ahí que *GS* pierda el 40% de la citación de revistas científicas de *WoS* y *Scopus* (Meho; Yang, 2007). Por último conviene señalar que mientras en *WoS* la se-

lección de revistas se basa en un riguroso proceso de identificación de las más relevantes del mundo, *GS* no parece seguir ninguna directriz, por lo que tiene cabida cualquier revista, independientemente de su calidad.

6. El problema de la normalización y las búsquedas en *GS*

Consecuencia de la amplia cobertura, la variedad de fuentes de información empleadas y el procesamiento automático de la información es la ausencia de normalización en los datos de *GS*. Si en las bases de datos en general ya hay enormes limitaciones en la normalización de campos tan básicos como los autores o las instituciones, el problema cobra aquí mayores dimensiones. Si comparamos la información altamente estructurada de productos como *WoS* o *Scopus* podríamos decir que *GS* es un auténtico banco de datos tóxico que le resta credibilidad y le aleja de ser un producto consistente. Con el fin de orientar ante su posible uso evaluativo recopilamos algunos de sus errores más comunes y que aparecen bien documentados en la literatura científica sobre el tema (Jacsó, 2005a, 2005b, 2008a, 2008b).

El principal inconveniente de *GS* radica en que su herramienta de indización intenta detectar los campos que componen los documentos de forma automática, pero este proceso de identificación de estructuras no siempre funciona. En ocasiones se toman como autores de un trabajo elementos constitutivos del cuerpo del texto: por ejemplo, si desde la búsqueda avanzada introducimos como autor el texto "introducción" devolvemos un total de 7.160 trabajos (figura 4). En este caso se toma el inicio de un epígrafe como autor. Una búsqueda por autor con "estado de la cuestión" da 1.330 resultados; igualmente ocurre con "índice" (4.320) o "contenido" (6.180). Inexplicablemente el campo autor presenta otros errores: el término "i-netlibrary" aparece en 12.200 ocasiones como firmante cuando el término ni siquiera aparece referenciado en los textos. El problema no sólo radica en una indexación automática sino en la absoluta ausencia de vocabularios controlados e índices (Schroeder, 2007). No hay siquiera un control de los títulos de las revistas (por ejemplo aparece tanto *BMJ* como *British Medical Journal*, *JAMA* y *Journal of the American Library Association*) ni de palabras ni términos clave como por ejemplo los *Medical Subject Headings (MeSH)* de *Medline* (Shultz, 2007), herramientas fundamentales para la recuperación pertinente de información científica.

También el campo del año de publicación provoca errores, y la búsqueda avanzada acotada por años genera resultados incomprensibles. En el momento de la realización de este trabajo si buscamos simplemente el período 2006-2008 *GS* devuelve un total de 93.900 documentos; sin embargo, al ampliar a 2005-2008 se



Figura 4. Problemas de indexación del nombre de los autores en GS

reduce a 89.800. Una búsqueda para el período 2004-2008 devuelve una cifra de 139.000 documentos, sin embargo una búsqueda entre 2000-2008 reduce los documentos a 109.000. Otro de los inconvenientes de las fechas es que a veces son identificadas erróneamente tomando como fecha de publicación del documento la fecha de depósito del mismo en un repositorio o incluso en otras ocasiones el número ISSN de las revistas. En general, cualquier número con cuatro dígitos es susceptible de ser confundido por GS con el año de publicación.

Otro de los inconvenientes encontrados es la enorme presencia de trabajos duplicados en los resultados, lo que crea confusión. Los duplicados se producen básicamente por las diferentes versiones que un artículo puede tener y que GS no ha sabido reagrupar bajo un mismo encabezamiento de título, por lo que aparecen como trabajos diferentes. Se pueden duplicar citas al estar presente la versión en preprint de un trabajo indizado en un repositorio y/o en la página personal del autor y la versión final publicada en una revista científica. Como consecuencia si comparamos la producción de una misma revista en WoS y en GS los resultados difieren bastante, situación que no se produce en otras bases de datos (WoS o Scopus). Si buscamos la producción en 2008 de una revista como *Lancet* GS devuelve 3.250 referencias, *Scopus* 1.653 y *WoS* 1.688. Es decir GS casi duplica los registros.

La consecuencia de estas incoherencias es un enorme aumento del coste en el tratamiento de datos derivados de GS. Para realizar un ranking bibliométrico (trabajos, citas, h-index) de 24 profesores del campo de la documentación, con WoS el procesamiento de los datos nos lleva 1 hora, 2

horas con Scopus y 30 horas si decidimos emplear GS (Meho; Yang, 2007). Esta situación pone en evidencia que, por el momento, su utilización a media y gran escala como herramienta de evaluación científica supone un consumo de recursos tan grande que la inhabilita. Sin embargo, ya empieza a haber algunas solu-

ciones, como el software *Publish or Perish*, que facilita en alguna medida la recopilación y la manipulación de datos extraídos de GS. Este software ideado por **Harzing** proporciona indicadores asociados a los resultados, y hace posible la elaboración de rankings por diferentes campos y exportarlos a otros formatos como xls. La limitación principal radica en el elevado coste de limpieza de datos (normalización, eliminación de duplicados...) y en que no permite descargar las citas.

7. Consideraciones finales

Finalmente en la tabla 2, a modo de guía, se recopilan algunas de las particularidades de GS presentadas a lo largo de este trabajo frente a los índices de citas WoS y Scopus. En la misma se evidencia que GS es un producto ambicioso desde el punto de vista de su cobertura pero mal resuelto en el plano del procesamiento de la información y su presentación. En cualquier caso conviene señalar que ninguna base de datos tiene una cobertura completa de las citas que se emiten y cada una de ellas presenta un universo completamente diferente. Por esta situación los índices de citas disponibles en la actualidad son productos complementarios entre



Figura 5. Diferentes versiones de un mismo trabajo agrupadas por GS bajo encabezamiento de título único

Google Scholar		Índices de citas multidisciplinares (Web of Science; Scopus)	
PRECIO			
▲	Libre acceso	▼	Pago de licencias
COBERTURA GENERAL			
▼	Falta de transparencia en la cobertura. No se declaran acuerdos con editoriales ni las fuentes que se indizan	▲	Transparencia absoluta en las fuentes que componen las bases de datos. Disponibilidad de <i>Master Lists</i> actualizadas
▲	Cobertura de una amplia tipología de fuentes de información: repositorios, bases de datos, sociedades científicas, catálogos online de bibliotecas, institutos de investigación, productos de Google (<i>Google Patents</i> y <i>Google Books</i>)		
▲	Posibilidad de encontrar diversos tipos documentales: preprints, artículos de revistas, libros, tesis, informes, comunicaciones a congresos...	▼	Sólo cubren los contenidos de revistas científicas y recientemente libros de actas de congresos (<i>WoS: Conference Proceedings Citation Index</i>)
▼	Cobertura de documentos que podrían no ser de carácter académico: guías de biblioteca, temarios, etc.	▲	Contenidos exclusivamente científicos y mayoritariamente sometidos a revisión
▲	Buena cobertura de literatura en lenguas nacionales europeas	▼	Dominio de la literatura de carácter anglosajón, especialmente en <i>WoS</i>
▲	Acceso directo a publicaciones científicas a texto completo y gratuitas	▼	Acceso sólo a la referencia de los artículos
▲	Acceso directo al documento si la biblioteca lo tiene contratado	▲	Acceso directo al documento si la biblioteca lo tiene contratado
▲	Localiza citas emitidas por documentos no cubiertos por otras bases de datos, especialmente desde preprints, congresos o tesis doctorales. Esta característica lo hace especialmente útil para las siguientes disciplinas: humanidades, ciencias sociales e ingenierías	▼	Sólo localiza citas de revistas y congresos
COBERTURA DE REVISTAS CIENTÍFICAS			
▼	No existe ningún tipo de control en la selección de las revistas que indiza, por lo que todo tipo de revistas tiene cabida	▲	Rigurosos proceso de selección de las revistas científicas, especialmente en <i>WoS</i>
▼	Mala cobertura de las revistas de humanidades y ciencias sociales presentes en otras bases de datos (<i>MLA Bibliography, Philosopher's Index, PsycInfo, Sociological Abstracts...</i>)	▼	Tradicional mala cobertura de revistas de humanidades y ciencias sociales, aunque en la actualidad tienen una política de expansión en estos campos del conocimiento
INTERFAZ, BÚSQUEDAS y RESULTADOS			
▼	Sólo ofrece tres campos de búsqueda (autor, revista y año de publicación)	▲	Posibilidad de buscar en 12 campos diferentes en <i>WoS</i> y 17 en <i>Scopus</i>
▼	No tiene ninguna herramienta para analizar resultados	▲	Herramientas de análisis bibliométrico on-line como <i>Citation Report</i> en <i>WoS</i>
▼	Los resultados se presentan directamente ordenados y no existen otras opciones	▲	Permiten ordenar los resultados según diferentes opciones (título, nº de citas, fecha de publicación, primer autor...)
▼	Sólo permite exportar los resultados, uno a uno, a un software bibliográfico	▲	Exportación de los resultados en diferentes formatos (RIS, txt, tabulados, etc.)
▼	Gran coste en el procesamiento de los datos, lo que hace difícil su uso en estudios de gran escala	▲	Procesamiento de la información con menores costes en horas
▼	Presenta gran variedad de resultados duplicados		
▲	Posibilidad de exportar los resultados a software de análisis de datos: <i>Publish or Perish</i>	▲	Posibilidad de exportar los resultados a software de análisis de datos: <i>Histcite, Refviz, NWB, BibExcel</i>
▼	Sólo se muestran los 1.000 primeros documentos recuperados en cada consulta	▲	Se pueden consultar todos los resultados que genera una búsqueda
▲	Localiza las diferentes versiones de un documento y las agrupa bajo un mismo encabezamiento de título.		
▼	No identifica ante qué tipo documental nos encontramos. Tan sólo identifica los libros	▲	Cada registro está clasificado en un tipo documental (artículo, revisión, carta, nota, recensión, etc.)
▼	Sólo incluye el filtrado por 7 disciplinas	▲	Incluyen diversas opciones de filtrado (disciplina, año, tipo documental) que permiten refinar las búsquedas
CONTROL DE LA INFORMACIÓN			
▼	No existe normalización de los autores.	▲	No existe normalización pero tienen herramientas para identificación de autores (<i>WoS=Author Finder</i>)
▼	Ausencia de cualquier tipo de vocabulario controlado. No existe control de las revistas científicas; éstas pueden aparecer indizadas de diferente forma	▲	Control absoluto de las revistas científicas

Tabla 2. Comparación de las principales características de GS con las bases de datos multidisciplinares WoS y Scopus

sí. Mientras que hay un cierto consenso entre la comunidad científica en el uso de *WoS* como herramienta de evaluación, *GS* se muestra por el contrario como un producto inmaduro. Por esta razón se desaconseja su utilización como única fuente de información para la evaluación de la ciencia, especialmente en trabajos de media-gran escala (instituciones, países).

“GS es un producto ambicioso desde el punto de vista de su cobertura pero mal resuelto en el plano del procesamiento de la información y la presentación de resultados”

Ahora bien, creemos que *GS* es útil a nivel micro, como ayuda a los autores e investigadores concretos en la búsqueda rápida, fácil y directa de documentos a texto completo, y en la identificación de citas a sus trabajos. Sobre todo es útil para la literatura no anglosajona, que es la peor controlada por los sistemas de información dominantes en el mundo de la ciencia, para las disciplinas que no emplean preferentemente las revistas como medio de comunicación (ingenierías, humanidades, ciencias sociales...) y para localizar citas a libros, tesis, informes y a artículos publicados en revistas secundarias no incorporadas a la llamada “corriente principal de la ciencia”.

8. Referencias

- Bakkalbasi, Nisa; Bauer, Kathleen; Glover, Janis; Wang, Lei.** “Three options for citation tracking: *Google Scholar*, *Scopus* and *Web of Science*”. *Biomedical digital libraries*, 2006, v. 3, n. 7. <http://www.bio-diglib.com/content/3/1/7>
- Bauer, Kathleen; Bakkalbasi, Nisa.** “An examination of citation counts in a new scholarly communication environment”. *D-Lib magazine*, 2005, v. 11, n. 9. <http://www.dlib.org/dlib/september05/bauer/09bauer.html>
- Bornmann, Lutz; Marx, Werner; Schier, Hermann; Rahm, Erhard; Thor, Andreas; Daniel, Hans-Dieter.** “Convergent validity of bibliometric *Google Scholar* data in the field of chemistry. Citation counts for papers that were accepted by *Angewandte Chemie International Edition* or rejected but published elsewhere, using *Google Scholar*, *Science Citation Index*, *Scopus*, and *Chemical Abstracts*”. *Journal of informetrics*, 2009, v. 3, n. 1, pp. 27-35. <http://lips.informatik.uni-leipzig.de/files/2009-0.pdf>
- Friend, Frederick.** “*Google Scholar*: potentially good for users of academic information”. *Journal of electronic publishing*, 2006, v. 9, n. 1. http://eprints.ucl.ac.uk/1771/1/JEP_OA_GS.pdf
- Hartman, Karen; Mullen, Laura-Bowering.** “*Google Scholar* and academic libraries: an update”. *New library world*, 2008, v. 109, n. 5-6, pp. 211-222. <http://eprints.rclis.org/13820/1/GSfinalupdate.pdf>
- Harzing, Anne-Wil K.; Van-der-Wal, Ron.** “*Google Scholar* as a new source for citation analysis”. *Ethics in science and environmental politics*, 2008, v. 8, n. 1, pp. 61-73. <http://www.int-res.com/articles/esepp2008/8/e008p061.pdf>
- Jacsó, Péter.** “As we may search - Comparison of major features of the *Web of Science*, *Scopus*, and *Google Scholar* citation-based and citation-enhanced databases”. *Current science*, 2005a, v. 89, n. 9, pp. 1537-1547. <http://www.ias.ac.in/currensci/nov102005/1537.pdf>
- Jacsó, Péter.** “*Google Scholar*: the pros and the cons”. *Online information review*, 2005b, v. 29, n. 2, pp. 208-214. <http://www.jacso.info/PDFs/jacso-google-scholar-pros-and-cons.pdf>
- Jacsó, Péter.** “*Google Scholar* revisited”. *Online information review*, 2008a, v. 32, n. 1, pp. 102-114. <http://www.jacso.info/PDFs/jacso-GS-revisited-OIR-2008-32-1.pdf>
- Jacsó, Péter.** “The pros and cons of computing the h-index using *Google Scholar*”. *Online information review*, 2008b, v. 32, n. 3, pp. 437-452. <http://www.jacso.info/PDFs/jacso-pros-and-cons-of-computing-the-h-index.pdf>
- Kousha, Kayvan; Thelwall, Mike.** “*Google Scholar* citations and *Google web/url* citations: a multi-discipline exploratory analysis”. *Journal of the American Society for Information Science and Technology*, 2007, v. 58, n. 7, pp. 1055-1065.
- Kousha, Kayvan; Thelwall, Mike.** “Sources of *Google Scholar* citations outside the *Science Citation Index*: a comparison between four science disciplines”. *Scientometrics*, 2008, v. 74, n. 2, pp. 273-294.
- Mayr, Philipp; Walter, Anne-Kathrin.** “An exploratory study of *Google Scholar*”. *Online information review*, 2007, v. 31, n. 6, pp. 814-830.
- Meho, Lokman I.; Yang, Kiduk.** “Multi-faceted approach to citation-based quality assessment for knowledge management”. *En: World library and information congress: 72nd IFLA General conference and council, 2006*.
- Meho, Lokman I.; Yang, Kiduk.** “Impact of data sources on citation counts and rankings of LIS faculty: *Web of Science* versus *Scopus* and *Google Scholar*”. *Journal of the American Society for Information Science and Technology*, 2007, v. 58, n. 13, pp. 2105-2125.
- Neuhaus, Chris; Neuhaus, Ellen; Asher, Alan; Wrede, Clint.** “The depth and breadth of *Google Scholar*: an empirical study”. *Libraries and the Academy*, 2006, v. 6, n. 2, pp. 127-141.
- Noruzi, Alireza.** “*Google Scholar*: the new generation of citation indexes”. *Libri*, 2005, v. 55, n. 4, pp. 170-180. <http://www.librijournal.org/pdf/2005-4pp170-180.pdf>
- Price, Gary.** *Google Scholar documentation and large pdf files*. 2004. <http://blog.searchenginewatch.com/041201-105511>
- Robinson, Mary L.; Wusteman, Judith.** “Putting *Google Scholar* to the test: a preliminary study”. *Program*, 2007, v. 41, n. 1, pp. 71-80. <http://www.ucd.ie/wusteman/articles/robinson-wusteman.pdf>
- Schroeder, Robert.** “Pointing users toward citation searching: using *Google Scholar* and *Web of Science*”. *Libraries and the academy*, 2007, v. 7, n. 2, pp. 243-248.
- Shultz, Mary.** “Comparing test searches in *PubMed* and *Google Scholar*”. *Journal of the Medical Library Association*, 2007, v. 95, n. 4, pp. 442-445. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2000776>

Daniel Torres-Salinas

Grupo Evaluación de la Ciencia y la Comunicación Científica (EC3), Centro de Investigación Médica Aplicada, Universidad de Navarra, Avda. Pío XII, 31008 Pamplona (España).

torressalinas@gmail.com

Rafael Ruiz-Pérez; Emilio Delgado-López-Cózar

Grupo Evaluación de la Ciencia y la Comunicación Científica (EC3), Departamento de Biblioteconomía y Documentación, Universidad de Granada, Campus Cartuja, 18071 Granada (España)

r Ruiz@ugr.es

edelgado@ugr.es

Utilización de las bases de datos de patentes como instrumento de vigilancia tecnológica

Por Antonio Hidalgo-Nuchera, Santiago Iglesias-Pradas y Ángel Hernández-García

Resumen: Ante la importancia que adquieren la innovación y el desarrollo tecnológico en la economía de un país, es fundamental contar con indicadores que permitan medir el grado de innovación de sus empresas. Los indicadores relativos a las patentes proporcionan una excelente medida de esta capacidad y son una destacada fuente de información para organizaciones innovadoras, aunque se han llevado a cabo pocos estudios que midan esta utilidad. Se plantea, en base a un análisis realizado a lo largo de 2007, el nivel de uso de las bases de datos de patentes orientado a la vigilancia tecnológica por parte de más de medio millar de empresas españolas que presentaron al menos una patente en el período 1988-1998. Los resultados se han analizado tanto globalmente como en función del tamaño y sector de actividad económica de la empresa.



Antonio Hidalgo-Nuchera es profesor titular de Organización de empresas en la Escuela Técnica Superior de Ingenieros Industriales de la Univ. Politécnica de Madrid y director del grupo de investigación en Innovación, Propiedad Industrial y Política Tecnológica (Innopro). Ha participado en diferentes proyectos financiados por la Comisión Europea y organismos nacionales sobre innovación tecnológica, transferencia de tecnología y desarrollo de políticas tecnológicas.



Santiago Iglesias-Pradas es profesor de Organización de empresas de la Universidad Politécnica de Madrid. Imparte asignaturas de grado y postgrado en las áreas de dirección de empresas y de tecnologías y sistemas de información empresariales en distintas universidades. Es doctor ingeniero de telecomunicación y MBA por la UPM, ha participado en diferentes proyectos de investigación nacionales e internacionales y pertenece al grupo Innovación, Propiedad Industrial y Política Tecnológica (Innopro).



Ángel Hernández-García es ingeniero de telecomunicación por la Univ. Politécnica de Madrid y trabaja como investigador doctoral en la misma universidad. Sus líneas de investigación están orientadas a la innovación tecnológica, la aplicación de modelos de aceptación de la tecnología en mercados electrónicos, caracterización de contenidos digitales en dispositivos móviles y análisis de factores de influencia en el rendimiento académico en sistemas e-learning.

Palabras clave: Patentes, Innovación, Vigilancia tecnológica, Bases de datos, Servicios de información, Indicadores.

Title: Use of patent databases as a tool for technology watch

Abstract: Given the importance of innovation and technology development in a country's economy, there has been an increasing need for selecting indicators to measure the degree of innovation of its companies. The indicators related to patents offer an excellent measure of this capability and can be used as an important source of information by innovative organizations, although the studies assessing this utility are scarce. This study, conducted throughout 2007, proposes a technology watch oriented analysis of the level of use of patent databases from more than 500 Spanish companies that have issued at least one patent in the period 1988-1998, followed by discussion of the study results. The analysis took into account global data, company size and sector of activity.

Keywords: Patents, Innovation, Technology watch, Databases, Information services, Indicators.

Hidalgo-Nuchera, Antonio; Iglesias-Pradas, Santiago; Hernández-García, Ángel. "Utilización de las bases de datos de patentes como instrumento de vigilancia tecnológica". *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 511-519.

DOI: 10.3145/epi.2009.sep.04

1. Introducción

Un factor destacado en la competitividad de la economía de un país es su capacidad de estimular la innovación y el desarrollo tecnológico. La innovación

constituye una de las claves del crecimiento económico a largo plazo, por lo que es necesario identificar en qué medida es de origen nacional o transferida mediante licencias, importaciones, imitaciones o inversiones extranjeras directas.

Artículo recibido el 09-06-09

Aceptación definitiva: 11-08-09

Este interés se ha consolidado debido a que los elementos que intervienen en el proceso de desarrollo tecnológico, cuando éste se estudia desde la óptica de la acumulación de conocimientos y el aprendizaje, abarcan un campo de variables mucho más amplio (Freeman; Soete, 1982; Dosi, 1988). También ha contribuido a ello la consideración, a partir de mediados de los ochenta, de la innovación como un proceso independiente e interactivo (Kline; Rosenberg, 1986), con lógica propia y al que contribuyen de forma relevante pero no exclusiva, las tareas propias de investigación. El *Manual de Oslo* (OCDE, 2005) ha mejorado la interpretación de la medición de iniciativas de innovación tecnológica, y ha propiciado la elaboración de las recientes *Encuestas de Innovación Tecnológica* del CIS apoyadas por la *Comisión Europea*.

Las patentes constituyen los indicadores del output de la organización que aportan la información más relevante sobre el conjunto del proceso de innovación tecnológica. Desde los primeros trabajos realizados utilizando patentes como indicadores de actividades tecnológicas (Schmookler, 1966), el uso de la información que contienen ha experimentado un cambio más que significativo, de forma que en la actualidad es casi imposible encontrar trabajos sobre los procesos de innovación o las capacidades tecnológicas de cualquier organización (empresa, universidad, centro de investigación) sin que se utilicen las patentes como indicadores.

No obstante, este uso lleva asociado de forma implícita un conjunto de dificultades que es necesario tener en cuenta (Basberg, 1987; Pavitt, 1998; Archibugi, 1992):

- Sólo una parte de las innovaciones se patentan, ya que hay otros mecanismos diferentes para proteger las invenciones, por ejemplo, el secreto industrial.
- Algunas invenciones no son patentables, o porque la legislación las excluye expresamente o porque no cumplen alguno de los requisitos necesarios (novedad, actividad inventiva, aplicación industrial).
- En determinados sectores el ciclo de vida de los productos es muy corto y no es interesante patentar.

En todo caso y a pesar de las dificultades enumeradas, son muchas las ventajas que conlleva utilizar las patentes como indicadores de la actividad tecnológica de la organización, ya que representan la culminación de un esfuerzo por parte de ésta que pone de manifiesto no sólo la intensidad de las labores inventivas, sino la capacidad de implementar las propias competencias tecnológicas. Entre estas ventajas se encuentra el hecho de que se dispone de series de datos que cubren períodos de tiempo relativamente largos y que contie-

nen información detallada (incluyendo los datos de la empresa y del sector industrial). Estos datos pueden ser utilizados para analizar la actividad nacional de innovación con un elevado nivel de precisión.

“Las patentes aportan la información más relevante sobre el conjunto del proceso de innovación tecnológica”

Esta información tecnológica se encuentra en las bases de datos de organismos nacionales e internacionales como la *Oficina Española de Patentes y Marcas* (Oepm), la *Oficina Europea de Patentes* (EPO), la *Oficina Norteamericana de Patentes* (Uspsto), la *Organización Mundial de la Propiedad Intelectual* (Ompi), etc. El análisis de los datos bibliográficos extraídos de un documento de patente puede proporcionar información muy valiosa. Por ejemplo, un análisis de la relación entre la tasa de denegación de patentes frente al número de solicitudes realizadas y sobre las causas que provocaron dicha denegación (qué requisito de patentabilidad falló), puede aportar una excelente información acerca de hasta qué punto las empresas, las universidades y los centros de investigación llevan a cabo labores de vigilancia tecnológica y cómo utilizan esta información antes de aportar fondos a proyectos de investigación. Asimismo, este análisis hace posible la obtención de información relacionada con otro conjunto de aspectos, entre los que destacan:

- La detección de las empresas que, debido a sus altas tasas de concesión frente a solicitudes, son compañías ejemplares en su sector de actividad en la utilización de la propiedad industrial como forma de proteger la innovación, así como en el uso de fuentes de información tecnológica para evitar la duplicidad de esfuerzos.
- Cómo y cuándo las compañías utilizan las bases de datos de patentes, es decir, se puede medir hasta qué punto la información de las patentes es utilizada antes y durante la investigación.
- La calidad de una patente en función de los requisitos de patentabilidad.

2. Objetivos y metodología

Teniendo en cuenta estas consideraciones, la finalidad del proyecto fue analizar el grado en que el colectivo de empresas españolas con al menos una patente concedida y que fuera solicitada en el período 1988-

1998 ha utilizado las bases de datos de patentes hasta el año 2007 como instrumento de vigilancia tecnológica. Para ello, se ha tomado como punto de partida la siguiente información:

– Patentes con prioridad española (o sea, España primer país donde se solicita la patente) que, habiendo sido solicitadas en la *Oficina Europea de Patentes* entre los años 1988 y 1998, han sido concedidas hasta la fecha de realización del estudio (se ha de tener en cuenta que la concesión de estas patentes conlleva un plazo medio de aproximadamente 5 años). En este colectivo se incluye también a empresas e inventores extranjeros que protegen sus invenciones en España en primer lugar, bien por estrategia de protección o bien porque la invención es realizada en España.

– Solicitudes de patentes realizadas por sociedades españolas en el período 1988-1998 que han sido concedidas hasta 2007 (sin prioridad española).

“Son muchas las ventajas que conlleva utilizar las patentes como indicadores de la actividad tecnológica”

Para la realización del análisis se han utilizado los datos procedentes de la *Oficina Europea de Patentes*, ya que sus bases de datos *Epat* y *Espacenet* contienen información con la precisión y calidad suficiente para los propósitos fijados. La primera ofrece información básica de las solicitudes (fechas, solicitantes, organismos, etc.) y facilita la consulta de más de 830.000 referencias. Por su parte, *Espacenet* contiene los datos bibliográficos y los documentos completos en formato pdf de las patentes publicadas en los últimos años en cualquier estado miembro de la *Organización Europea de Patentes*, así como los de la *Oficina Europea de Patentes* y de la *Organización Mundial de la Propiedad Intelectual*. Su volumen es de unos 30 millones de documentos.

Del análisis de las bases de datos *Epat* y *Espacenet*, y de acuerdo con los criterios de selección mencionados, se ha obtenido para el período 1988-1998 un total de 5.046 patentes concedidas correspondientes a 546 empresas. La distribución de estas patentes en función del tipo de información referenciada ha sido la siguiente:

- concedidas con prioridad española o solicitante español: 4.197 registros;
- concedidas a inventores españoles fuera de España: 849 registros.

Para llevar a cabo el análisis se diseñó un cuestionario en el que se solicitaba a las empresas que realizaran una valoración cualitativa sobre el uso que hacen de los servicios de información sobre patentes mediante un conjunto de nueve preguntas dirigidas a los responsables del área tecnológica de la organización:

- a. ¿Cuál es la importancia de las patentes obtenidas por su empresa para su negocio?
- b. ¿Utiliza o dispone su empresa de servicios de información sobre patentes?
- c. ¿Con qué frecuencia utiliza su empresa estos servicios?
- d. ¿Para qué se utilizan estos servicios de información?
- e. ¿Cuáles son los servicios de información normalmente utilizados?
- f. En caso de acceder directamente a información sobre patentes, ¿puede especificar qué proveedores de bases de datos o servicios en internet utiliza?
- g. ¿Piensa su empresa comenzar a utilizar o incrementar la utilización de los documentos de patentes como fuente de información?
- h. ¿Podrían existir factores que impulsasen a su empresa a hacer más efectiva la búsqueda de información en las patentes?
- i. ¿Son importantes para el avance de su empresa otras formas de protección de la propiedad industrial?

El cuestionario fue enviado a las 546 empresas españolas, obteniéndose 123 respuestas, lo que representa el 22,5% de los cuestionarios enviados. El porcentaje más alto (58,5%) corresponde a compañías de más de 100 empleados, lo que resulta lógico desde la perspectiva de la gestión innovadora y de las capacidades tecnológicas, pues son las que poseen mayor nivel de recursos (tabla 1).

Nº empleados	<10	10-50	51-100	101-250	>250
% respuestas	9,7	13,0	18,8	21,1	37,4

Tabla 1. Distribución de las respuestas según el número de empleados (%)

En función de los sectores económicos, los datos obtenidos reflejan que el mayor nivel de respuesta se concentra en química-farmacia e ingeniería mecánica y maquinaria (tabla 2). Este resultado es consistente con el hecho de que estas industrias están condicionadas por la frecuente aparición de rupturas tecnológicas que implican la necesidad de patentar para protegerse de las mismas (Hidalgo; Molero, 2009).

Área	Número de empresas	Porcentaje (%)	Tamaño (%)				
			< 10	10-50	51-100	101-250	> 250
Ingeniería civil, bienes de consumo	21	17	14	0	29	29	29
Electricidad-electrónica	19	15	11	11	26	16	37
Ingeniería de proceso	7	6	0	29	14	0	57
Ingeniería mecánica y maquinaria	34	28	15	18	9	18	41
Instrumentación	6	5	33	0	0	33	33
Química-farmacia	36	29	0	17	22	25	36
Total	123	100	10	13	19	21	37

Tabla 2. Distribución de las respuestas según el sector económico (para mayor claridad los porcentajes que figuran en las tablas de este artículo se han redondeado)

3. Resultados obtenidos

3.1. Importancia de las patentes obtenidas para el negocio

De entre las seis opciones que se plantearon a las empresas para que indicaran su importancia a la hora de tomar una decisión para solicitar una patente, éstas consideraron como más crítica la relativa a que la patente es un instrumento de defensa de la propiedad industrial, en el 70% de las respuestas. Consideran con una importancia media la opción que indica que la patente proporciona mayor capacidad de negociación con otras compañías (42%), mientras que el resto de posibilidades son consideradas de forma mayoritaria como poco relevantes. Es de destacar que las empresas no consideran la patente como un mecanismo válido para fomentar la exportación de tecnología ni para llevar a cabo su licencia y obtener más ingresos. La tabla 3 refleja el resumen de estas opciones con la valoración mejor considerada (alta-3; media-2; baja-1).

Atendiendo al tamaño de las sociedades, la tabla 4 refleja para cada tipo la opción más valorada, destacando la importancia que conceden las grandes (>250 empleados) a la patente como elemento que ayuda a la exportación de tecnología y a la formalización del

conocimiento, factor este último que es muy bien valorado por las que tienen entre 11 y 50 empleados. Por su parte, las microempresas (5-10 empleados) valoran de forma positiva la patente como factor crítico de exportación.

El análisis de esta variable en función del sector económico al que pertenece la organización (tabla 5) pone de relieve que los sectores electricidad-electrónica, ingeniería de proceso, ingeniería mecánica y maquinaria, e ingeniería civil-bienes de consumo presentan

Opciones	Importancia	%
Instrumento de defensa de la propiedad industrial	Alta	70
Proporcionar a la empresa mayor capacidad para solicitar créditos	Baja	77
Proporcionar a la empresa mayor capacidad de negociación con otras empresas	Media	42
Permitir a la empresa exportar tecnología	Baja	37
Licenciar la patente para obtener mayores ingresos	Baja	53
Formalizar el conocimiento de la empresa	Baja	38

Tabla 3. Importancia de las patentes obtenidas por la empresa para su negocio

Opciones	5-10	11-50	51-100	101-250	> 250
Instrumento de defensa de la propiedad industrial	78 (alta)	56 (alta)	57 (alta)	84 (alta)	71 (alta)
Proporcionar a la empresa mayor capacidad para solicitar créditos	78 (baja)	69 (baja)	70 (baja)	80 (baja)	80 (baja)
Proporcionar a la empresa mayor capacidad de negociación con otras empresas	44 (media)	44 (baja)	39 (media)	36 (media)	50 (media)
Permitir a la empresa exportar tecnología	44 (alta)	44 (baja)	57 (baja)	40 (baja)	36 (media)
Licenciar la patente para obtener mayores ingresos	33 (baja)	44 (baja)	57 (baja)	60 (baja)	55 (baja)
Formalizar el conocimiento de la empresa	67 (baja)	38 (alta)	39 (baja)	40 (media)	32 (media)

Tabla 4. Importancia de las patentes en función del tamaño de la empresa (%)

Opciones	Bienes de consumo, ingeniería civil	Electricidad-electrónica	Ingeniería de proceso	Ingeniería mecánica y maquinaria	Instrumentación	Química-farmacia
Instrumento de defensa de la propiedad industrial	71 (alta)	74 (alta)	71 (alta)	62 (alta)	67 (alta)	75 (alta)
Proporcionar a la empresa mayor capacidad para solicitar créditos	86 (baja)	84 (baja)	57 (baja)	74 (baja)	50 (media)	86 (baja)
Proporcionar a la empresa mayor capacidad de negociación con otras empresas	52 (media)	42 (media)	29 (media)	47 (media)	50 (alta)	36 (media)
Permitir a la empresa exportar tecnología	38 (baja)	53 (baja)	43 (baja)	41 (baja)	83 (alta)	44 (media)
Licenciar la patente para obtener mayores ingresos	71 (baja)	58 (baja)	43 (baja)	59 (baja)	50 (alta)	39 (baja)
Formalizar el conocimiento de la empresa	57 (baja)	47 (baja)	29 (baja)	38 (media)	33 (alta)	33 (alta)

Tabla 5. Importancia de las patentes en función del sector económico (%)

una estructura de comportamiento semejante, caracterizada por la mayor relevancia del uso de la patente como instrumento de defensa de la propiedad industrial y la consideración de que ésta proporciona mayor capacidad de negociación con otras empresas. En los sectores de instrumentación y química-farmacia las compañías conceden además una gran importancia a la patente como elemento que permite la formalización del conocimiento de la empresa y la exportación de tecnología, lo que se debe en gran medida al mayor tamaño relativo de las sociedades que configuran estos sectores.

3.2. Uso de servicios de información sobre patentes por la empresa

Las respuestas obtenidas sobre uso o disponibilidad de algún servicio de información sobre patentes por parte de las empresas son afirmativas en su mayoría (90%), lo que refleja el alto nivel de interés en acceder a la información contenida en los documentos de patentes.

“Se concede una gran importancia a las patentes, fundamentalmente como instrumento de defensa de la propiedad industrial”

Respecto al tamaño de la empresa, las que tienen más de 51 empleados son las que más utilizan estos servicios, con un elevado porcentaje de respuestas afirmativas (por encima del 93%). Las organizaciones con menos de 50 empleados tienen un nivel de respuesta positiva del 75%, lo que resulta importante para este colectivo y pone de relieve el aumento del uso de la patente como instrumento de vigilancia tecnológica.

3.3. Frecuencia de utilización de los servicios de información sobre patentes

El conocimiento de la frecuencia en el uso de servicios de información sobre patentes por parte de las sociedades constituye un indicador del interés que tienen en actualizar su nivel tecnológico y acceder a información crítica para la evolución de su actividad, así como de la aceptación de una cultura basada en la vigilancia tecnológica (Escorsa; Maspons, 2001).

El análisis de las respuestas obtenidas pone de manifiesto que el 50% de las empresas utilizan los servicios de información sobre patentes con una frecuencia mensual (lo que puede considerarse suficiente), mientras que tan sólo el 6% hacen un uso poco frecuente de estos servicios (frecuencia inferior a un año). El resto categorizan la frecuencia de sus búsquedas en semanal (16%) y anual (21%). Por su parte, el análisis por sectores económicos muestra que en todos ellos la media relativa a la frecuencia de búsqueda más usada es de carácter mensual.

En función del tamaño de la empresa se observa que las de más de 51 empleados llevan a cabo sus procesos de búsqueda con una frecuencia mensual, mientras que esa frecuencia es menor para las más pequeñas, lo cual es coherente con el menor nivel de recursos de este colectivo (tabla 6).

Tamaño de la empresa	Frecuencia de uso (%)			
	Semanal	Mensual	Anual	>Anual
5-10	11	33	11	22
11-50	0	25	44	6
51-100	9	65	26	0
101-250	24	52	16	0
>250	21	59	14	2

Tabla 6. Frecuencia de uso de los servicios de información sobre patentes en función del tamaño de la empresa

3.4. Finalidad de los servicios de información sobre patentes

Del conjunto de compañías que afirman utilizar servicios de información sobre patentes, el 65% lo hace para realizar búsquedas previas al comienzo de un proyecto de I+D en su propia empresa evitando invertir en algo ya existente. Un porcentaje similar responde que los utiliza también para informarse sobre los avances de los competidores (63%) y para comprobar si una invención que la empresa ha realizado, está ya hecha con anterioridad (59%). Estas actuaciones se complementan además con el uso de las patentes para informarse acerca de nuevos avances tecnológicos o ideas innovadoras (tabla 7). En el lado opuesto se encuentran las respuestas que hacen referencia a que los servicios de información sobre patentes apenas son utilizados para llevar a cabo la divulgación de las innovaciones y solventar un problema tecnológico concreto de la empresa.

Del análisis de las respuestas tanto en función del número de empleados de las empresas como de los sectores económicos a los que pertenecen se deduce que, en general, el uso de los servicios de información sobre patentes se realiza fundamentalmente para conseguir los tres fines mejor valorados: informarse acerca de los avances de los competidores, realizar búsquedas previas al comienzo de un proyecto de I+D y comprobar la existencia de una invención realizada.

Finalidad	%
Comprobar si existe una invención que ha realizado la empresa	60
Informarse acerca de nuevos avances tecnológicos o ideas innovadoras	51
Informarse acerca de los avances de los competidores	63
Solventar un problema tecnológico concreto de la empresa	14
Obtener información del mercado	24
Realizar búsquedas previas al comienzo de un proyecto de I+D en la empresa evitando invertir en algo ya existente	65
Divulgar las innovaciones realizadas	9

Tabla 7. Utilización de los servicios de información sobre patentes

3.5. Servicios de información sobre patentes más utilizados

En la actualidad hay diversas fuentes de acceso a la información de patentes, por lo que las empresas pueden diseñar su propia estrategia de búsqueda en base a sus necesidades y diversificar su acción. Las respuestas ponen de manifiesto que las empresas suelen utilizar un número medio de 1,8 fuentes.

De las diferentes opciones posibles, la mayor parte de las organizaciones encuestadas utilizan un agente de la propiedad industrial como principal fuente de información (60%), lo que probablemente refleja también el uso de los servicios de estos agentes para preparar las solicitudes de patentes. Esta tendencia se observa para todas las empresas, a excepción de las más pequeñas (5-10 empleados) que prefieren utilizar los servicios que proporciona la *Oficina Española de Patentes y Marcas* (56%), lo que se explica por los mayores costes que implica utilizar los servicios profesionales de los agentes de la propiedad industrial.

Las siguientes opciones más utilizadas para acceder a la información contenida en las patentes son la *Oficina Española de Patentes y Marcas* (40%) y el acceso directo por la propia empresa desde internet, cd-rom o bases de datos de patentes online. El resto de posibilidades no tiene un uso extendido, tal y como se refleja en la tabla 8, destacando en este aspecto el relativamente escaso papel que desempeñan las agencias regionales de desarrollo, los centros tecnológicos y las asociaciones empresariales en este proceso.

Servicios de información	%
<i>Oficina Española de Patentes y Marcas</i>	40
Agente de la propiedad industrial	60
Directamente por la empresa (internet, cd-rom, etc.)	36
Directamente por la empresa o desde una biblioteca de documentos de patentes	15
Desde una asociación empresarial	3
Desde la empresa matriz	13
Desde agencias de desarrollo regional o centros tecnológicos	8

Tabla 8. Servicios de información sobre patentes utilizados por las empresas

Al realizar el análisis en función de los sectores económicos se pone de manifiesto que los sectores de ingeniería civil-bienes de consumo, electricidad-electrónica, ingeniería de proceso e ingeniería mecánica y maquinaria utilizan en mayor medida a los agentes de propiedad industrial y a la *Oficina Española de Patentes y Marcas* como servicios de información. Por su parte, el sector instrumentación, aunque también utiliza estos servicios, se apoya más en los recursos de acceso directo (cd-rom, internet), mientras que en química-farmacia se utilizan con elevada frecuencia las tres opciones de forma indistinta.

3.6. Proveedores de bases de datos o servicios en internet utilizados

La principal fuente directa de acceso a los servicios de información sobre patentes es mediante internet a la

base de datos *Espacenet*, según un 24% de las respuestas; si bien un 17% de las empresas prefieren el proveedor comercial *Derwent* (*World Patent Index*) y un 13% utiliza los servicios de *IBM* (que han experimentado un gran crecimiento en los últimos años) y la base de datos *Oepmpat*. Por el contrario, el otro gran proveedor comercial europeo de servicios de información sobre patentes *Questel-Orbit* apenas es utilizado por las empresas españolas (4%), al igual que el sistema *Dialog* (7%) (tabla 9).

Proveedor/Servicio	%
<i>Espacenet</i>	24
<i>Oepmpat</i>	13
<i>Cibepat</i>	7
<i>Derwent</i>	17
<i>Questel-Orbit</i>	4
<i>STN</i>	10
<i>Micropatent</i>	3
<i>IBM</i>	13
<i>Dialog</i>	7
<i>Reuter</i>	1

Tabla 9. Bases de datos y servicios en internet utilizados por las empresas

El mismo esquema de utilización presenta el análisis de las empresas en función de su tamaño, si bien se puede resaltar que las grandes prefieren el proveedor *Derwent* (25%) a la base de datos *Espacenet* (23%).

Cuando el análisis se realiza en función de los sectores económicos se pone de relieve que el sector de química-farmacia es el que hace mayor uso del proveedor *Derwent*, mientras que la base de datos *Oepmpat* es requerida principalmente en ingeniería de proceso e ingeniería civil-bienes de consumo. Las compañías del resto de sectores (instrumentación, electricidad-electrónica e ingeniería mecánica y maquinaria) utilizan más la base de datos *Espacenet*.

3.7. Incremento del uso de los documentos de patentes como fuente de información

Una gran parte de las sociedades que tienen patentes (62%) expresa su intención de empezar a utilizar o incrementar el uso de los documentos de patentes como fuente de información en un futuro próximo (vigilancia tecnológica), mientras que el 38% responde negativamente a la misma cuestión, debido principalmente a que ya utilizan sistemas de búsqueda. Son las empresas que tienen más de 50 empleados las que expresan con mayor claridad su deseo de incrementar este uso, lo que pone de manifiesto el interés en apoyarse en la patente

como estrategia específica para impulsar y consolidar el proceso de innovación. Desde la perspectiva de los sectores económicos es el de química-farmacia el que en mayor medida expresa ese interés.

“La innovación tecnológica no resulta fácilmente medible”

3.8. Factores que pueden impulsar a hacer más efectiva la búsqueda de información en las patentes

Esta pregunta es complementaria a la anterior y las respuestas señalan que sería posible fomentar un mayor uso de las patentes como fuente de información tecnológica si se consiguiera una mayor facilidad de acceso directo a las mismas desde internet (48%) y llevaran asociado una mayor cantidad de información (42%). También se reconoce, aunque en menor medida, que a ello contribuiría de forma positiva una mayor accesibilidad a los documentos de patentes (37%). Se observa que hay coincidencia en las respuestas por parte de las empresas que tienen más de 50 empleados, mientras que aquellas de menor tamaño no valoran tanto esta facilidad de acceso, sino los precios.

El análisis por sectores económicos refleja un perfil semejante al global, a excepción de las compañías de ingeniería de proceso que, en lugar de mayor nivel de información, exigen descuentos por mayor uso, lo que es coherente con el nivel de información que se obtiene de las solicitudes de patentes, más enfocado al producto que al proceso (tabla 10).

Factores	%
Precios más flexibles	26
Descuentos por mayor uso	11
Mayor facilidad de acceso	37
Mayor nivel de cualificación en el personal	11
Acceso directo desde internet	49
Mayor nivel de información	42

Tabla 10. Factores que pueden impulsar o hacer más efectiva la búsqueda de información en las patentes

3.9. Importancia de otras formas de propiedad industrial para el avance de la actividad empresarial

Si bien el proyecto se basa en el análisis del papel que desempeñan las patentes como elemento de referencia para identificar patrones de innovación empresarial, resulta interesante conocer si las empresas españolas

que tienen patentes utilizan además otros instrumentos de propiedad industrial.

En relación con este aspecto, las empresas consideran que también hay otras formas de propiedad industrial que son relevantes para incrementar la actividad innovadora y que complementan en buena medida a la patente como instrumento específico. Estas son las marcas (76%) y los modelos de utilidad (56%). Por el contrario, el *copyright* (11%), los secretos industriales (17%) y los dibujos industriales (20%) son considerados importantes por un reducido número de empresas, en su mayor parte por aquellas que tienen menos de 250 empleados.

En función del tamaño, las grandes compañías utilizan principalmente las opciones de los dibujos industriales (25%) y los secretos industriales (30%) como consecuencia de su mayor capacidad innovadora y la fuerte competencia a la que están sometidas. También, resalta el hecho de que sean las más pequeñas (5-10 empleados), las que conceden mayor relevancia al *copyright* como forma eficaz de protección (33%) (tabla 11).

El análisis por sectores económicos refleja la consolidación de la marca y los modelos de utilidad como formas de protección de la propiedad industrial más utilizadas y valoradas. Tan sólo destaca el sector ingeniería de proceso que concede igual importancia relativa al modelo de utilidad y al dibujo industrial. Por su parte, los secretos industriales son más valorados en química-farmacia.

Opciones	Tamaño de la empresa				
	5-10	11-50	51-100	101-250	>250
Marcas	56	88	74	80	75
Modelos de utilidad	33	25	52	60	68
Dibujos industriales	11	0	22	20	25
Secretos industriales	33	6	9	4	30
Copyright	33	0	9	8	16

Tabla 11. Importancia de otras formas de propiedad industrial para las empresas (%)

4. Conclusiones

En las dos últimas décadas se ha puesto de manifiesto que la innovación tecnológica es un proceso complejo que influye de manera determinante en la capacidad de producir nuevos productos y, por tanto, constituye un elemento esencial para comprender el nivel de desarrollo tecnológico de un país. No obstante, la realidad es que no resulta fácilmente medible, aunque está

contrastada su relación con indicadores asociados a los procesos de investigación y desarrollo (I+D) y con la existencia de capacidades tecnológicas.

“Los servicios de información sobre patentes más utilizados son los agentes de la propiedad industrial y la Oficina Española de Patentes y Marcas”

Del conjunto de indicadores disponibles, aquellos que más información aportan sobre los resultados de los procesos de innovación son las patentes, si bien su uso también se enfrenta a un conjunto de limitaciones que son superadas por las ventajas que conllevan. La principal novedad que incorpora este análisis se encuentra en la identificación de los elementos cualitativos que aportan las solicitudes de patentes y que no han sido utilizados en la mayor parte de los análisis realizados en España con base en estos indicadores.

El nivel de utilización de la información descrita por las patentes pone de relieve que se les concede una gran importancia fundamentalmente como instrumento de defensa de la propiedad industrial y, en menor medida, como instrumento que proporciona a la empresa mayor capacidad de negociación con otras compañías. Tan sólo el segmento de grandes empresas considera importantes las patentes como instrumento de ayuda a la exportación de tecnología y a la formalización del conocimiento.

Desde la perspectiva de la frecuencia en el uso de los servicios de información sobre patentes se obtiene que las sociedades de tamaño medio y grande los utilizan de forma mensual, mientras que las pequeñas y micro empresas tienen una frecuencia inferior. La principal finalidad en el uso de los servicios de información sobre patentes es la realización de búsquedas previas al comienzo de un proyecto de I+D en la organización, evitando invertir en algo ya existente, y también la obtención de información acerca de los avances de los competidores.

Entre los servicios de información sobre patentes más utilizados se encuentran los agentes de la propiedad industrial y la *Oepm*. Otras opciones como centros tecnológicos y asociaciones empresariales apenas tienen valoración.

La principal fuente directa de acceso a los servicios de información sobre patentes es la base de datos *Espace-net* accediendo directamente por internet, y también el proveedor comercial *Derwent*. Por último, se pone de manifiesto que sería posible fomentar en mayor me-

dida el uso de las patentes como fuente de información tecnológica si se consiguiera un acceso directo más fácil desde internet.

Nota

Para mayor claridad los porcentajes que figuran en este artículo se han redondeado

Bibliografía

Archibugi, Daniele. "Patenting as an indicator of technological innovation: a review". *Science and public policy*. 1992, December, v. 19, n. 6, pp. 357-368.
http://www.danielearchibugi.org/downloads/papers/Archibugi_Patenting_S&PP_92.pdf

Arundel, Anthony; Kabla, Isabelle. "What percentage of innovations are patented? Empirical estimates for European firms". *Research policy*, 1998, v. 27, n. 2, pp. 127-141.

Basberg, Bjorn. "Patents and the measurement of the technological change: a survey of the literature". *Research policy*. 1987, v. 16, n. 2-4, p. 131-141.

Díaz-Pérez, Maidelyn; De-Moya-Anegón, Félix. "El análisis de patentes como estrategia para la toma de decisiones innovadoras". *El profesional de la información*, 2008, mayo-junio, v. 17, n. 3, pp. 293-302.

Dosi, Giovanni. "Sources, procedures, and microeconomics effects of innovation". *Journal of economic literature*, 1988, September, v. 26, n. 3, pp. 1120-1171.

Escorsa, Pere; Maspons, Ramón. *De la vigilancia tecnológica a la inteligencia competitiva*. Madrid: Prentice Hall, 2001.

Freeman, Chris; Soete, Luc. *The economics of industrial innovation*. London, Frances Pinter, 1982.

Glazier, Stephen. *Patent strategies for business*. USA, Law & Business Institute, 2003.

Griliches, Zvi. "Patent statistics as economic indicators: a survey". *Journal of economic literature*, 1990, December, v. 17, pp. 1661-1707.

Hidalgo, Antonio; Molero, José. "Technology and growth in Spain (1950-1960): an evidence of Schumpeterian pattern of innovation based on patents". *World patent information*, in press, v. 31, n. 3, September 2009, pp. 199-206.

Kline, Stephen; Rosenberg, Nathan. "An overview of innovation". En: Landau, Ralph; Rosenberg, Nathan. *The positive sum strategy*. Washington DC: National Academy Press, 1986, pp. 273-305.

Muñoz-Durán, Javier; Marín-Martínez, María; Vallejo-Triano, José. "La vigilancia tecnológica en la gestión de proyectos de I+D+i: recursos y herramientas". *El profesional de la información*, 2006, noviembre-diciembre, v. 15, n. 6, pp. 411-419.

Narin, Francis; Elliot, Perry, Ross. "Patents as indicators of corporate technological strength". *Research policy*. 1987, v. 16, pp. n. 2-4, pp. 143-155.

Newton, David. *How to find information: patents on the internet*. Great Britain: The British Library, 2000.

OCDE. *Oslo manual. Guidelines for collecting and interpreting innovation data*. Paris: Oecd, Eurostat, European Commission, 2005.

Pavitt, Keith. "Do patents reflect the useful research output of universities?". *Research evaluation*, 1998, August, v. 7, n. 2. pp. 105-112.

Rojo, Raquel; Gómez, Isabel. "Análisis de la producción científica y tecnológica de la industria española en el sector de las tecnologías de la información y de las comunicaciones". *El profesional de la información*, 2006, mayo-junio, v. 15, n. 3, pp. 190-201.

Schmookler, Jacob. *Invention and economic growth*. Cambridge (MA): Harvard University Press, 1966.

Agradecimientos

Los autores agradecen su colaboración en la realización de este análisis a la *Oficina Española de Patentes y Marcas* mediante el proyecto *Análisis estadístico de datos básicos de patentes como indicadores de innovación*.

Antonio Hidalgo-Nuchera, Depto. de Ingeniería de Organización, Administración de Empresas y Estadística, Escuela Técnica Superior de Ingenieros Industriales, Univ. Politécnica de Madrid.
ahidalgo@etsii.upm.es

Santiago Iglesias-Pradas y Ángel Hernández-García, Depto. de Ingeniería de Organización, Administración de Empresas y Estadística, Escuela Técnica Superior de Ingenieros de Telecomunicación, Univ. Politécnica de Madrid.
s.iglesias@upm.es
angel.hernandez@upm.es

Spanish Meeting Point

La revista *El profesional de la información* organiza una nueva edición del *Spanish Meeting Point*, dentro de la *Online Information Conference*, en Londres - la más importante del mundo en materia de contenidos y bases de datos científicas y profesionales, así como en software de gestión de contenidos.

Online Information Conference, Londres 1-3 diciembre 2009
<http://www.online-information.co.uk>



¡¡ Nada más...



... y nada menos !!

Mi Biblioteca

La revista del mundo bibliotecario

Suscríbete a *Mi Biblioteca* y recibirás cada año, de manera gratuita, el *Calendario de la Lectura* y el *Anuario de Bibliotecas Españolas* de la Fundación Alonso Quijano.

Tfno. 952 23 54 05
www.mibiblioteca.org

La base de datos *ISOC* como sistema de información y fuente para el análisis de las ciencias humanas y sociales en España

Por Teresa Abejón-Peña, Angeles Maldonado-Martinez, Luis Rodríguez-Yunta y María-Cruz Rubio-Liniers



Teresa Abejón-Peña es documentalista especializada en el CSIC desde 1989 desempeñando principalmente tareas relacionadas con las bases de datos *ISOC* de humanidades. En el Cindoc ha participado en varios proyectos sobre lenguajes documentales y terminología especializada de las ciencias humanas y sociales. Es la coordinadora de la Unidad de Análisis Documental y Producción de Bases de Datos *ISOC*, que depende del Centro de Ciencias Humanas y Sociales del CSIC.



Ángeles Maldonado-Martínez, documentalista del CSIC desde 1989 como encargada del Área de Urbanismo y Ordenación del Territorio. Experta en el tratamiento y la gestión de documentación de carácter científico. Autora en colaboración de diversos tesauros especializados, entre ellos el "Tesauro de Urbanismo" y el "Tesauro *ISOC* de Topónimos" (edición abreviada). En la actualidad es responsable del Área de Ciencias Sociales de la base de datos *ISOC* del CCHS del CSIC.



Luis Rodríguez-Yunta ha ido perdiendo el pelo, que no el buen humor, en labores de documentalista en el CSIC desde 1986, en los sucesivos centros *ISOC*, Cindoc y ahora CCHS. Trabaja en la base de datos *ISOC* y el portal Americanismo (un proyecto cooperativo de la red Redial). Desde 2007 es además profesor asociado en la Facultad de Documentación de la Universidad Complutense de Madrid. Ha colaborado con frecuencia en actividades de la asociación profesional Sedic.



María-Cruz Rubio-Liniers, responsable del área de humanidades de la Unidad de Análisis Documental y Producción de Bases de Datos *ISOC*, del CSIC, es especialista en información y documentación histórica, evaluación de revistas científicas, lenguajes documentales, tesauros, bases de datos, y estudios de género. Editora de la serie "Bibliografías de historia de España". Participa en el proyecto de investigación (CSIC-Univ. Carlos III) "La historia moderna de España a través de su bibliografía".

Resumen: Se presenta la base de datos *ISOC* como sistema de información de gran utilidad para la difusión y el análisis de la producción científica de las ciencias humanas y sociales en España desde 1975. Se comienza con una exposición sobre la evolución histórica de la base desde sus inicios hasta el momento presente. En segundo lugar se describen los elementos de calidad que la diferencian de otros sistemas documentales. A continuación se argumenta sobre su idoneidad para la realización de análisis bibliométricos sobre producción científica en cualquiera de las disciplinas que componen las ciencias sociales y humanas en España. Finalmente, se aportan datos acerca de su uso por la comunidad científica y académica.

Palabras Clave: Bases de datos bibliográficas, *ISOC* (base de datos), Revistas científicas, Bibliometría, Ciencias humanas, Ciencias sociales.

Title: *ISOC* database as information system: a major source for the analysis of Human and Social Sciences in Spain

Abstract: The *ISOC* database is presented as an information system that is very useful for the dissemination and the study of scientific production in human and social sciences in Spain since 1975. First of all, an historical overview is proposed. Secondly, quality elements that distinguish *ISOC* from other documentary systems are described. Thirdly, its suitability for bibliometric studies on the scientific production of these subjects published in Spain is discussed. Finally, some data about its use by the scientific and scholarly community are presented.

Keywords: Bibliographic databases, *ISOC* (database), Scientific journals, Bibliometrics, Human sciences, Social sciences.

Abejón-Peña, Teresa; Maldonado-Martinez, Angeles; Rodríguez-Yunta, Luis; Rubio-Liniers, María-Cruz. “La base de datos ISOC como sistema de información y fuente para el análisis de las ciencias humanas y sociales en España”. *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 521-528.

DOI: 10.3145/epi.2009.sep.05

1. La base de datos ISOC, ¿cómo hemos llegado hasta aquí?

LA BASE DE DATOS ISOC forma parte de las bases de datos bibliográficas del CSIC. Al igual que ICYT e IME, se trata de productos documentales que nacieron en la primera fase de evolución del mercado de la información en España (García-Moreno, 1999; Rodríguez-Yunta, 2009).

Durante muchos años estos productos fueron recursos únicos, indispensables para la realización de búsquedas de la bibliografía científica editada en publicaciones españolas, así como para la realización de estudios bibliométricos.

Las siglas *ISOC* se corresponden con la denominación que tuvo el centro productor, el *Instituto de Información en Ciencias Humanas y Sociales*, que llevó a cabo su labor en el período 1975-1991. En 1992 el *ISOC* se anexiona al *Instituto de Información en Ciencia y Tecnología (ICYT)*, para constituir el *Centro de Información y Documentación Científica (Cindoc)*, hasta 2007. Actualmente, la *Unidad de Producción de Bases de Datos ISOC* se encuentra integrada en el *Centro de Ciencias Humanas y Sociales (CCHS)* del *CSIC*.

La mayor parte de las bases documentales surgidas en las décadas de 1970 y 1980 se originaron a partir de repertorios bibliográficos impresos. En el caso de *ISOC*, entre 1978 y 1997 se elaboraron dos publicaciones periódicas: el *Índice español de ciencias sociales (IECS)* y el *Índice español de humanidades (IEH)*. La primera comercialización en línea de las bases de datos del *CSIC* se ofreció desde el *Centro*

de Proceso de Datos del *Ministerio de Educación y Ciencia* a partir de 1985. El producto enriquecía la edición impresa con la incorporación de datos sobre el lugar de trabajo de los autores, la identificación del idioma, la traducción de títulos, la apuesta por el control del vocabulario en la indización y la utilización de clasificaciones diseñadas a medida.

En enero de 1991 se inicia la comercialización de las bases de datos en línea por el propio *CSIC*, desde un servidor gestionado por el *Centro Técnico de Informática* y con la creación de un servicio específico de distribución (**Álvaro-Bermejo**, 1991). Este paso fue la culminación de un proceso iniciado en 1988 a partir del cambio de software y la introducción de nuevas mejoras en la base *ISOC*: la grabación de los resúmenes de autor o la división de los campos de indización en descriptores, identificadores y topónimos, a fin de facilitar la búsqueda mediante tesauros.

“En 1985 comienza la comercialización de las bases de datos del CSIC por el Ministerio de Educación y Ciencia. En enero de 1991 se inicia esta comercialización por el propio CSIC”

Sin embargo, durante la década de los noventa la principal difusión de las bases documentales se realizó mediante la edición en cd-rom. El disco de las bases de datos del

CSIC realizado por la empresa *Micronet*, salió al mercado en 1990 y reunía el catálogo colectivo de la *Red de Bibliotecas del CSIC* junto con las bases de datos de artículos de revistas científicas, *ICYT*, *ISOC* e *IME*. Este formato fue valorado como “una de las bases de datos más representativas e importantes en el panorama de la industria de las bases de datos en España” (**Fernández-Marcial**, 2001, p. 322).

La versión en línea de *ISOC* se llevó a cabo a partir de 1991 como un sistema multibase ya que incluye ficheros especializados por disciplinas. Con ello, los productores han querido resaltar la utilidad de poder realizar búsquedas más precisas, acotadas a un ámbito disciplinar. Sin embargo, el conocimiento por los usuarios de estos productos especializados ha sido escaso, en gran parte debido a que el cd-rom de *Micronet* no incluía esta opción de división por subficheros.

En 1997, para adaptarse a los cambios producidos por la expansión de internet, el *CSIC* asume la distribución mediante una interfaz web con una doble versión, gratuita para un contenido básico y comercial para su versión completa. Entre 2002 y 2004 se renueva esta interfaz, coincidiendo con una nueva etapa en la apuesta por una política de calidad más estricta en la selección de las fuentes (**Alcain; Urdín**, 2002; **Urdín; Alcain**, 2004). El nuevo diseño presenta mejoras importantes: combinación de resultados del historial de búsqueda, enlaces SFX, posibilidad de filtrar los resultados para seleccionar sólo los registros con enlaces al texto completo, opción de búsqueda previa conjunta en las tres bases de datos, relación entre las

referencias y los datos del directorio de revistas y nuevas opciones de exportación.

2. Elementos de calidad de la base de datos ISOC

La base de datos *ISOC* es el principal sistema analítico de información científica en ciencias humanas y sociales disponible en España. Por una parte cumple los principales requisitos de calidad referentes a la cobertura temática y la selección de documentos, y por otra ofrece un conjunto de valores añadidos que van desde el análisis de contenido hasta el diseño y las posibilidades de recuperación de su interfaz de consulta.

2.1. Cobertura, selección y evaluación de revistas

La gestión de los productos documentales debe tener muy en cuenta la calidad de las publicaciones incorporadas (Abejón-Peña, 2009).

A partir de 2006 para la base de datos *ISOC* se aplica un sistema riguroso de selección y evaluación de las revistas, que remarca su carácter de sistema de información científica. Esta evaluación la realiza el “Grupo de investigación de evaluación de publicaciones científicas” del *Iedcyt (CSIC)*, que a partir de 1999 utiliza como base de su sistema los criterios de calidad *Latindex*, la presencia de las publicaciones en bases de datos internacionales y la evaluación de los artículos por parte de “pares” (Giménez-Toledo; Román-Román; Alcain-Partearroyo, 2007).

“La base de datos *ISOC* es el principal sistema analítico de información científica en ciencias humanas y sociales disponible en España”

Áreas	Revistas A	Revistas B	Revistas C	A+B+C
Ciencias humanas	34	116	350	500
Ciencias sociales	27	84	380	491
Multidisciplinares	4	3	56	63
Total base <i>ISOC</i>	65	203	786	1.054

Tabla 1. Datos de la categorización de revistas en la base de datos *ISOC*

Como resultado de dicha evaluación las revistas españolas de ciencias humanas y sociales se han categorizado como A, B o C en función de los parámetros de valoración obtenidos. Para que una revista sea vaciada en la base *ISOC* tiene que alcanzar como mínimo la categoría C, que exige cumplir con 13 requisitos básicos de calidad. En la tabla 1 puede verse la situación de la categorización de revistas en la base *ISOC* en enero de 2009.

El total de revistas con presencia en el directorio *ISOC* es de 2.440 títulos, de los cuales 783 son de revistas “muertas” y 1.658 se siguen publicando; de estas últimas, 604 han dejado de analizarse en la base de datos por no cumplir los requisitos mínimos de evaluación.

El *Directorio de Revistas de Ciencias Sociales y Humanidades* ofrece información sobre estos parámetros de evaluación, con datos actualizados en 2008 de 1.658 revistas vivas. Esta revisión permite ofrecer una información fiable para los usuarios sobre la categorización, el soporte y la accesibilidad de las publicaciones en internet. El sistema de selección y evaluación de revistas es un valor añadido e importante de la base de datos *ISOC*, y la convierte en una herramienta de apoyo a la política científica y en un instrumento de diagnóstico de la producción española en ciencias humanas y sociales.

2.2. Profundidad en la descripción de los documentos

El segundo aspecto de calidad que cabe destacar es el enriquecimiento de la estructura de campos

de contenido, que hace a la base *ISOC* única entre las bases de datos nacionales. Este sistema de análisis es una apuesta por una recuperación más eficaz, ya que hace posible la obtención de resultados más concretos y ajustados a la especialidad científica, garantizando un equilibrio entre exhaustividad y pertinencia.

Clasificación especializada

Las clasificaciones aplicadas son de elaboración propia, de acuerdo con las peculiaridades metodológicas y taxonómicas de cada disciplina. El indizador clasifica cada artículo, con independencia de la revista donde aparezca, de modo que el documento queda encuadrado en la sub-base más pertinente.

Las clasificaciones especializadas utilizadas en *ISOC* facilitan una búsqueda y recuperación mucho más pertinente y ayudan en bibliometría a la elaboración de “mapas” disciplinares, así como a análisis temáticos más ajustados a la realidad conceptual de las investigaciones.

Análisis de contenido en *ISOC*: diversidad de campos para la indización

ISOC ha apostado por el análisis de contenido de los artículos y el control del lenguaje de indización. Para cada documento incorporado, se analiza y describe su contenido mediante conceptos basados en la significación y el contexto.

Se ha tratado de mantener un vocabulario lo suficientemente específico y cuidadoso con la terminología especializada de cada una de las áreas.

Los campos para la indización de contenido que ofrece *ISOC* son:

- Descriptores o conceptos representativos del tema del trabajo.
- Identificadores: nombres propios, instituciones con sede, títulos de obras, etc.
- Topónimos: nombres geográficos de lugares, tanto físicos como administrativos.
- Período histórico, décadas y siglos: datos numéricos que se utilizan para analizar y recuperar los artículos de tema histórico.
- Legislación y jurisprudencia: leyes y sentencias, objeto de estudio en los trabajos jurídicos.

La consulta de los índices y campos correspondientes, permite elaborar estrategias de recuperación más ajustadas y precisas.

“ISOC ha apostado decididamente por el análisis de contenido de los artículos y el control del lenguaje de indización”

Los resúmenes en la base de datos *ISOC*

Otro elemento destacado de calidad para la recuperación en bases de datos es la incorporación de resúmenes de autor. Es interesante destacar su considerable aumento desde la creación de la base *ISOC* hasta la actualidad, como consecuencia de la mejora en la normalización y presentación de artículos en las revistas científicas españolas. En los documentos publicados de 1975 a 1989 en la base hay 7.585 resúmenes sobre un total de 171.478 referencias (1,4%); entre 1990-1999 se contabilizan 63.058 resúmenes sobre 223.678 referencias (28,2%), y desde 2000 a 2008 se incorpora un total de 91.671 de 164.931 referencias (55,6%), lo que representa un

espectacular crecimiento a lo largo de los años. Un aumento todavía mayor se percibe en los artículos de revistas publicadas en 2009 e incorporadas a la base hasta junio de este año, de los cuales el 80,7% tienen resumen de autor.

Otros campos de valor añadido

ISOC incorpora otros datos que aportan una mayor versatilidad a la recuperación de información y al análisis de la producción científica:

- Idioma de los documentos.
- Lugar de trabajo de los autores, siempre que figure en la fuente.
- Denominación del congreso, cuando se trata de una comunicación o ponencia, aun cuando haya sido publicado en una revista.

“Un elemento destacado de calidad para la recuperación en bases de datos es la incorporación de resúmenes de autor”

2.3. Acceso al texto completo de los artículos

La base de datos *ISOC* incluye el campo “url”, en el que se reco-

ge la dirección del sitio web correspondiente al texto completo de cada artículo. Así, siempre que las revistas difundan sus contenidos en acceso abierto, es posible consultar el texto original desde los registros de *ISOC*.

El 52% de las revistas que se incluyen en la base de datos son electrónicas o tienen una versión en internet, pero no todas aplican la misma política de acceso a sus contenidos. En la tabla 2 puede apreciarse un aumento continuado del porcentaje de enlaces url de los últimos años. El ligero descenso en el año en curso se relaciona con la frecuente práctica editorial de embargar los últimos números publicados.

2.4. Amplias posibilidades de recuperación de información

ISOC se caracteriza por su versatilidad a la hora de recuperar información. Se puede realizar una búsqueda conjunta en toda la base de datos o dirigirla exclusivamente a una de las sub-bases. En cualquiera de los dos casos se ofrecen cuatro opciones de búsqueda: simple, por campos, por índices y por comandos. La búsqueda simple es la más sencilla, pero también la más imprecisa. La búsqueda por campos permite combinar diversos criterios y cruzarlos con una clasi-

Año publicación	Nº enlaces	Total documentos	Enlaces (%)
2000	1.873	23.654	8
2001	1.970	22.426	9
2002	2.494	22.365	11
2003	2.984	21.306	14
2004	3.235	20.971	15
2005	3.629	19.030	19
2006	4.585	16.992	27
2007	4.247	13.318	32
2008	3.790	10.302	37
2009	286	950	30

Tabla 2. Porcentaje de enlaces al texto completo en los registros de la base de datos *ISOC* por año de publicación

ficación temática y/o con un año o intervalo temporal. La búsqueda por índices hace posible la visualización en forma de listado alfabético del contenido de casi la totalidad de los campos que componen el registro. Esta es una opción útil para conocer el vocabulario utilizado en el análisis de los documentos y mejorar la precisión en la recuperación. La modalidad de búsqueda por comandos es recomendable para la realización de las búsquedas más complejas, generalmente ejecutadas por los profesionales de la documentación.

También es interesante resaltar que gracias al hipertexto se pueden realizar nuevas búsquedas a partir de la visualización de un registro. El hipertexto funciona en los campos de autor, descriptores, identificadores y topónimos. Para el campo de revista enlaza con su registro correspondiente en el directorio de publicaciones.

Cualquiera que haya sido la búsqueda elegida y una vez que aparece en pantalla el listado de los documentos que cumplen los criterios explicitados, se puede activar un filtro sobre los mismos con el fin de visualizar exclusivamente las referencias con enlace al texto completo.

El resultado de una búsqueda puede ser descargado en txt con o sin etiquetas de campo, pdf, xml o formatos de exportación específicos para *Reference Manager* y *RefWorks*. La base de datos *ISOC* puede ser integrada en un gestor de recursos electrónicos y participar de los beneficios de las búsquedas conjuntas en diferentes aplicaciones que posibilita este tipo de herramientas informáticas. Además, también está adaptada a la norma *NISO Z39.88-2004*, lo que le permite funcionar con un servidor dinámico de enlaces. Gracias a ello, los suscriptores pueden conectar fácilmente la consulta de las bases de datos del *CSIC* con su catálogo.

3. *ISOC*, recurso para la realización de estudios bibliométricos sobre producción científica

ISOC es una base de datos que recoge la producción científica española en ciencias humanas y sociales en los últimos 35 años contenida en revistas de calidad editadas en nuestro país. Es una herramienta imprescindible para la realización de estudios bibliométricos en estas áreas del saber. Los motivos fundamentales que la hacen imprescindible son básicamente dos: la exhaustividad, dada su vocación de reunir todas las revistas científicas de calidad editadas en España, y la multidisciplinariedad, ya que todas las disciplinas académicas que componen las ciencias humanas y sociales están representadas en la base.

ISOC cuenta a 10 de junio de 2009 con un total de 565.698 documentos, de los que 537.192 son artículos de revistas, 19.375 comunicaciones y ponencias de actas de congresos y 9.212 se corresponden con otros tipos documentales.

“*ISOC* recoge la producción científica en ciencias humanas y sociales en los últimos 35 años contenida en revistas de calidad editadas en España”

La cobertura temporal se inicia en 1975, aunque hay en la base 3.899 documentos anteriores a esta fecha. A partir de este año el número total aumenta progresivamente hasta 1993, donde se alcanza el máximo anual con 24.494 (figura 1). Desde este año hasta 2005 la cifra oscila entre los 20.000 y los 24.000 documentos. A partir de 2005 la cifra se sitúa por debajo de los 20.000. La tendencia descendente que se observa desde 2005 puede deberse a la suma de dos circunstancias: por una parte a la reducción del número de títulos de revistas vaciadas, como consecuencia del proceso de evaluación; y por otra parte al retraso en la edición de algunas revistas, sobre todo en el campo de las ciencias humanas. Las cifras proporcionadas en este momento probablemente aumentarán en el futuro, dado que se recuperarán documentos con fecha de publicación anterior al año en curso, no incorporados en su momento por diversos motivos, y que las nuevas revistas que son dadas de alta en *ISOC* son vaciadas con carácter retrospectivo siempre que es posible.

Las referencias hacen posible el estudio de la producción de sus autores, pero además el campo “lugar de trabajo” sirve de fuente para la realización de análisis bibliométricos sobre la producción de instituciones, comunidades autónomas y países. En total un 45,8% de los documentos presentes en

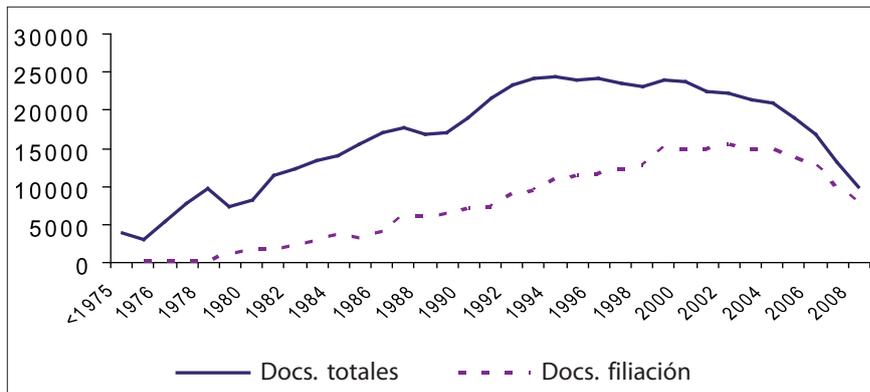


Figura 1. Distribución temporal de los documentos incorporados a la base de datos *ISOC* y existencia de datos de filiación institucional

ISOC aporta el lugar de trabajo de los autores, pero se observa un incremento positivo sostenido con el paso del tiempo, ya que entre 1975 y 1989 lo tienen un 22,4% de los trabajos, elevándose el porcentaje hasta un 46,2% en la década de los 90, para pasar a un 70,3% en el período 2000-2008 (ver figura 1). En los registros de documentos publicados en 2009 este porcentaje sube ya al 81%¹. Mediante este campo se puede estudiar, además, la colaboración científica en España, tanto entre autores, como entre instituciones o áreas geográficas.

La multidisciplinariedad de ISOC se hace patente con la presencia de revistas procedentes de quince disciplinas académicas de gran arraigo en la universidad española y dos más transversales (estudios sobre América Latina y urbanismo). En la figura 2 se pueden ver las disciplinas presentes y el número de revistas vivas de cada una de ellas (suma de las de categorías A, B y C).

En cuanto a la distribución de documentos por disciplinas científicas (ver figura 3) las que más volumen alcanzan son economía (82.335 documentos) e historia (80.513), a continuación se sitúa derecho (63.222), seguida de literatura (52.442) y educación (50.381).

La existencia del campo “clasificación temática” posibilita la realización de análisis bibliométricos dirigidos a disciplinas completas o a partes de ellas. También permite estudiar la relación entre las propias disciplinas y la detección de temas fronterizos entre unas y otras.

4. ISOC, cada vez más consultada desde internet

4.1. Política de distribución y características de los usuarios

En 2004 se establecieron las condiciones actuales del acceso en línea a las tres bases de datos bibliográficas del CSIC. En la ac-

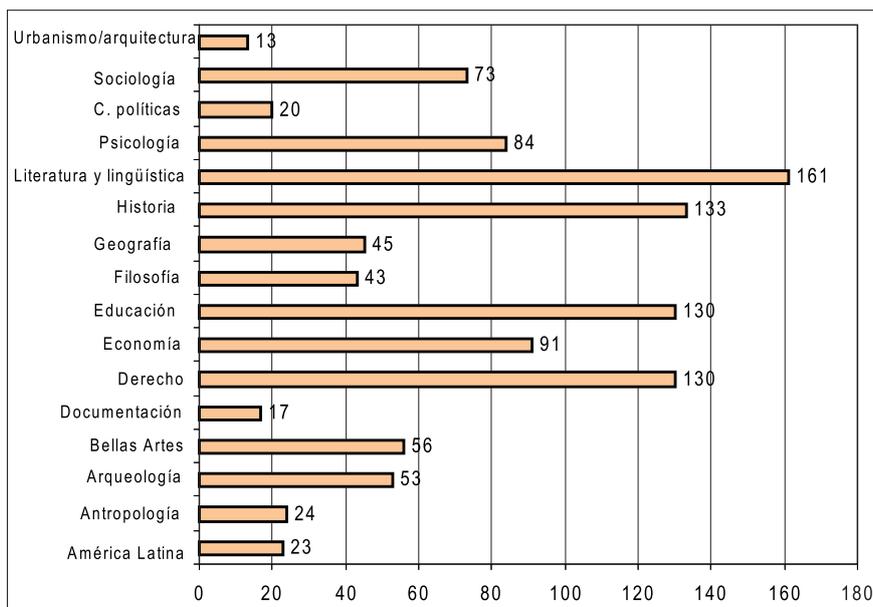


Figura 2. Distribución por disciplinas de las revistas vaciadas en ISOC

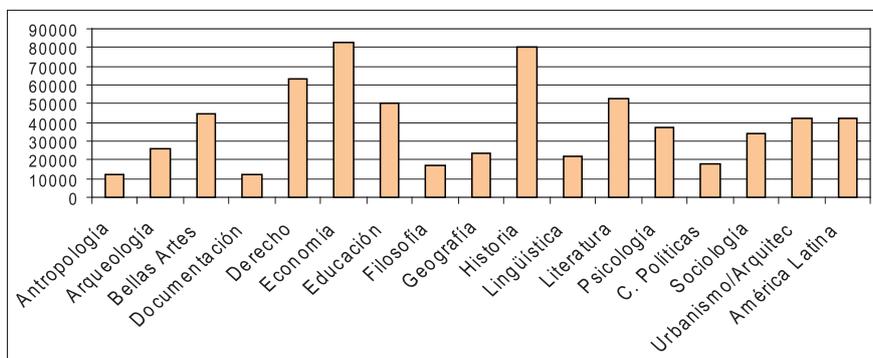


Figura 3. Distribución por disciplinas de los documentos incorporados en ISOC

tualidad, se ofrecen al usuario dos opciones: acceso para suscriptores y acceso gratuito. En la primera opción, se accede a tres productos principales que se pueden consultar conjuntamente o por separado: ISOC-Ciencias sociales y humanidades, ICYT-Ciencia y tecnología e IME-Biomedicina. La suscripción garantiza la búsqueda en todos los campos y la visualización del re-

gistro completo. El acceso gratuito facilita la consulta de los Sumarios ISOC-Ciencias sociales y humanidades, Sumarios ICYT-Ciencia y tecnología, Sumarios IME-Biomedicina, la base de datos ISOC-Biblioteconomía y documentación y los tres directorios de revistas españolas.

Aunque al usuario no se le ofrece la posibilidad de contratar exclusivamente la consulta de la base de datos ISOC, los registros de las sesiones abiertas por los suscriptores proporcionan datos seguros sobre la utilización de esta base de datos, que recibe un 60% del total de consultas a las bases de datos del CSIC².

Actualmente hay 154 entidades públicas y privadas suscritas al acceso en línea. Se trata principal-

“La multidisciplinariedad de ISOC se hace patente con la presencia de revistas procedentes de quince disciplinas académicas y dos más transversales”

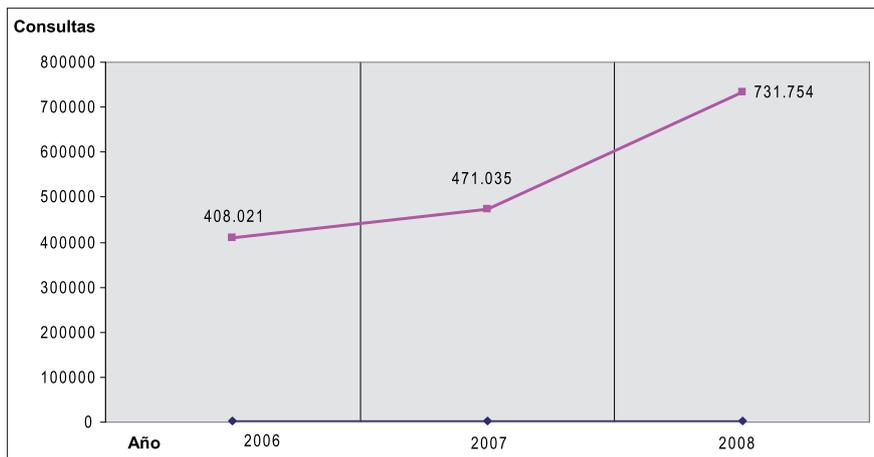


Figura 4. Uso de la base de datos ISOC

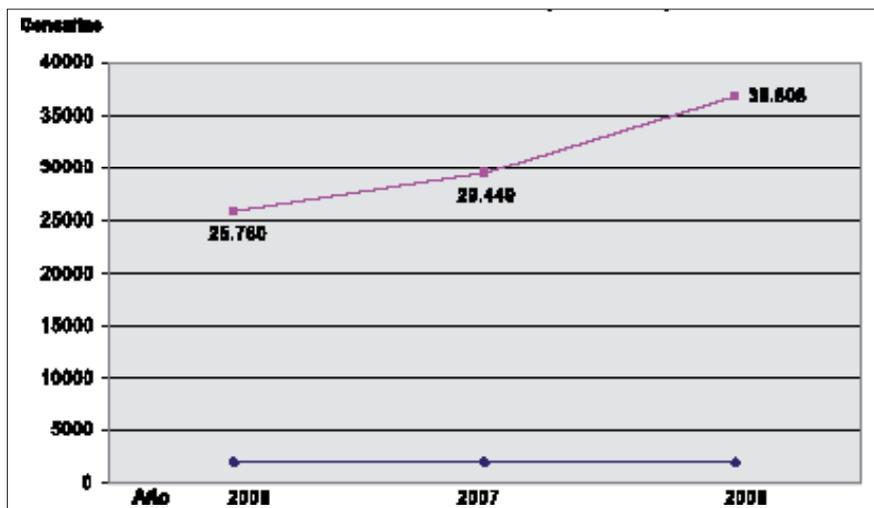


Figura 5. Uso del directorio de revistas de la base de datos ISOC

ISOC y 12.869 en el directorio de revistas), esta tendencia ascendente se confirmaría también durante el año en curso.

La figura 6 muestra la distribución de las consultas por tipos de usuarios⁴.

Estos datos de uso ilustran el alto nivel de implantación que tiene la base de datos ISOC entre la comunidad científica española. El 96% de las universidades españolas está suscrito al acceso en línea, y es el tipo de usuario que más utiliza este servicio de información.

En cuanto al acceso gratuito desde internet, el cómputo de los “logs” de 2008 indica que el volumen de consultas triplica el del acceso por suscripción. En 2009 se ha implantado la herramienta *Google Analytics* para obtener más información sobre la utilización de las dos modalidades de acceso en línea.

El 76% de las consultas a las bases de datos del CSIC provienen de Andalucía, Comunidad de Madrid y Cataluña. El hecho de que el 94% de las mismas⁵ se realicen desde España indica que se deberían difundir mucho más en América Latina y entre los departamentos de hispanismo de las universidades de América del Norte y Europa. Hay dos vías que fomentarían su utilización fuera de España: incluir la base de datos en la oferta de los distribuidores comerciales internacionales y/o incorporar la opción de consulta en inglés.

mente de universidades, organismos públicos de investigación, servicios especializados de la administración autonómica, empresas y otras entidades privadas. La mayor parte de estas entidades tienen su sede en España, pero también utilizan ISOC algunas universidades europeas y americanas³. Los usuarios de los centros y bibliotecas del CSIC utilizan ISOC gratuitamente con los mismos derechos de acceso que los suscriptores.

4.2. Estadísticas de uso

Los datos de uso se obtienen contabilizando los “logs” o sesiones que quedan registrados en el servidor que soporta las conexiones a las bases de datos del CSIC. No se contabilizan las visitas a la web que no llegan a ser una consulta efectiva mediante los formularios de búsqueda.

queda. Las figuras 4 y 5 muestran la evolución de las consultas a ISOC y a su directorio de revistas en los últimos tres años.

En este período se observa un continuo crecimiento en la utilización de la base de datos ISOC. Si atendemos a los datos del primer trimestre de 2009 (231.899 consultas registradas en la base de datos

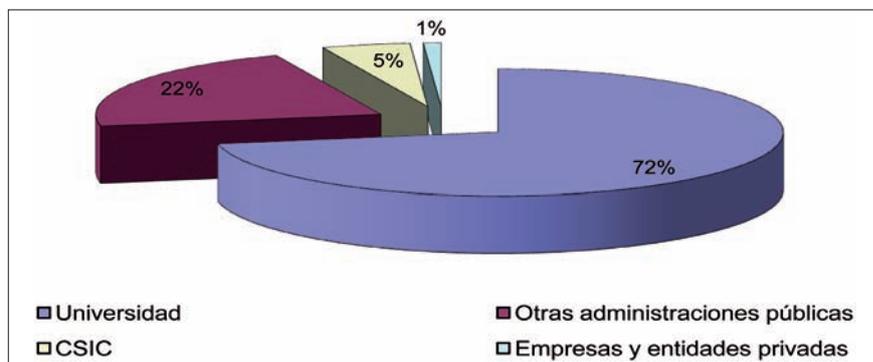


Figura 6. Distribución de las consultas por tipo de usuarios

5. Conclusiones

La base de datos *ISOC* es una herramienta de gran utilidad para la difusión y el estudio de las ciencias humanas y sociales en España, dadas sus características y peculiaridades.

“Hay que fomentar su uso fuera de España: incluyendo la base de datos en la oferta de los distribuidores internacionales y/o incorporando la opción de consulta en inglés”

Durante muchos años la base de datos *ISOC* ha representado el único medio para realizar búsquedas bibliográficas de carácter multidisciplinar en publicaciones periódicas editadas en España. En la actualidad, aunque como sistema de información tiene la competencia de otros recursos que han ido surgiendo con el paso del tiempo, como son las bases de datos de sumarios y las plataformas y agregadores de revistas electrónicas, sigue siendo una herramienta fundamental de obligada consulta dadas las ventajas competitivas que presenta frente a todos ellos. Dichas ventajas son las siguientes:

- Extensa cobertura temporal, ya que tiene documentos desde 1975 hasta la actualidad.
- Multidisciplinariedad, dado que están representadas todas las disciplinas de las ciencias humanas y sociales.
- Selección de revistas de calidad.
- Presencia de revistas con soportes y modos de acceso diversos: revistas impresas, revistas electrónicas gratuitas y comerciales.

– Análisis de contenido y control del vocabulario.

– Consignación de la afiliación institucional de los autores.

– Diversidad en las opciones de búsqueda: simple, por campos, por índices y por comandos.

– Alto grado de implantación entre la comunidad académica española, dado el número de consultas y su continuo aumento.

Por otra parte, al limitar su cobertura a la edición realizada en España, *ISOC* es el producto ideal para poder valorar la producción nacional específica en las diferentes disciplinas de ciencias humanas y sociales.

En este trabajo hemos expuesto la larga trayectoria de la base de datos *ISOC* y su constante transformación en consonancia con los avances de los sistemas de información. Esperamos que esta evolución continúe con la incorporación de nuevos valores añadidos, la optimización general de la consulta y una mejor actualización. En el futuro deben introducirse cuantas mejoras sean necesarias para que siga siendo un buen sistema de información bibliográfica. La base de datos *ISOC* tiene elementos únicos para convertirse en una fuente de obtención de datos estables y fiables sobre la calidad de las publicaciones que recoge. Una futura ampliación de los tipos documentales en función de los resultados de investigación más representativos de las ciencias humanas y sociales, especialmente las monografías, junto con la incorporación directa de citas en *ISOC*, harían de esta base de datos el recurso principal para evaluar con datos objetivos la producción científica de estas áreas publicada en España.

6. Notas

1. Dato calculado a 10 de junio de 2009.
2. Datos de consultas del período 2006-2008.
3. De estas universidades extranjeras proviene un 4% de las consultas a *ISOC*.
4. Los datos corresponden a las consultas efec-

tuadas entre el 1 enero de 2008 y el 31 de marzo de 2009.

5. Datos correspondientes a las tres bases de datos del *CSIC* en los primeros meses de 2009.

7. Bibliografía citada

Abejón-Peña, Teresa. “Criterios de selección de revistas para la producción de bases de datos: la evaluación previa como garantía de calidad”. En: Rodríguez-Yunta, Luis; Giménez-Toledo, Elea (coords.). *La documentación como servicio público. Estudios en homenaje a Adelaida Román*. Madrid: CSIC, 2009, pp. 199-226.

Alcain, María-Dolores; Urdín, Carmen. “Las bases de datos del *Cindoc* inician una nueva etapa”. *El profesional de la información*, 2002, v. 11, n. 4, pp. 275-277.
<http://www.elprofesionaldelainformacion.com/contenidos/2002/julio/6.pdf>

Álvaro-Bermejo, Concepción. “Las bases de datos documentales del *CSIC*”. *Cuadernos de documentación de cajas de ahorro*, 1991, v. 6, n. 11, pp. 31-36.

Fernández-Marcial, Viviana. *Diagnóstico y mejora de la producción de bases bibliográficas desde la perspectiva de la gestión total de la calidad: Estudio de la base de datos ISOC en cd-rom*. Tesis doctoral leída en la Universidad Carlos III de Madrid, Getafe, 2001.

García-Moreno, María-Antonia. *De la teledocumentación a internet. La industria española de las bases de datos*. Madrid: Editorial Fragua, 1999.

Giménez-Toledo, Elea; Román-Román, Adelaida; Alcain, María-Dolores. “De la experimentación a la coordinación en la evaluación de revistas científicas españolas de ciencias sociales y humanidades”. *Research evaluation*, 2007, v. 16, n. 2, pp. 137-148.

Rodríguez-Yunta, Luis. “Las bases de datos documentales del *CSIC* en el desarrollo histórico del mercado de la información en España (desde sus antecedentes hasta 2008)”. En: Rodríguez-Yunta, Luis; Giménez-Toledo, Elea (coords.). *La documentación como servicio público. Estudios en homenaje a Adelaida Román*. Madrid: CSIC, 2009, pp. 131-172.

Urdín, Carmen; Alcain, María-Dolores. “Bases de datos del *CSIC* en la red”. *Hipertext.net*, 2004, n. 2.
<http://www.hipertext.net/web/pag217.htm>

Teresa Abejón-Peña, Angeles Maldonado-Martinez, Luis Rodríguez Yunta, María-Cruz Rubio-Liniens. *Centro de Ciencias Humanas y Sociales (CCHS). Consejo Superior de Investigaciones Científicas (CSIC). c/ Albasanz, 26-28. Desp. 3A10, 28037 Madrid, España.*
teresa.abejon@cchs.csic.es
angeles.maldonado@cchs.csic.es
luis.ryunta@cchs.csic.es
mariacruz.rubio@cchs.csic.es

Retrieval of very large numbers of items in the *Web of Science*: an exercise to develop accurate search strategies

By Ricardo Arencibia-Jorge, Loet Leydesdorff, Zaida Chinchilla-Rodríguez, Ronald Rousseau and Soren W. Paris



Ricardo Arencibia-Jorge, máster en bibliotecología y ciencia de la información, es jefe del Departamento de Información Científica del Centro Nacional de Investigaciones Científicas (CNIC) de La Habana, Cuba. Desarrolla su investigación doctoral en el análisis cuantitativo de la actividad científica cubana. Coordinador del proyecto Red de Estudios Cuantitativos sobre la Educación Superior cubana (Redec). Editor para Cuba del repositorio de información E-LIS y premio de la Asociación Cubana de Bibliotecarios en 2009.



Loet Leydesdorff, doctor en sociología y profesor titular del Departamento de Dinámica de la Ciencia y la Tecnología en la Universidad de Amsterdam, Holanda, posee una vasta experiencia en diferentes campos como la sociología de la ciencia, el análisis de redes sociales y la sociología de la innovación. En el 2003 recibió el premio Derek de Solla Price por su contribución a la cuantimetría. Autor de múltiples libros y artículos que abarcan los dominios de la ciencia de la información y la comunicación.



Zaida Chinchilla-Rodríguez es doctora en documentación e información científica y científica titular del Consejo Superior de Investigaciones Científicas (CSIC), en el Instituto de Políticas y Bienes Públicos (IPP) de Madrid, España. Miembro del Grupo de Investigación SCImago, desarrolla su investigación en el análisis cuantitativo de dominios del conocimiento, la representación y visualización de información y redes de colaboración científica, y propuestas metodológicas para el diseño de sistemas de información.



Ronald Rousseau, doctor en bibliotecología y ciencia de la información, profesor de la Escuela Católica de Educación Superior de Brujas-Ostende (KHBO), e investigador asociado de la Univ. Católica de Leuven, Bélgica. Presidente de la Sociedad Internacional de Cuantimetría e Informetría (ISSI). Investiga sobre estudios métricos de la información y evaluación de la investigación. Ha recibido importantes galardones como el premio de la Academia de Ciencias de Bélgica en 1979, y el Derek J. de Solla Price en 2001.

Abstract: *The Web of Science interface counts at most 100,000 retrieved items from a single query. If the query results in a dataset containing more than 100,000 items the number of retrieved items is indicated as >100,000. The problem studied here is how to find the exact number of items in a query that leads to more than 100,000 items. One way to achieve this objective is presented. The retrieval of the entire scientific production from the United States in a specific year (2007) is counted and an advanced search strategy is designed. Different sections of items can be retrieved using the Source field of the database. A Boolean statement was created with the aim of eliminating overlapped sections and improving the accuracy of this search strategy.*

Keywords: *Information retrieval, Search strategies, Databases, Web of Science, Scientific production, USA.*

Título: **Recuperación de grandes cantidades de registros en la *Web of Science*: un ejercicio para realizar estrategias de búsqueda precisas**

Resumen: *La interfaz de la Web of Science permite recuperar como máximo 100.000 registros en una búsqueda simple. Si el resultado de la búsqueda tiene más de 100.000 registros, el número de registros recuperados se indica como >100.000. En este artículo se presenta una forma de encontrar el número total de registros en una búsqueda que supera los 100.000 registros. Concretamente, se contabiliza la producción científica total de los Estados Unidos en un año específico (2007). Se diseña una estrategia de búsqueda avanzada para recuperar conjuntos diferentes de registros usando el campo Source*



Soren W. Paris, licenciado en lengua inglesa por la Universidad de West Chester, Estados Unidos, en 2001. Actualmente cursa la Maestría en Gestión y Conservación de Recursos de la Universidad de Antioch, en New Hampshire. Desde el 2002, ha sido investigador asistente del Dr. Eugene Garfield en el área de tecnologías de información. Ambos han trabajado en el desarrollo del programa HistCite, una herramienta para el análisis bibliométrico y la visualización de los resultados de las búsquedas de información en el Web of Science.

de la base de datos. Se crea una instrucción booleana con el fin de eliminar los solapos y mejorar la precisión de la estrategia de búsqueda.

Palabras clave: Recuperación de información, Estrategias de búsqueda, Bases de datos, Web of Science, Producción científica, USA.

Arencibia-Jorge, Ricardo; Leydesdorff, Loet; Chinchilla-Rodríguez, Zaida; Rousseau, Ronald; Paris, Soren W. "Retrieval of very large numbers of items in the Web of Science: an exercise to develop accurate search strategies". *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 529-533.

DOI: 10.3145/epi.2009.sep.06

1. Introduction

SOMETIMES INFORMATION PROFESSIONALS FACE SINGULAR PROBLEMS related to the use of information technology and the management of digital environments. Changes and improvements offered by online providers present users with new tools and different interfaces, requiring continual re-learning (Martínez, 2008).

Often an apparently simple and easy activity requires the practical knowledge of specialists. As many retrieval tasks are team work, each member of the team must clearly communicate objectives, solutions and experiences to the rest. Such working habits lead to a global increase in knowledge and skills.

"WoS data related to geopolitical domains with large numbers of items must be searched using a combination of search statements"

The problem presented here came up during a work session of the *SCImago* research team. Specialists from the Spanish group were doing a scientometric study of the world scientific production in 2007 using *Scopus* and *Web of Science* (WoS) interfaces, when they noticed an inconvenience which at first sight appeared easy to resolve.

Researchers needed an accurate number of papers produced in the USA and the United Kingdom, but a precise number over 100,000 items using the WoS interface was not available.

Recently, one of the multiple papers of **Péter Jacsó** on search strategies and techniques in the most widely used citation-enhanced databases called attention to this topic (**Jacsó, 2009**). WoS data related to geopolitical domains with large numbers of items must be searched using a combination of search statements. The clearest examples were countries such as the USA or the United Kingdom, or blocks of countries such as the European Union, with a scientific production in mainstream journals of over 100,000 articles during a year.

The identification of items from the United Kingdom does not present major difficulties. The construction of two statements including and excluding the word "London" in the *Author Address* field can easily solve the problem. For example, using all databases comprising the WoS (*SCI-Expanded*, *SSCI*, *A&HCI*, *CPCI-S*, *CPCI-SSH*, *IC*, *CCR-Expanded*), and selecting all years in *Timespan*, a user can obtain the total output of this nation through the sum of the items retrieved by the following search statements:

1. PY=2007ANDCU=(England OR Scotland OR Wales OR North Ireland) AND AD=LONDON

2. PY=2007ANDCU=(England OR Scotland OR Wales OR North Ireland) NOT AD=LONDON

"To restrict the search to the year an article is published, it is necessary to use the *Publication Year* (PY) field"

Note that using the *Timespan Limits* the user is in fact restricting the search to the year the data were entered into the database. Therefore, to restrict the search to the year an article is published, it is necessary to use the *Publication Year* (PY) field.

As of June 18, 2009 (the date of this query), there were 33,043 articles in 2007 signed by authors from at least one London scientific or scholarly institution, and there were 98,802 in which there was no author from this English city. A total of 131,845 articles compose the sum total output of the United Kingdom in the WoS that year.

But, what about the USA? The scientific production from this country in a year far exceeds 100,000 articles. How to obtain the total output of the USA using the WoS interface? That question gave rise to a practical and interesting exercise, which required the united efforts of various specialists from different research groups.

2. In search of a solution

At first, a series of search strategies developed by the *SCImago* group was oriented towards the identification of the states of the Union in the *Author Address* (AD) field, with the aim of obtaining different sections of fewer than 100,000 items. But the design of this kind of advanced search strategies, based on the AD field, became very complex in this case. The extensive collaboration between institutions from different states made it difficult to construct a logical operation in the search strategy that would eliminate duplicates. **Ronald Rousseau** devised the most complete strategy, but the results required a very complex validation process. All the strategies and results were sent

to **Eugene Garfield** and his assistant **Soren W. Paris**, who validated the results with their own results obtained from their direct searches in *Thomson Reuters* databases. In this case, there were still significant differences between the AD-based search strategy and the statistics compiled by **Paris**.

Based on previous personal experiences, **Loet Leydesdorff** proposed the use of a less problematic field to develop the search strategy: the *Source* (SO) field (**Zhou & Leydesdorff**, 2006). Thus, using the initial of the journal/proceedings title plus an asterisk (a truncation designed to retrieve all titles with the selected initial), the process of division into sections of fewer than 100,000 items was effective. The only problem was the

existence of journals belonging to series, which were retrieved not only by the journal title, but also by the series title. In any case, there were only two possibilities to obtain duplicated data; that is, a journal could be covered by no more than two sections of fewer than 100,000 items. For this purpose, a Boolean statement in the search strategy with the aim to eliminate duplicates could be developed. **Leydesdorff's** proposal was further developed by the *SCImago* research group, which finally devised a more accurate search strategy and developed the validation procedure.

3. Proposed search strategy

Table 1 shows the complete procedure devised to obtain the to-

Search strategy	Items	Sum
1. PY=2007 AND CU=USA AND (SO=A* OR SO=B*)	91,122	91,122
2. PY=2007 AND CU=USA AND (SO=C* OR SO=D* OR SO=E* OR SO=F* OR SO=G*)	91,920	183,042
3. PY=2007 AND CU=USA AND (SO=H* OR SO=I* OR SO=K* OR SO=L* OR SO=M*)	82,897	265,939
4. PY=2007 AND CU=USA AND (SO=N* OR SO=O* OR SO=P* OR SO=Q* OR SO=R*)	84,783	350,722
5. PY=2007 AND CU=USA AND (SO=S* OR SO=T* OR SO=U* OR SO=V* OR SO=W* OR SO=X* OR SO=Y* OR SO=Z* OR SO=1* OR SO=2* OR SO=3* OR SO=4* OR SO=5* OR SO=6* OR SO=7* OR SO=8* OR SO=9*)	58,751	409,473
6. PY=2007 AND CU=USA AND SO=J* AND AD=CA	17,064	426,537
7. PY=2007 AND CU=USA AND SO=J* NOT AD=CA	92,976	519,513
Statement to find overlapping	Items	Sum
8. (#1 AND #2) OR (#1 AND #3) OR (#1 AND #4) OR (#1 AND #5) OR (#1 AND #6) OR (#1 AND #7) OR (#2 AND #3) OR (#2 AND #4) OR (#2 AND #5) OR (#2 AND #6) OR (#2 AND #7) OR (#3 AND #4) OR (#3 AND #5) OR (#3 AND #6) OR (#3 AND #7) OR (#4 AND #5) OR (#4 AND #6) OR (#4 AND #7) OR (#5 AND #6) OR (#5 AND #7) OR (#6 AND #7)	23,026 (Overlapping)	496,487 (Σ 1-7) - 8
New Search Strategy (Excluding overlapping)	Items	Sum
9. #1 NOT #8	85,586	85,586
10. #2 NOT #8	87,535	173,121
11. #3 NOT #8	69,457	242,578
12. #4 NOT #8	75,516	318,094
13. #5 NOT #8	45,551	363,645
14. #6 NOT #8	17,008	380,653
15. #7 NOT #8	92,808	473,461
Sum 9-15 plus articles excluded by overlapping		496,487 (Σ 9-15) + 8

Table 1. Search strategy to obtain the total number of articles from the United States of America in 2007 through the WoS interface (Databases = SCI-Expanded, SSCI, A&HCI, CPCI-S, CPCI-SSH; Timespan = All years; All document types). May 18, 2009

tal number of articles produced by institutions from the United States.

The first 7 statements were created with the aim of dividing the results into sections of fewer than 100,000 items. In each statement, the necessary journal initials, in alphabetical order, to obtain an upper limit of fewer than 100,000 results were used. Note that statements #6 and #7 were shaped with the same philosophy as the United Kingdom output retrieval procedure. There were more than 100,000 USA articles published in journals whose titles begin with "J". Therefore, the AD field was used to divide this specific section in two: articles published in these journals including authors belonging to institutions from California (CA), and excluding them. In the end, a total number of 519,513 articles was obtained.

Then, a Boolean statement (#8) was created to identify overlapping and to improve the accuracy of the search strategy. Removing these 23,026 overlapping items from the previously calculated number, a final number of 489,487 items was obtained.

With the purpose of identifying inaccuracies in this calculation process, the first 7 statements were implemented again (#9 to #15), but excluding items in overlapping sections. This gave a result of 473,461 items. The items in the overlapping sections were added, and 489,487 items were once again obtained. This number established a hypothetical total number of articles published by institutions from the USA during the year 2007.

4. Validation process

The total scientific output of a less productive country than the USA or the United Kingdom was tested. This could have been any nation from the rest of the world, but we used Cuba as a test case. A

Search strategy	Items	Sum
1. PY=2007 AND CU=CUBA	910	910

Table 2. Search strategy to obtain the total number of Cuban articles in 2007 through the WoS interface: direct method (Databases = SCI-Expanded, SSCI, A&HCI, CPCI-S, CPCI-SSH; Timespan = All years; All document types)

direct method was used to find the Cuban scientific production in WoS during the year 2007 (Table 2).

A total of 910 items were identified using the word "Cuba" in the *Affiliation Country* (CU) field. So, the second step was to use the same strategy as the one developed to retrieve the total USA output. If the search strategy was correctly developed, the final number obtained by either of the two indirect methods (including and excluding overlapping sections) had also to be precisely 910. The table 3 confirms, finally, the accuracy of data obtained through the search strategy developed during the exercise.

Furthermore, results obtained from the WoS following this search strategy were in complete accordance with results reported independently by the Thomson Reuters team.

5. Final considerations

This exercise provided a methodology to obtain the same result in two different ways: a) searching with overlapping, and subtracting items in overlapping sections at the end; and b) searching without overlapping and adding the items in overlapping sections at the end. The use of a small country during the validation procedure allowed us to obtain the same total number not only through the proposed strategies, but also using a direct method, confirming the accuracy of the results and the efficacy of the search method.

This kind of SO-based search strategy is probably not the only alternative to retrieve USA scientific production in the WoS, and

it may be that its implementation does not solve other problems related to large numbers of items to be retrieved using the WoS interface. In any case, for scientometric purposes, a fast and well described method to obtain reliable data is always welcome. In this sense, the method devised is an accurate and validated search strategy to be used by any specialist around the world, and the procedure presented shows the importance of team work in the development of advanced search strategies for information retrieval.

6. Acknowledgments

To Eugene Garfield, for all the support and advices. To Félix de Moya Anegón, Carmen López Illescas, Elena Corera Álvarez, María Benavent Pérez (SCImago Research Group, Institute of Public Goods and Policies, CSIC), for the team work. Thomson Reuter's databases were available in Spain thanks to the Spanish Foundation for Science and Technology and the Ministry of Science and Innovation of the Spanish government.

7. References

Jacsó, Péter. "Errors of omission and their implications for computing scientometric measures in evaluating the publishing productivity and impact of countries". *Online information review*, 2009, v. 33, pp. 376-385.

Martínez, Luis-Javier. "La nueva versión de ISI Web of Knowledge: calidad y complejidad". *El profesional de la información*, 2008, v. 17, pp. 331-339.

Zhou, Ping; Leydesdorff, Loet. "The emergence of China as a leading nation in science". *Research policy*, 2006, v. 35, pp. 83-104.

Ricardo Arencibia-Jorge^{a,c}, Loet Leydesdorff^b, Zaida Chinchilla-Rodríguez^c, Ronald Rousseau^{d,e}, and Soren W. Paris^f

Search strategy	Items	Sum
1. PY=2007 AND CU=CUBA AND (SO=A* OR SO=B*)	140	140
2. PY=2007 AND CU=CUBA AND (SO=C* OR SO=D* OR SO=E* OR SO=F* OR SO=G*)	216	356
3. PY=2007 AND CU=CUBA AND (SO=H* OR SO=I* OR SO=K* OR SO=L* OR SO=M*)	161	517
4. PY=2007 AND CU=CUBA AND (SO=N* OR SO=O* OR SO=P* OR SO=Q* OR SO=R*)	193	710
5. PY=2007 AND CU=CUBA AND (SO=S* OR SO=T* OR SO=U* OR SO=V* OR SO=W* OR SO=X* OR SO=Y* OR SO=Z* OR SO=1* OR SO=2* OR SO=3* OR SO=4* OR SO=5* OR SO=6* OR SO=7* OR SO=8* OR SO=9*)	91	801
6. PY=2007 AND CU=CUBA AND SO=J* AND AD=Havana	108	909
7. PY=2007 AND CU=CUBA AND SO=J* NOT AD=Havana	35	944
Statement to find overlapping	Items	Sum
8. (#1 AND #2) OR (#1 AND #3) OR (#1 AND #4) OR (#1 AND #5) OR (#1 AND #6) OR (#1 AND #7) OR (#2 AND #3) OR (#2 AND #4) OR (#2 AND #5) OR (#2 AND #6) OR (#2 AND #7) OR (#3 AND #4) OR (#3 AND #5) OR (#3 AND #6) OR (#3 AND #7) OR (#4 AND #5) OR (#4 AND #6) OR (#4 AND #7) OR (#5 AND #6) OR (#5 AND #7) OR (#6 AND #7)	34 Overlapping	910 (Σ 1-7) - 8
New search strategy (excluding overlapping)	Items	Sum
9. #1 NOT #8	127	127
10. #2 NOT #8	205	332
11. #3 NOT #8	139	471
12. #4 NOT #8	177	648
13. #5 NOT #8	86	734
14. #6 NOT #8	108	842
15. #7 NOT #8	34	876
Sum 9-15 plus articles excluded by overlapping		910 (Σ 9-15) + 8

Table 3. Search strategy to obtain the total number of Cuban articles in 2007 through the WoS interface: indirect methods (Databases = SCI-Expanded, SSCI, A&HCI, CPCI-S, CPCI-SSH; Timespan = All years; All document types). May 19, 2009.

^a National Scientific Research Center CNIC, Avenue 25 and 158 street, AP 6414 Havana, Cuba.

^b Amsterdam School of Communication Research, University of Amsterdam, Kloveniersburgwal 48, 1012 CX Amsterdam, The Netherlands.

^c CSIC, Institute of Public Goods and Policies, SCImago Research Group, Albasanz 26-28, 28037 Madrid, Spain.

^d K.U.Leuven, Department of Mathematics, Celestijnenlaan 200B, 3001 Leuven (Heverlee), Belgium.

^e KHBO, Department Industrial Sciences and Technology, Zeedijk 101, 8400 Oostende, Belgium.

^f The Scientist, 400 Market St. Philadelphia, PA 19106 USA.

Address for correspondence:

Ricardo Arencibia-Jorge, National Scientific Research Center, Avenue 25 and 158 street, AP 6414 Havana, Cuba.

ricardo.arencibia@cnic.edu.cu

Si te interesan los

INDICADORES EN CIENCIA Y TECNOLOGÍA,

y todos los temas relacionados con la medición de la ciencia, tales como:

Análisis de citas, Normalización de nombres e instituciones, Impacto de la ciencia en la sociedad, Indicadores, Sociología de la ciencia, Política científica, Comunicación de la ciencia, Revistas, Bases de datos, Índices de impacto, Políticas de open access, Análisis de la nueva economía, Mujer y ciencia, etc.

Entonces **INCYT** es tu lista. Suscríbete en:

<http://www.rediris.es/list/info/incyt.html>

Estrategia para mejorar la difusión de los resultados de investigación con la Web 2.0

Por Daniel Torres-Salinas y Emilio Delgado-López-Cózar

Resumen: La comunicación científica se está viendo enriquecida por la utilización de nuevos modos de almacenamiento, publicación y difusión de los resultados. Entre ellos se encuentran los servicios de la Web 2.0, que aún son desconocidos en gran medida por los investigadores. En este contexto se indica cómo podemos emplear estratégicamente estos servicios para difundir y dar mayor visibilidad a las publicaciones científicas. Para ello se presenta una serie de aplicaciones de interés para la comunicación científica (Blogger, Twitter, Facebook, Slideshare y E-Lis) y se explica la función que pueden desempeñar para comunicar resultados científicos. A continuación se muestra cómo deben interrelacionarse lógicamente en lo que se ha denominado ciclo estratégico de difusión. Para concluir se recopila un conjunto de indicadores webmétricos clasificados en tres grupos (influencia social, utilización y reconocimiento) para conocer el éxito de los trabajos en la Web 2.0 y de la estrategia de difusión.

Palabras clave: Web 2.0, Ciencia 2.0, Publicación científica, Comunicación científica.

Title: Strategies to improve the dissemination of research results with the Web 2.0

Abstract: Scientific communication is being enriched by the introduction of new ways of storage, publication and dissemination of the results. These include the services of the Web 2.0 which are still largely unknown to researchers. In this context the objective of this paper is to illustrate how we can strategically use Web Services 2.0 to disseminate and give greater visibility to scientific publications. To this end we present a series of services of interest to scientific communication (Blogger, Twitter, Facebook, Slideshare y E-Lis) and explain the role they can play in communicating scientific results. It also shows how these services should be logically interrelated, which we call the strategic dissemination cycle. In the last section, the authors collected a set of webmetric indicators classified into three groups (social influence, use, recognition) to evaluate the success of the dissemination strategy and the visibility of a work in the Web 2.0.

Keywords: Web 2.0, Science 2.0, Scientific publication, Scientific communication.

Torres-Salinas, Daniel; Delgado-López-Cózar, Emilio. "Estrategia para mejorar la difusión de los resultados de investigación con la Web 2.0". *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 534-539.

DOI: 10.3145/epi.2009.sep.07



Daniel Torres-Salinas, doctor en documentación científica, trabaja como técnico de gestión de la investigación en la Universidad de Navarra, donde realiza auditorías sobre la calidad y el impacto de la investigación. Asimismo es miembro del grupo EC3 (Evaluación de la Ciencia y de la Comunicación Científica) de la Universidad de Granada, donde participa en diferentes proyectos.



Emilio Delgado-López-Cózar es profesor de metodología de la investigación en la Facultad de Comunicación y Documentación de la Universidad de Granada y miembro del grupo EC3. Sus líneas de investigación se centran en la evaluación de revistas científicas y de la ciencia, el estudio de la investigación en ByD, y la evaluación del rendimiento científico. Es uno de los promotores del índice In-Recs.

Introducción

EN LOS ÚLTIMOS AÑOS la comunicación científica se está viendo enriquecida por la utilización de nuevos modos de almacenamiento, publicación y difusión de resultados.

Nos encontramos en un momento de fragmentación de la comunicación científica (Brown;

Boulderstone, 2008) donde otros modos de transmitir los resultados de investigación empiezan a participar en el espacio reservado anteriormente a las revistas. Bohlin (2004), que habla de *regímenes de comunicación*, señala al menos tres tendencias bien consolidadas: repositorios, autoarchivo y revistas científicas. A éstas tendríamos que añadir recientemente las platafor-

mas de edición y publicación nacidas en torno a la Web 2.0 o Web social.

Todos estos medios, especialmente los relacionados con el mundo 2.0, han traído un gran número de aplicaciones en su mayoría desconocidas por los investigadores, más apegados a las prácticas de publicación tradicionales. Es evidente que la comunidad científica aún

“La comunicación científica se está viendo enriquecida por los nuevos modos de almacenamiento, publicación y difusión de resultados”

no tiene conciencia de su utilidad y hay enormes reticencias por la falta de reconocimiento que tengan sus resultados publicados en estos canales (Waldrop, 2008).

Sin embargo todo investigador debería tener claro que uno de los propósitos esenciales de la ciencia es la publicación de resultados ya que, como propugna Ziman (1998), la ciencia es ante todo un sistema de comunicación y estos canales son una oportunidad de fortalecer y optimizar dicho sistema. Según este autor la ciencia actual se caracteriza también por redes de trabajos donde la información se transmite entre los nodos de las mismas. La Web social no es otra cosa que un entramado digital de nodos que proporciona una cobertura tecnológica al concepto de Ziman. Los nuevos canales 2.0 pueden ser por tanto un escenario perfecto donde difundir las publicaciones.

El propósito de este trabajo es ilustrar sobre los múltiples usos y aprovechamientos que se pueden dar a las herramientas 2.0, en interacción con los repositorios, para difundir y comunicar resultados científicos, independientemente de que éstos sean inéditos o publicados, para alcanzar una mayor visibilidad y difusión.

En definitiva presentamos una estrategia para una transmisión más efectiva de los resultados científicos. Para ello tratamos en primer lugar de las aplicaciones generalistas 2.0 que podemos emplear; en segundo lugar cómo debemos interrelacionarlas y por último cómo

podemos monitorizar métricamente esta estrategia para conocer si estamos obteniendo éxito.

Servicios 2.0 útiles para la publicación científica

Del numeroso conjunto de servicios que se ofertan en la Web 2.0 hemos seleccionado cuatro, además de los repositorios en acceso abierto, todos ellos de carácter gratuito. La selección se ha realizado en función del número de usuarios, es decir del tamaño de la comunidad y por tanto de su audiencia potencial; además, casi todos ellos actúan como estándares en el mundo 2.0 y en última instancia todos ofrecen indicadores que nos permitirán monitorizar cuantitativamente nuestra estrategia. Estas aplicaciones son las siguientes:

– *Blogger*: servicio de alojamiento de blogs que hace posible la publicación periódica y secuencial de entradas (posts) con elementos textuales y multimedia.

– *Twitter*: servicio que facilita a sus usuarios el lanzamiento de mensajes cortos de carácter textual (*tweets*) que no superen los 140 caracteres, por lo que se conocen también como microblogs.

– *Facebook*: red social donde los usuarios pueden crear una cuen-

“Las herramientas 2.0 permiten difundir y comunicar resultados científicos para alcanzar una mayor visibilidad”

Servicios de alojamientos de blogs	
<i>Blogger</i>	https://www.blogger.com
<i>Wordpress</i>	http://es.wordpress.com
<i>La Coctelera</i>	http://www.lacoctelera.com
<i>Bitácoras.com</i>	http://bitacoras.com
Servicios de microblogging	
<i>Twitter</i>	http://twitter.com
<i>Jaiku</i>	http://www.jaiku.com
<i>Khaces</i>	http://www.khaces.com
Servicios de redes sociales	
<i>Facebook</i>	http://www.facebook.com
<i>Tuenti</i>	http://www.tuenti.com
<i>Hi5</i>	http://hi5.com
<i>Orkut</i>	http://www.orkut.com
Servicios de alojamiento de presentaciones	
<i>Slideshare</i>	http://www.slideshare.net/
<i>Zoho Show</i>	http://show.zoho.com
<i>Google docs</i>	http://docs.google.com
Repositorios temáticos	
<i>E-Lis</i> [documentación]	http://eprints.rclis.org/
<i>arXiv.org</i> [física, matemáticas]	http://arxiv.org
<i>RePEc</i> [economía]	http://repec.org
<i>PubMed Central</i> [ciencias de la salud]	http://www.pubmedcentral.nih.gov

Tabla 1. Breve directorio de servicios 2.0 y repositorios en acceso abierto

ta con contactos y compartir todo tipo de materiales (vídeo, texto, etc.). Cuenta con 200 millones de usuarios y una amplia penetración en el mundo universitario.

– *Slideshare*: plataforma de alojamiento de presentaciones que pueden ser visualizadas mediante un reproductor e integradas en otros sitios web.

– *Repositorios*: archivos digitales donde las comunidades científicas suben materiales poniéndolos a libre disposición. En el campo de la documentación contamos con *E-Lis*, que utilizaremos como ejemplo.

Sin embargo aunque éstas son las aplicaciones que proponemos, hay que tener claro que existen otras similares (en la tabla 1 se presentan sólo algunos ejemplos) por lo que la selección de una u otra dependerá en última instancia de la comunidad a la que nos dirijamos y habrá que optar por aquellas donde se produzcan las interacciones sociales de nuestra disciplina. Así por ejemplo, es evidente que en el caso de los repositorios debemos escoger aquel que más se ajuste al tema de nuestra especialidad. Una primera forma de comprobar la audiencia potencial de los servicios de redes sociales es la búsqueda automática que suelen realizar estas aplicaciones de los contactos que tenemos en nuestras cuentas de correo para ver cuántos de ellos están suscritos al servicio. Se recomienda asimismo utilizar en todas las aplicaciones el mismo nombre de usuario y que éste sea fácilmente identificable con nuestra firma científica (por ejemplo si nuestra firma científica es A. López-López utilizar como usuario *Alopezlopez*).

Función de las aplicaciones 2.0 y ciclo estratégico de difusión

Cada servicio 2.0 tiene una función específica en el proceso de difusión de los resultados de inves-



Figura 1. Función de las aplicaciones 2.0 en la difusión de resultados y ciclo estratégico de difusión

tigación. Podemos distinguir al menos tres usos (figura 1):

– *Aplicaciones de depósito*. Las vamos a aprovechar en primer lugar para el depósito y almacenamiento abierto de nuestras publicaciones. Es el caso de los repositorios como *E-Lis*, que emplearemos para depositar nuestros artículos, y *Slideshare* donde fundamentalmente colgaremos el material en pdf o ppt de las aportaciones realizadas en congresos o cursos.

– *Canal central*. Será el blog el que juegue el papel de canal central en el proceso de difusión ya que nos servirá para conectar las aplicaciones de depósito con los servicios de redes sociales. Será nuestra imagen en internet al ser recuperable desde cualquier buscador y desde él podemos anunciar todas nuestras actividades o resultados. Es además un medio que empieza a ser aceptado entre la comunidad científica (Torres-Salinas; Cabezas-Clavijo, 2008).

– *Canales de difusión*. Aunque todas las aplicaciones en última instancia sirven para dar difusión y visibilidad, las redes sociales recomendadas (*Twitter* y *Facebook*) son

idóneas para el anuncio inicial de las publicaciones y para comenzar a introducirlas en el circuito científico y profesional. Si mantenemos una ecología de contactos adecuada, seleccionando únicamente contactos profesionales presumiblemente interesados en nuestras líneas de investigación, ellos serán los primeros en conocer nuestros resultados iniciando el proceso de difusión entre los nodos y jugando un rol similar al del envío de las tradicionales separatas por correo electrónico.

Una vez determinada la función de cada uno de estos servicios hay que tener en cuenta la interrelación entre los mismos y el momento en el que hay que usar cada uno de ellos. Es lo que denominamos *el ciclo estratégico de difusión* (figura 1), que tendría diferentes etapas.

El ciclo debe comenzar con el depósito de nuestro trabajo en las aplicaciones recomendadas. Hay que señalar que cuando se trate de artículos aceptados para ser publicados en una revista científica el autor debe estar atento a las políticas de autoarchivo de la misma ya que, dependiendo del editor, la

posibilidad de poner en libre acceso el texto podría estar sometida a diferentes restricciones. Buena parte de las grandes editoriales científicas (*Kluwer*, *Springer* o *Elsevier*) no impiden el depósito de preprint y postprint aunque debe realizarse bajo ciertas condiciones. Por ejemplo, esta revista *El profesional de la información* permite el autoarchivo inmediato del artículo postprint siempre que se cite el url de su web. Antes de realizar el depósito, recomendamos consultar el directorio *Sherpa-Romeo* donde se detallan las políticas de copyright y autoarchivo de las editoriales científicas.

<http://www.sherpa.ac.uk/romeo>

Después de depositar el trabajo se redactará una entrada en el blog con el fin de informar a la comunidad de la libre disposición del mismo. Esta entrada debe tener un título descriptivo y el contenido debe recoger como mínimo la descripción bibliográfica completa (resumen incluido) y, si se desea y se cuenta con tiempo para ello, algunos comentarios del autor en un tono más divulgativo, explicando aspectos destacados de los resultados. No debemos olvidar incluir un enlace directo al documento en su lugar de depósito. Además si se trata de presentaciones depositadas en *Slideshare* podemos incrustar la presentación en la entrada gracias al código fuente que nos ofrece esta aplicación.

“Los trabajos divulgados por cauces tradicionales también tienen cabida en el puzzle de las tecnologías 2.0”

Por último redactaremos un mensaje corto en *Twitter* y *Facebook* para anunciar la publicación

a nuestros contactos ofreciendo en esta ocasión un enlace que remita a los lectores directamente a la entrada del blog.

Con el ciclo propuesto habremos conseguido que todos los servicios queden lógicamente conectados entre sí, con una función específica para cada uno de ellos.

La medición de la visibilidad

Uno de los objetivos principales de la difusión de resultados mediante este tipo de aplicaciones es conseguir una mayor visibilidad en internet que no tiene que traducirse necesariamente en un mayor impacto científico (medido mediante la citación) pero sí en un aumento directo en la utilización de los materiales. Para comprobar si nuestra estrategia funciona, tanto las aplicaciones 2.0 como la mayor parte de los repositorios nos permiten comprobar el interés despertado por nuestros resultados e incluso conocer por quién, dónde y cuándo es consultado el material. Podemos clasificar los indicadores ofrecidos por estos servicios en tres grupos (figura 2):

Indicadores de influencia social:

- Número de nodos: los contactos profesionales que hemos agregado en nuestras cuentas de las diferentes redes sociales como *Twitter* y *Facebook*. También podemos considerar como nodos el número de usuarios suscritos al feed de nuestro blog desde agregadores como *Bloglines* o *Google Reader*.

- Número de comentarios y número de réplicas: son las opiniones a modo de comentarios dejadas por otros usuarios en las diferentes entradas realizadas en las aplicaciones. Se pueden medir en el blog pero también en *Facebook* y en *Slideshare*. En *Twitter* podemos considerar como réplica la opción *reply to* y en *Facebook* la opción *like*.

Indicadores de utilización:

- Visitas: se pueden contabilizar mediante los múltiples sistemas de contadores y estadísticas existentes, que ofrecen además otro tipo de informaciones de gran utilidad (origen, intermediarios, comportamiento, etc.) (**Ortega; Aguillo**, 2009). Utilizaremos estos sistemas, por ejemplo *Google Analytics*, para monitorizar la utilización del blog y sus diferentes entradas.

- Número de reproducciones y número de reutilizaciones de las presentaciones: ambos indicadores son ofrecidos por *Slideshare* y nos permiten, en primer lugar, saber el número de usuarios que han reproducido nuestra presentación y, en segundo, aquellas presentaciones que han sido incluidas en otros sitios webs siendo reutilizadas y difundidas por otros usuarios.

- Número de descargas y número de lecturas del resumen: la cantidad de usuarios que han descargado nuestros trabajos es una información facilitada tanto por *Slideshare* como por los repositorios, como es el caso de *E-Lis*. Además, este último nos indica desde qué país se ha realizado. Junto con las descargas, *E-Lis* nos ofrece también el número de lecturas o visualizaciones del abstract.

Indicadores de reconocimiento:

- Número de enlaces: tradicionalmente los links se han considerado el equivalente de la citas en la Web (situation). Se pueden considerar los links recibidos por la entrada de nuestro blog en la que hemos anunciado la publicación, y pueden ser recuperados mediante servicios como *Technorati* o *Google Blogs*. También se deberían tener en cuenta los enlaces que reciben los documentos en sus lugares de depósito.

- Número de citas recibidas en *Google Scholar*: otra de las ventajas de la mayor parte de los repositorios, como ocurre en *E-Lis*, es

que su contenido es indizado y recuperable por otros servicios como *Google Scholar* permitiéndonos rescatar las citas recibidas.

“La publicación científica no debe acabar con la aceptación de un manuscrito por parte de una revista”

Como ejemplo real de estos indicadores podemos ilustrar cómo los autores de este trabajo en una aportación realizada en una mesa redonda titulada *Qué y cómo se edita una revista científica 2.0* ante un público presencial de 50 personas alcanzaron los siguientes resultados en la Web 2.0: la presentación ha sido reproducida en 809 ocasiones, descargada en 31, la entrada tiene una media de 29 visitas mensuales, recibió 2 comentarios, 1 enlace y la información llegó directamente a sus 30 contactos en *Twitter* y los 103 de *Facebook*. Gracias a la estrategia de difusión hemos pasado claramente a una audiencia más amplia y variada de la original conseguida con la suma de las microaudiencias de las diferentes aplicaciones. Analizando los indicadores mostrados, estas aplicaciones nos pueden ayudar a conocer cuáles son aquellos de nuestros trabajos que más respaldo e interés despiertan, lo que en cierta medida nos puede orientar temáticamente. En definitiva, estas aplicaciones están generando una serie de nuevos indicadores web-métricos que podíamos denominar de segunda generación y la suma de estas medidas estarían muy cerca de lo que **Cronin** (2005) denomina *Presence Density* o *Scholarly Saliency*.

Consideraciones finales

Cuando se habla de modelos de comunicación se tiende a tomar una

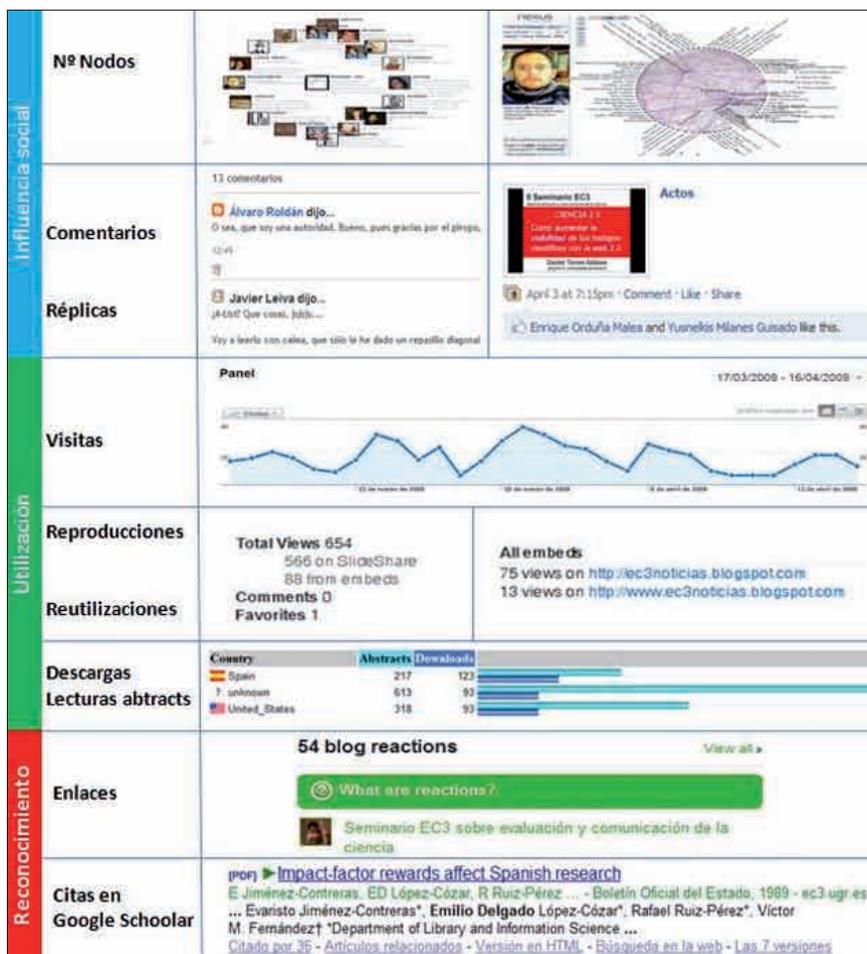


Figura 2. Ejemplo de algunos de los indicadores para monitorizar la visibilidad de nuestros trabajos en la Web 2.0

postura determinista y dicotómica en la que se cuenta con la inadecuación de la postura contraria o con la elección al menos por parte del investigador de una ellas (acceso abierto/acceso por pago, o publicación 2.0/revistas tradicionales). Sin embargo, tal como hemos expuesto, nos encontramos en un momento en el que hay diferentes regímenes de comunicación y ninguno de ellos acabará monopolizando la publicación científica. El escenario al que nos enfrentamos es el de una coexistencia pacífica entre distintos modelos no necesariamente excluyentes.

Con la estrategia propuesta se propugna que los trabajos divulgados por cauces tradicionales puedan ser reaprovechados en otros circuitos ya que también tienen cabida en el puzzle de las tecnologías 2.0. De esta forma, un trabajo científico di-

fundido en medios formales llegará a su comunidad natural, siendo reconocido su valor mediante la citación; pero además, difundiéndolo con las herramientas de la Web social cobra una nueva vida con una mayor audiencia potencial y un público más diversificado, como pueden ser los colectivos profesionales.

Por tanto creemos que la publicación científica no debe acabar con la aceptación de un manuscrito por parte de una revista y que es responsabilidad del científico, individual y social, dar la mayor publicidad posible a su trabajo.

Bibliografía

Ortega, José-Luis; Aguillo, Isidro F. “Minería del uso de webs”. *El profesional de la información*, 2009, v. 18, n. 1, pp. 20-26.

Bohlin, Ingemar. “Communication regimes in competition: the current transition in scholarly communication seen through the lens of the so-

ciology of technology". *Social studies of science*, 2004, v. 34, n. 3, pp. 365-391.

Brown, David J.; Boulderstone, Richard. *The impact of electronic publishing*. München: KG Saur, 2008. ISBN 978-3598115158.

Cronin, Blaise. *The hand of science: academic writing and its rewards*. Lanham, Toronto, Oxford: Scarecrow Press, 2005, ISBN 978-0810852822.

Torres-Salinas, Daniel; Cabezas-Clavijo, Álvaro. "Los blogs como nuevo medio de comunicación científica". En: *III Encuentro ibérico de docentes e investigadores en información y documentación*, 2008, pp. 773-781.

Ziman, John. *Real science: what it is, and what it means*. Cambridge: Cambridge University Press, 1998, ISBN 0521471710.

Waldrop, M. Mitchell. "Science 2.0 – Is open access science the future?". *Scientific American*, 2008, v. 298, n. 5, pp. 68-73.

Daniel Torres-Salinas

Grupo Evaluación de la Ciencia y la Comunicación Científica (EC3), Centro de Investigación Médica Aplicada, Universidad de Navarra,

Av. Pío XII, 31008 Pamplona (España).

torressalinas@gmail.com

Emilio Delgado-López-Cózar

Grupo Evaluación de la Ciencia y la Comunicación Científica (EC3), Departamento de Biblioteconomía y Documentación, Universidad de Granada, Colegio Máximo, Campus Cartuja, 18001 Granada (España). edelgado@ugr.es

Don't Federate – Integrate EBSCOhost® Integrated Search™

Hasta la fecha, los servicios de búsquedas federadas han sufrido cuatro problemas principales:

- Pocas funciones y herramientas
- Lentitud en las búsquedas
- Poco servicio al cliente
- Altos costes y cuotas de mantenimiento



EBSCOhost® Integrated Search™ lleva las búsquedas federadas a un nivel más alto, ofreciendo excelentes soluciones a estos problemas:

1 Avanzadas herramientas y funciones

Simultáneamente búsqueda de bases de datos de EBSCO y otros recursos electrónicos usando la plataforma EBSCOhost y sus herramientas.

2 Respuesta rápida a las búsquedas realizadas

Los resultados de las búsquedas son presentados por niveles, resultados de niveles más altos son presentados rápidamente, mientras que los más lentos o menos relevantes al tema, van buscando mientras se muestran los primeros resultados.

3 Servicio al Cliente de primer nivel.

Soporte técnico en el país, así como soporte en línea los 7 días de la semana las 24 horas.

4 Grandes ahorros en precio

EBSCO no cobrará los conectores a las bases de datos que estén a través de EBSCOhost, lo que ahorrará considerablemente los precios por conectores creados, al momento de solicitar el presupuesto.

Para más información visite www.ebscohost.com/integrate, o escriba un correo a request@ebscohost.com

Meridià: un portal para la difusión de la ciencia desde una óptica integral y cooperativa

Por Llorenç Arguimbau, Aldara Cervera, Robert Latorre y Mercè Martí

Resumen: El buen funcionamiento del ciclo de la investigación, el desarrollo y la innovación (I+D+I) requiere un tratamiento eficaz de la información científica. El Institut d'Estudis Catalans (IEC) dispone de un observatorio científico que, recientemente, ha presentado el portal web Meridià. Desde un punto de vista global, este proyecto tiene la voluntad de ser una iniciativa cooperativa para integrar y compartir información con los otros agentes del sistema de la ciencia, facilitándoles un conocimiento exhaustivo de su entorno y de la evolución de las diferentes áreas científicas y tecnológicas.

Palabras clave: Investigación, Desarrollo, Innovación, Información científica, Observatorio, Portal Meridià.

Title: Meridià: a web portal to disseminate science from a comprehensive and cooperative view

Abstract: The proper functioning of the Research, Development and Innovation (R&D&I) cycle requires an efficient processing of scientific information. The Institute for Catalan Studies (IEC) possesses a scientific observatory which recently presented the web portal Meridià. From a global point of view, this project aims to be a cooperative initiative to integrate and share information with other agents of the scientific system, providing a thorough understanding of their environment and the evolution of the different areas in science and technology.

Keywords: Research, Development, Innovation, Information Science, Observatory, Meridià portal.

Arguimbau, Llorenç; Cervera, Aldara; Latorre, Robert; Martí, Mercè. "Meridià: un portal para la difusión de la ciencia desde una óptica integral y cooperativa". *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 540-544.

DOI: 10.3145/epi.2009.sep.08



Llorenç Arguimbau es licenciado en historia del arte y diplomado en biblioteconomía y documentación por la Univ. de Barcelona. Postgrado en gestión cultural por la Univ. Pompeu Fabra (UPF) y diploma de estudios avanzados por la Univ. Autònoma de Barcelona. Ha trabajado como bibliotecario en la UPF. Es profesor asociado del Área de Documentación de la Univ. Autònoma de Barcelona y coordinador de l'Observatori de la Recerca del Institut d'Estudis Catalans.

Aldara Cervera es licenciada en documentación y en periodismo por la Universitat Autònoma de Barcelona, máster en gestión de contenidos digitales por la Universitat de Barcelona y la Universitat Pompeu Fabra. Trabaja como documentalista en el Observatori de la Recerca de l'Institut d'Estudis Catalans y, previamente, ha colaborado en medios de comunicación y en el ámbito bibliotecario.

Robert Latorre es licenciado en informática por la Universitat Politècnica de Catalunya. Ha trabajado en empresas tecnológicas como Centrisa, Teleinformàtica y Azertia, participando en la consultoría e implantación de diferentes sistemas de información. Es jefe de proyecto en el Servei d'Informàtica de l'Institut d'Estudis Catalans.

Mercè Martí es licenciada en documentación por la Universitat Autònoma de Barcelona, diplomada en magisterio por la Universitat de Vic y titulada en información para la gestión de la innovación por la Universitat Oberta de Catalunya. Ha trabajado en la Oficina de Gestió de Projectos del Internet Interdisciplinary Institute. Actualmente, trabaja como documentalista del Observatori de la Recerca del Institut d'Estudis Catalans.

Introducción

EL CICLO DE LA INVESTIGACIÓN, el desarrollo y la innovación (I+D+I) ocupa una posición estratégica al potenciar el conocimiento científico y el crecimiento económico. Para el buen funcionamiento de este ciclo, es imprescindible llevar a cabo un tratamiento eficaz de la información científica.

Hacer ciencia es sinónimo de

hacer públicos los resultados (Rodríguez-López, 2005). Por lo tanto, los análisis documentales sobre la ciencia se han centrado en áreas como el movimiento *Open Access* (OA), la bibliometría o, más recientemente, la Ciencia 2.0 (Cabezas-Clavijo; Torres-Salinas; Delgado-López-Cózar, 2009; Codina, 2009). Cabe reivindicar una perspectiva más amplia que reúna los recursos, los resultados y las

formas de organización involucrados en la creación de conocimiento (Callon; Courtial; Penan, 1995). Para impulsar la investigación, ésta ha de estudiarse científicamente, aplicando con rigor el mismo método científico.

A partir de 1990, con la creación en Francia del *Observatoire des Sciences et des Techniques* (OST), este tipo de organismos ha proliferado en diversos países. Los

observatorios de ciencia y tecnología pueden definirse como organizaciones concebidas para diseñar, integrar y producir información, indicadores y análisis sobre el sistema nacional de I+D+I bajo los estándares y metodologías internacionales (OCDE, 2003; De-la-Vega, 2007).

La tarea de descripción y valoración de un sistema de I+D+I no resulta nada fácil (Fuentes; Arguimbau, 2008). Hay una gran atomización de los datos, derivada de la enorme cantidad de agentes ejecutores y financiadores, que se caracterizan por su pluralidad (naturaleza jurídica, sector de trabajo, dimensiones) y por su interacción dinámica (redes, depósitos de conocimiento, etc.). Además, surgen constantemente nuevas áreas, conceptos, métodos, técnicas y espacios interdisciplinares.

Marco organizativo

En los últimos años, el *Institut d'Estudis Catalans (IEC)* ha ejercido un papel cada vez más importante como entidad consultora de los agentes públicos, mediante la elaboración de informes y dictámenes sobre aspectos estratégicos de política científica. El *IEC*, la academia catalana de las ciencias y las humanidades, se creó en 1907 como un instrumento al servicio de las políticas de investigación y difusión del conocimiento. Después de un largo período de clandestinidad, un Real decreto aprobado el 26 de noviembre de 1976 le otorgaba reconocimiento oficial como "corporación académica que tiene por objeto la alta investigación científica". Su espacio de actuación se extiende a los territorios de lengua y cultura catalanas.

En 1995, por encargo de la *Generalitat de Catalunya*, el *IEC* puso en marcha el proyecto *Reports de la recerca a Catalunya*. Se trata de una serie de informes por materias específicas sobre el estado de la investigación en Catalunya y redacta-

dos por equipos de investigadores de prestigio. Por ahora se han publicado dos ediciones, que abarcan los períodos 1990-1995 (24 informes) y 1996-2002 (27 informes).

Para facilitar un apoyo documental a las tareas de elaboración de la segunda edición de los *Reports*, el *IEC* creó en 2003 el *Observatori de la Recerca de l'Institut d'Estudis Catalans (OR-IEC)*. Desde entonces, el *OR-IEC* ha ido incorporando datos detallados sobre el sistema I+D+I de las zonas de lengua y cultura catalanas, con comparativas nacionales e internacionales.

El portal Meridià

Durante 2008 el *IEC* puso en marcha el portal web *Meridià (Medición de la investigación, el desarrollo y la innovación)* (figura 1), un servicio ideado para diferentes tipos de clientes. Entre los puntos fuertes del proyecto destaca la globalidad del público (universidades, empresas, administraciones públicas, entidades privadas sin ánimo de lucro, etc.) y de los temas (ciencias, ciencias de la vida y de la salud, ingenierías y arquitectura, ciencias sociales y humanidades). Con la finalidad de organizar y facilitar la comparación de los datos, se trabaja con la clasificación de las áreas científicas de los *Reports* y se presentan equivalencias con otras taxonomías estandarizadas.

El propósito básico de *Meridià* es impulsar y difundir la ciencia a partir de un conocimiento detallado de todo el ciclo de la investigación científica, desde los *inputs* recibidos hasta los *outputs* obtenidos: contexto socioeconómico, organismos de investigación, recursos (humanos, económicos y materiales) y, finalmente, resultados (artículos, patentes y tesis). El enfoque evolutivo del proyecto parte de la recopilación de los datos dispersos y heterogéneos, continúa con su transformación en información incorporándola a la

plataforma de *Meridià* y acaba con el tratamiento de la información para la generación de conocimiento y la consiguiente toma de decisiones. La utilidad de la información contenida en el proyecto es muy diversa: desde respuestas a preguntas concretas (servicio de referencia virtual) hasta la elaboración de informes estratégicos.

El portal se encuentra disponible en catalán, castellano e inglés y presenta la siguiente información (tabla 1):

- Informes estadísticos y dinámicos, con datos metodológicos sobre alcance geográfico y temporal, fuente de datos, etc.
- Documentos (memorias, informes, planes de I+D+I, etc.).
- Motor de búsqueda en la base de datos y los documentos.
- Noticias de actualidad.
- Enlaces web.

Categorías	Ítems
Clasificaciones	26
Entidades	12.765
Recursos económicos (expedientes)	68.500
Recursos humanos	37.429
Resultados (revistas)	1.092
Resultados (tesis doctorales)	11.621
Resultados (patentes)	8.795

Tabla 1. Información de la base de datos Meridià

Meridià dispone de diversos niveles de consulta de la información, desde el privilegiado para los organismos con los que se establecen convenios, hasta la información básica para el público en general. Así, el sistema resulta flexible para establecer perfiles específicos para cada grupo de usuarios potenciales:

- Cargos políticos de las administraciones públicas.



Figura 1. Meridià. <http://meridia.iec.cat>

– Rectores y vicerrectores de las universidades públicas y privadas.

– Directores y responsables de organismos, institutos y centros de investigación.

– Personal gestor y técnico en política científica y tecnológica.

– Miembros, personal y socios de las sociedades filiales del IEC.

– Personal investigador especializado en la descripción, el análisis y la prospectiva de las diferentes actuaciones de I+D+I.

– Comunidad científica y sociedad en general.

Este proyecto es una iniciativa cooperativa para integrar y compar-

“Se ofrecen desde respuestas a preguntas concretas hasta la elaboración de informes estratégicos”

tir información con el resto de los agentes del sistema de I+D+I, facilitándoles un conocimiento exhaustivo de su entorno y de la evolución de las diferentes áreas científicas y tecnológicas (no se realizan análisis comparativos entre entidades ni investigadores). Todos los datos sobre recursos y resultados de que se dispone son fruto de la tarea documental llevada a cabo por el *OR-IEC*, que contempla la identificación, selección y vaciado sistemático de múltiples fuentes de información (figura 2) y el tratamiento informático de los datos recopilados, que provienen básicamente de dos tipos de fuentes:

– Convenios para la transferencia de información con organismos públicos y privados vinculados a la investigación.

– Procesamiento sistemático de fuentes de información primarias (boletines oficiales, registros administrativos, etc.) y secundarias (memorias, informes, etc.).

Como se puede comprobar, las

fuentes son públicas y oficiales y eso determina que se pueda hablar de datos sólidos y fiables que, una vez tratados y normalizados se incorporan a una base de datos para su posterior difusión. A modo de aplicación interna, el *OR-IEC* también dispone de una intranet donde se conservan documentos y fuentes vinculados con la investigación (informes, memorias, resoluciones de convocatorias, sitios web de interés, etc.).

“Se trata de mejorar el conocimiento de las áreas científicas y tecnológicas de los agentes del sistema de I+D+I”

Además del portal web que muestra los resultados generados por el *OR-IEC*, el diseño tecnológico de *Meridià* presenta una plataforma con diferentes módulos:

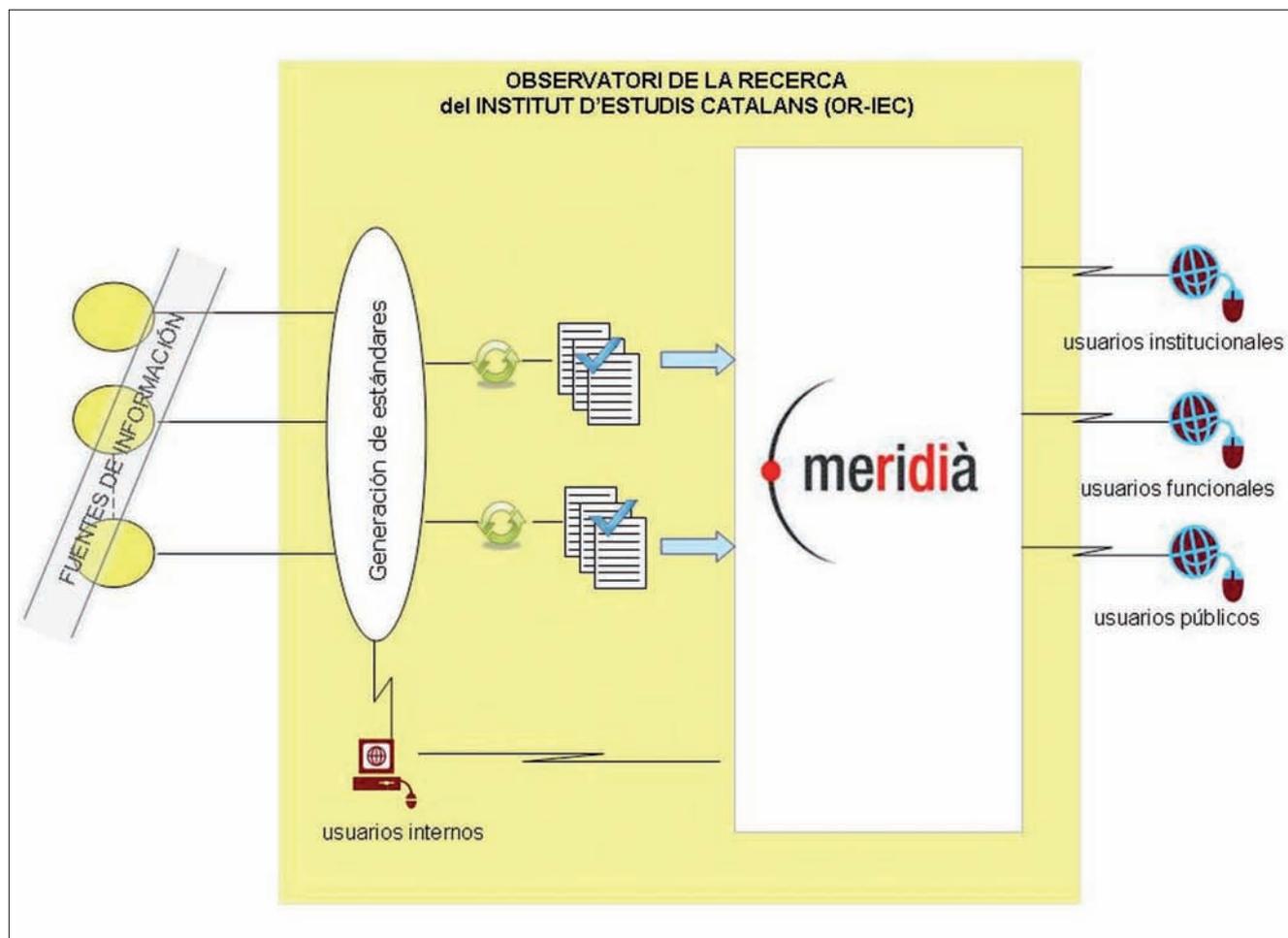


Figura 2. Meridià. Proceso de adquisición y carga de datos

- Base de datos construida sobre un esquema conceptual flexible para adaptarse a la complejidad de las múltiples fuentes de información.

- Entorno interno que facilita las tareas de administración de la información. Esta aplicación y el portal se han implementado en lenguaje *Java* con el *Framework Struts* y se han desplegado en un servidor de aplicaciones *Apache Tomcat 5.5*. La capa de datos utiliza la base de datos *Oracle 10 GR* y ficheros *xml* donde se almacenan los contenidos no estructurados (noticias, enlaces, documentos, etc.).

- Entorno analítico que garantiza la explotación de los datos para generar informes. Técnicamente, se ha confeccionado con *MicroStrategy 8.0* y con el componente de visualización de informes *IEPortal*.

A partir de la experiencia adquirida por el *OR-IEC* y de los comentarios recibidos, se prepara la fase 2 del portal que contempla una serie de nuevos servicios y mejoras funcionales que se agrupan en las siguientes líneas de trabajo:

- Rediseñar la interfaz *Meridià* y orientarla hacia un usuario no experto.

- Abrir una parte sustancial de los contenidos a todo tipo de usuarios.

- Aumentar la interacción con usuarios y sistemas externos.

- Desplegar un motor de búsqueda integral.

Consideraciones finales

Meridià es un observatorio interactivo para facilitar los datos y los instrumentos de búsqueda y

análisis a un público potencial. Su finalidad es impulsar y difundir la ciencia a partir de un conocimiento de los fondos invertidos y de los resultados obtenidos. Se trata de una iniciativa del *IEC* con un gran potencial de crecimiento futuro a partir de la cooperación con el resto de agentes científicos y tecnológicos. *Meridià* aspira a ser un eslabón más en la función original de internet, es decir, potenciar la red de redes como una comunidad real para el conocimiento científico.

Bibliografía

Arguimbau, Llorenç; Cervera, Aldara; Latorre, Robert; Martí, Mercè; Guerrero, Ricard. "El portal Meridià de l'Observatori de la Recerca (OR-IEC): anàlisi, promoció i difusió del coneixement científic català". *Coneixement i societat*, 2008, n. 15, pp. 80-105. <http://eprints.rclis.org/15676>

Cabezas-Clavijo, Álvaro; Torres-Salinas, Daniel; Delgado-López-Cózar, Emilio. "Ciencia 2.0: catálogo de herramientas e implicaciones

para la actividad investigadora". *El profesional de la información*, 2009, v. 18, n. 1, pp. 72-79.

Callon, Michel; Courtial, Jean-Pierre; Penan, Hervé. *Cienciometría: la medición de la actividad científica: de la bibliometría a la vigilancia tecnológica*. Gijón: Trea, 1995.

Codina, Lluís. "Ciencia 2.0: redes sociales y aplicaciones en línea para académicos". *Anuario hipertext.net*, 2009, n. 7.
<http://www.hipertext.net/web/pag295.htm>

De-la-Vega, Iván. "Tipología de observatorios de ciencia y tecnología. Los casos de América Latina y Europa". *Revista española de documentación científica*, 2007, v. 30, n. 4, pp. 545-552.

Fuentes-Pujol, Eulàlia; Arguimbau, Llorenç. "I+D+I: una perspectiva documental". *Anales de documentación*, 2008, n. 11, pp. 43-56.
<http://revistas.um.es/analesdoc/article/viewFile/24801/24121>

Ocde. *Manual de Frascati 2002: propuesta de norma práctica para encuestas de investigación y desarrollo experimental*. Madrid: Fecyt, 2003.

Rodríguez-López, Joaquín. "Ciencia y comunicación científica: edición digital y otros fundamentos del libre acceso al conocimiento". *El profesional de la información*, 2005, v. 14, n. 4, pp. 246-254.

Rovira-Fernández, Anna. "Les biblioteques universitàries i el suport a la recerca". *BiD: textos*

universitaris de biblioteconomia i documentació, 2007, n. 18.

<http://www.ub.edu/bid/pdf/18rovir1.pdf>

Llorenç Arguimbau, Aldara Cervera, Robert Latorre, Mercè Martí
Institut d'Estudis Catalans. C/ Carme, 47, 08001 Barcelona.
llarguimbau@iec.cat
acervera@iec.cat
rlatorre@iec.cat
mmarti@iec.cat



Licencia de acceso es el modelo flexible de NPG
que le permite seleccionar las suscripciones más importantes para su institución. Elige entre publicaciones de alta calidad de Nature Publishing Group y Palgrave Macmillan para crear una licencia que se ajuste a sus requerimientos y a las necesidades de sus usuarios.
Sin paquetes. Total flexibilidad. Su elección.

Para más información, diríjase a www.nature.com/libraries

palgrave macmillan nature publishing group npg

Servicio público de televisión y patrimonio audiovisual: el proyecto VideoActive

Por David Fernández-Quijada y Montse Fortino

Resumen: La necesidad de conservar los archivos televisivos y las posibilidades de difusión derivadas de la digitalización han generado un creciente interés debido al despliegue de plataformas desde las que proporcionar acceso a estos contenidos. Una de estas plataformas es VideoActive, un proyecto colaborativo paneuropeo con una aproximación temática tanto a la historia de Europa a través de la televisión como a la historia de la televisión europea, que tiene entre sus metas establecer puentes de colaboración con el mundo académico. El consorcio ha puesto en marcha su propia infraestructura técnica para suministrar hasta 10.000 piezas de diez archivos televisivos europeos, entre ellos Televisió de Catalunya, el operador público catalán que de esta manera busca proyectar internacionalmente la cultura catalana.

Palabras clave: Servicio público de radiodifusión, Patrimonio audiovisual, Archivos audiovisuales, Archivos televisivos, VideoActive, Televisió de Catalunya.

Title: The VideoActive project: Television as public service and audiovisual patrimony

Abstract: The need to preserve audiovisual archives and the possibilities for their dissemination resulting from digitisation have generated a growing interest in the development of platforms to provide access to their content. One of these platforms is VideoActive, a pan-European collaborative project that has developed a thematic approach to European history through TV and to the history of European TV. One of its key objectives is to encourage collaboration with the academic world. The consortium has developed its own technical infrastructure to offer at least 10,000 items from 10 different European archives, including that of Televisió de Catalunya, the Catalan public service broadcaster that seeks to contribute to international awareness of Catalan culture.

Keywords: Public service broadcasting, Audiovisual heritage, Audiovisual archives, Television archives, VideoActive, Televisió de Catalunya.

Fernández-Quijada, David; Fortino, Montse. "Servicio público de televisión y patrimonio audiovisual: el proyecto VideoActive". *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 545-551.

DOI: 10.3145/epi.2009.sep.09



David Fernández-Quijada es licenciado en comunicación audiovisual y doctor por la Univ. Autònoma de Barcelona, en cuyo Departament de Comunicació Audiovisual i de Publicitat ejerce como profesor ayudante doctor. Es investigador asociado al Grup de Recerca en Imatge, So i Síntesi (Griss), donde lleva a cabo investigaciones sobre tecnologías de la comunicación y el servicio público de radiodifusión.



Montse Fortino es diplomada en biblioteconomía y documentación por la Universitat de Barcelona. Trabaja como documentalista en Televisió de Catalunya (TVC) desde 1991, donde ha colaborado en distintos programas documentales como: *Vides Privades*, *Explica'ns la teva vida*, *Material sensible*, *Un lloc per viure* o *30 minuts*. Participa en el proyecto VideoActive desde enero de 2008.

1. Servicio público y patrimonio audiovisual

MEDIO DE CONSUMO MASIVO por excelencia, la televisión es el referente cultural, estético e informativo contemporáneo de más amplia difusión.

Contribuye por tanto a nuestra concepción del mundo, presente y pasado. De ahí la importancia de preservar sus fondos como memoria cultural e histórica. Una tarea que, además, les corresponde a los

operadores de televisión como parte de la función de servicio público que tienen adjudicada.

Esta idea no siempre se plasma en la realidad, pues los canales pri-

vados en general tienen poca predisposición a hacerlo. En cambio algunos operadores televisivos públicos están construyendo un archivo que se configura como parte fundamental de la memoria histórica del país.

“El archivo de un radiodifusor refleja la evolución de sus contenidos a lo largo del tiempo y cumple con la función de repositorio como parte de la memoria común de una sociedad”

Esto implica una serie de responsabilidades notables, desde su efectiva conservación con las máximas condiciones de seguridad hasta la posible consulta de ese archivo por investigadores especializados o por la ciudadanía en general.

La importancia de las televisiones públicas en este campo no radica solamente en el hecho de que posean por razones históricas los archivos audiovisuales más extensos, sino también en que la lógica utilitarista del libre mercado que siguen las televisiones privadas no garantiza una adecuada preservación del patrimonio audiovisual que, por otra parte, conlleva altos costes. Esta tarea que normalmente recae sobre los correspondientes departamentos de documentación, forma parte cada vez más de las funciones del servicio público que le son asignadas en ese contrato social que se establece entre medios públicos y ciudadanos.

Legalmente hay pocas obligaciones al respecto para los radiodifusores y nada cercano a la idea del depósito legal que rige por ejemplo para los libros. La principal novedad en los últimos años la constituye la *Ley 17/2006*, de 5 de junio, de la radio y la televisión de titularidad estatal, que regula al operador público *Radio Televisión Española*, y en la que se cita entre sus funciones de servicio público la de “velar por la conservación de los archivos históricos audiovisuales”, lo que se traduce en el artículo 52 de su preceptivo *Mandato-Marco*, dedicado al patrimonio audiovisual y en el que, entre otras cosas, se establece que “la *Corporación Rtve* tiene la obligación, en el período de vigencia del presente *Mandato-Marco*, de digitalizar, conservar y preservar en condiciones óptimas para su uso y consulta, el archivo histórico heredado de *Rtve*” y también que “la *Corporación Rtve* facilitará que esos archivos tengan presencia en bus-

cadore y puedan ser visualizados o escuchados a través de las redes telemáticas, fijas y móviles”.

<http://www.boe.es/boe/dias/2006/06/06/pdfs/A21207-21218.pdf>

<http://www.boe.es/boe/dias/2008/06/30/pdfs/A28833-28843.pdf>

El archivo de un radiodifusor refleja la evolución de sus contenidos a lo largo del tiempo y, como tal, cumple con la relevante función de repositorio. La memoria colectiva proporciona a una sociedad la posibilidad de satisfacer su necesidad simbólica de establecer un lazo con la historia de la comunidad (Lloyd, 2007). Los archivos televisivos más antiguos se corresponden con materiales cinematográficos utilizados en el medio, tales como noticiarios rodados en soporte químico. No fue hasta la invención en 1956 de la cinta magnética de dos pulgadas por la casa *Ampex* que se dio la posibilidad de sistematizar un archivo, aunque la falta de conciencia de memoria histórica ha privado en gran medida a los historiadores de esta fuente (Gómez-Alonso, 2004). En otros casos, la dificultad de acceder físicamente a los archivos ha restringido su posible consulta y utilización.

Consultar los archivos televisivos ha resultado históricamente una tarea complicada. Sin embargo, el tránsito al entorno digital se presenta como “el momento de que los archivos audiovisuales, en especial los de las televisiones públicas, se replanteen su misión, sus objetivos y lleven a cabo estrategias de comunicación” (Agirreazaldegui-Berriozabal, 2007, p. 440). En el caso de internet “significa liderar el proceso de coordinación de las diversas instituciones, sobre todo las públicas, en la gestión de la memoria digital, fuente de la información del futuro” (De-Moragas; Prado, 2000, p. 392). Además, la accesibilidad en línea constituye hoy en día una condición indispensable para optimizar los beneficios que pueden extraer

de estos contenidos los ciudadanos y los colectivos sociales. Así lo han entendido radiodifusores públicos como la *BBC* británica con su servicio *BBC Motion Gallery* o la *RAI* italiana, así como el gran centro documental francés, el *Institut National de l'Audiovisuel (INA)*.

<http://www.bbcmotiongallery.com>

<http://www.rai.tv>

<http://www.ina.fr/archivespourtous/>

Desde una perspectiva continental, la *Comisión Europea* ha venido trabajando en la conservación y expansión del patrimonio cultural europeo. A modo de ejemplo, la iniciativa *i2010* sobre bibliotecas digitales recuerda que “la presencia en línea de este material procedente de distintas culturas y expresado en diferentes lenguas permitirá a los ciudadanos apreciar más fácilmente su propio patrimonio cultural y el de los demás países europeos y utilizarlo con fines de estudio, trabajo u ocio. Contribuirá, de esta manera, a complementar y apoyar los objetivos de la actuación de la Unión Europea en el ámbito de la cultura” (*Comisión de las Comunidades Europeas*, 2005, p. 4). La multiculturalidad y el multilingüismo de las expresiones culturales afloran pues como elementos centrales de *i2010*.

“VideoActive pretende mostrar las similitudes y las diferencias en la presentación televisiva de la realidad social y cultural europea”

2. El proyecto VideoActive

En este contexto de progresiva concienciación sobre la necesidad de preservar y facilitar la difusión del patrimonio audiovisual, la llegada de la digitalización ha impulsado el surgimiento de varios proyectos

colaborativos. La colaboración entre diferentes organismos dedicados a la preservación de documentos ha sido una constante desde el siglo pasado, pero la aparición de las tecnologías digitales ha convertido en más aguda esta necesidad por las nuevas habilidades que requiere el tratamiento de estos materiales y las dificultades por asumirlas todas bajo un mismo techo (Besser, 2007). Esta voluntad de trabajo conjunto se reflejó por ejemplo en el proyecto *Presto* (Teruggi, 2004; Wright, 2004), centrado en el análisis de tecnologías específicas para la digitalización de archivos audiovisuales.

<http://www.prestospace.org>

En *Presto* participaban algunos radiodifusores que se volverían a encontrar en *Birth (Building an Interactive Research and Delivery Network for Television Heritage)*, otro proyecto de alcance continental nacido del mismo espíritu colaborativo (Hecht; O'Dwyer; Oomen; Scharinger, 2004) y que construyó la base técnica sobre la que trabaja el proyecto *VideoActive*.

<http://www.birth-of-tv.org>

<http://www.videoactive.eu>

VideoActive muestra las similitudes pero también las diferencias en la presentación televisiva de la realidad social y cultural europea, así como el rol histórico que ha jugado la televisión en la creación de esta experiencia común. Financiado con más de 2,5 millones de euros por el programa *eContentPlus* de la Comisión Europea, *VideoActive* comenzó el 1 de septiembre de 2006 para un período de 36 meses que finalizó el pasado 31 de agosto de 2009.

http://ec.europa.eu/information_society/activities/econtentplus/programme/index_en.htm

2.1 Contenido

Los documentos audiovisuales que se pueden consultar en *Video-*

Active pertenecen a los archivos de las emisoras públicas *British Broadcasting Corporation (BBC)*, *Deutsche Welle (DW)*, *Danmarks Radio (DR)*, *Österreichischer Rundfunk (ORF)*, *Radio Télévision Belge de la Communauté Française (RTBF)*, *Vlaamse Radio- en Televisieomroep (VRT)* y *Televisió de Catalunya (TVC)*. También aportan sus fondos otros entes como el *Nederlands Instituut voor Beeld en Geluid*, el húngaro *Neumann János Digitális Könyvtár és Multimédia Központ Kht.*, el italiano *Istituto Luce*, el sueco *Sveriges Nationalarkiv för Audiovisuella Medier*, el griego *Hellenic National Audiovisual Archive (Hena)*, el francés *Institut National de l'Audiovisuel (INA)* y la compañía británica *Moving Image Communications Ltd.* Junto a estos proveedores de contenidos, el peso técnico del proyecto recae sobre la compañía holandesa *Note-rik BV*, que ha diseñado el portal y gestiona el material audiovisual, y la *National Technical University of Athens*, que gestiona los metadatos y el tesoro. Además, las facultades de comunicación de la *Universiteit Utrecht* y del *Royal Holloway* de la *University of London* asesoran en cuanto a los contenidos.

La selección de contenidos ha sido un aspecto clave en el que se ha trabajado desde el primer momento dada la variedad de archivos participantes, así como las diferencias en materia de propiedad intelectual y de disponibilidad de metadatos. La aproximación a los contenidos ha sido doble: abrir una ventana a la historia de Europa en la televisión y, por otro lado, mostrar la historia de la televisión europea. Este último aspecto se complementa además con artículos y textos elaborados para ayudar a comprender la historia televisiva en el continente.

http://videoactive.files.wordpress.com/2007/10/23_content_selection_strategy_report.pdf

“VideoActive muestra la historia de Europa en la televisión, y la historia de la televisión europea”

Los materiales van acompañados de metadatos que ayudan en una tarea comparativa a la que también contribuye la propia política de selección de contenidos: priman aquellos que conllevan una comparación entre las diferentes culturas europeas, por ejemplo, el proceso de integración continental, la guerra fría o las distintas festividades nacionales. Así, esta aproximación temática se centra en aquellos materiales televisivos que tienen mayor capacidad de dar a conocer y ayudar a comprender la historia común de los europeos y las expresiones culturales del continente. Cada uno de los diez archivos participantes se ha comprometido a aportar hasta 1.000 piezas, algunas simplemente gráficas pero en su mayor parte imágenes en movimiento, lo que totaliza 10.000 piezas, de duración variable. En todo caso, una ínfima parte de lo disponible en sus archivos. El período de tiempo que abarcan los contenidos se inicia en 1907 con una filmación del día nacional danés, y llega hasta nuestros días.

2.2 Tecnología

La plataforma tecnológica del proyecto consta de tres partes bien diferenciadas: la aplicación con la que los participantes aportan sus contenidos, los servidores en los que se aloja el material digitalizado y el portal, el *front-end* con el público.

La solución diseñada para el proceso de contribución de archivos es el primer elemento a tener en cuenta. Los archivos se han gestionado con independencia de los metadatos. Los primeros se enviaban a los servidores de destino. Para los metadatos, la mayoría de los pro-

veedores de contenidos han optado por introducirlos directamente mediante una interfaz web implementada como *back-office* en el portal de *VideoActive*. Otros, como *TVC* o el *Istituto Luce*, han optado por preparar la información en una base de datos propia para después exportarla al portal mediante ficheros *excel*.

El segundo elemento técnico sobre el que se sostiene *VideoActive* son los servidores. *Noterik* mantiene un servidor central en Holanda, pero las diferencias en las regulaciones nacionales no han permitido a algunos archivos almacenar sus piezas en servidores ubicados en el extranjero, como es el caso del *Istituto Luce*, que gestiona

su propio servidor. La falta de una regulación común sobre propiedad intelectual ha obligado a flexibilizar este aspecto y, en general, ha originado algunos de los principales contratiempos en el cumplimiento del calendario previsto. Sea desde una localización u otra, el material videográfico se codifica a 25 fps en mpeg4/h.264 a una resolución de 640x480 píxeles, lo que asegura una buena calidad incluso si la reproducción se realiza a pantalla completa. El audio, por su parte, se codifica en mp3 a 96 Kb/s.

El portal constituye el tercer y último elemento de la plataforma tecnológica. La primera versión beta del mismo fue puesta en mar-

cha en febrero de 2008 y lanzada públicamente el 16 de septiembre del mismo año. Dado su carácter abierto, la consulta a sus contenidos puede beneficiar a cualquier ciudadano. No obstante, su utilización en entornos educativos ha sido uno de los motores que ha impulsado el proyecto. No sólo en la educación universitaria sino también como material de trabajo para la educación preuniversitaria.

La interfaz de consulta web de *VideoActive* es similar a la de otros sistemas de búsqueda ampliamente difundidos en la Red. La búsqueda avanzada, además, pone a disposición del público el contenido en función de idioma,

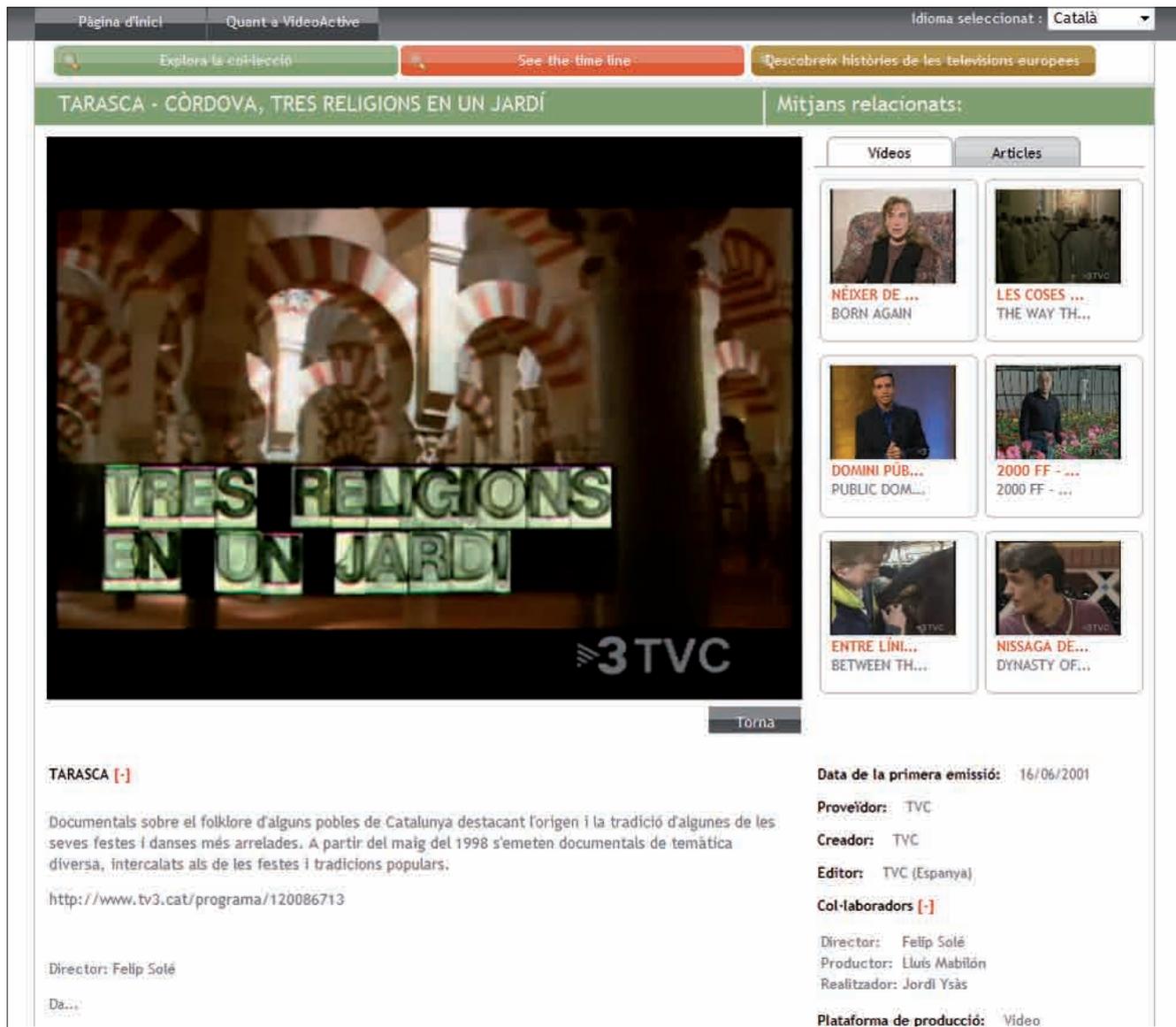


Figura 1. Interfaz web de VideoActive

tema, archivo propietario del material, género televisivo, década de producción, fecha de emisión, etc. La navegación se puede efectuar hasta en diez idiomas diferentes: inglés, alemán, francés, italiano, holandés, griego, húngaro, danés, sueco y catalán. El propósito declarado de esta aproximación plurilingüística es “reforzar la comprensión de las historias compartidas y las interrelaciones que han dado forma a la memoria y a la identidad europea colectiva, y a la vez conmemorar la dimensión multicultural que también ha forjado la ciudadanía europea”.

http://www.videoactive.eu/VideoActive/About.do?menu_page=menu-about

Para el tesoro se utiliza *TheSauriX*, que posibilita anotación multilingüe, de manera que los interesados pueden acceder a los contenidos con independencia del idioma en que éstos se encuentren. Los metadatos también se representan en rdf (*resource description framework*) y están disponibles públicamente mediante el protocolo OAI-MPH.

Más allá del propio portal de *VideoActive*, los contenidos son accesibles desde *Europeana*, la biblioteca digital paneuropea lanzada en 2008, en la que participa el consorcio que está detrás de *VideoActive*. Esta colaboración con una iniciativa que ha de convertirse en central en el entorno de internet contribuye a aumentar la visibilidad del material disponible en el portal.

2.3 Vínculo con el mundo académico

La selección del contenido se ha hecho en gran medida pensan-

do en posibles aplicaciones para el mundo académico, como se ha dicho. Particularmente para el campo de los estudios de comunicación, por las facilidades que supone para los análisis comparativos textuales y estéticos, así como por el material que se aporta sobre la historia de la televisión en el continente europeo. No es casual que *VideoActive* colabore con la *European Television History Network (ETHN)*, una red europea de académicos que investiga sobre la historia de la televisión y cuyos miembros han elaborado los textos del portal que contextualizan la historia televisiva europea. El material audiovisual disponible, además, puede resultar de gran utilidad para otros estudios desde la historia, las humanidades o la ciencia política (sobre el proceso de integración europea, por ejemplo).

La aproximación comparada a la historia televisiva no ha sido habitual en los estudios de comunicación, en gran parte por las diferencias idiomáticas y la dificultad para acceder a los materiales de base que harían posible este análisis comparativo. Al contar con *VideoActive* como fuente se abre una primera puerta a la elaboración de nuevas teorías y metodologías de aproximación a la televisión como objeto de estudio. Tal como se expone en la estrategia de selección del contenido del portal, “el tema del valor historiográfico [prioridades y necesidades a la hora de seleccionar los contenidos] es problemático y particularmente complejo ya que inevitablemente cambia en función de las fronteras nacionales, los sistemas radiotelevisivos y las disciplinas académicas”.

http://videoactive.files.wordpress.com/2007/10/23_content_selection_strategy_report.pdf

3. Contribución de Televisió de Catalunya

En este proyecto participa *Televisió de Catalunya (TVC)*, el operador público autonómico catalán, que en los últimos años se ha distinguido por una activa política de incorporación e impulso de novedades tecnológicas para el cumplimiento de su función de servicio público (Prado; Fernández-Quijada, 2006).

“En *VideoActive*, TVC pone al alcance del público parte de su fondo en un doble contexto: europeo y televisivo”

El archivo de *TVC* ha sido tradicionalmente de uso interno y se ha orientado desde sus orígenes a la preservación y reutilización del fondo en la propia producción. Siempre se ha mantenido abierto a la colaboración con investigadores y académicos y ha dispuesto del servicio de venta de copias físicas de programas. Desde 2004 funciona además un servicio de vídeo bajo demanda en internet, *3 a la carta*, que ya proporciona una parte de sus contenidos de manera gratuita, una tarea facilitada por el proceso de digitalización del fondo documental llevado a cabo en los últimos años y por la integración de los procesos digitales en el programa de gestión de contenidos implementado en el centro, *Digiton Suite*.

<http://www.tv3.cat/seccio/3alacarta/>

La participación de *TVC* en *VideoActive* es una manera de poner al alcance del público parte de su fondo desde una doble perspecti-

“Hay cooperación para la preservación del patrimonio audiovisual entre organizaciones de diferentes países, operadores y documentalistas e investigadores”

va: primero, en un contexto europeo, de manera que también sea una forma de proyectar la historia y la cultura catalanas al exterior y, en segundo lugar, en un contexto televisivo, gracias al trabajo específico que desde este operador se ha hecho para asignar metadatos a todos sus materiales. El trabajo se ha vehiculado mediante una documentalista encargada de seleccionar los ítems, añadir los metadatos y crear los clips de vídeo. Otras personas han colaborado en tareas auxiliares como las traducciones, el mantenimiento del tesoro o la transferencia de archivos.

Se han seleccionado piezas sobre las 34 materias fijadas, de forma más o menos equitativa entre las tres décadas de historia de TVC. La modernidad fue un factor diferenciador respecto al resto de los operadores del consorcio *VideoActive*, en su mayor parte antiguos monopolios fundadores de la televisión

–y de la radio– en sus respectivos países.

En la selección también se ha buscado un equilibrio entre géneros televisivos, de manera que se pudieran encontrar piezas informativas pero también documentales, infantiles, ficción, spots publicitarios, etc. Además, al tratarse de un proyecto europeo se han primado los enfoques comparativos que proponían algunos programas históricos como *Camí de l'euro* o *Estampes europees*.

4. Conclusiones

En el proyecto *VideoActive* se manifiesta una triple voluntad de cooperación para la preservación del patrimonio audiovisual: primero, transnacional entre organizaciones de diferentes países; en segundo término, entre operadores y organismos públicos y empresas comerciales, y finalmente, interdisciplinar entre documentalistas e

investigadores del campo de la comunicación.

Resulta también significativa la necesidad de crear un consorcio ad hoc para la puesta en marcha de un proyecto paneuropeo de esta singularidad, dada la existencia de una organización entre cuyas funciones cabría esperar precisamente la de generar estrategias para la conservación del patrimonio de sus asociados. Se trata de la *Unión Europea de Radiodifusión (UER)*, formada básicamente por los operadores públicos europeos –y algunos privados– pero que no ha desplegado ningún proyecto similar a *VideoActive*, a pesar de que la mayoría de los miembros de éste último pertenecen también a la *UER*. No es el caso de *TVC*, que figura como único representante español en el proyecto a pesar de que el mayor archivo televisivo del país es el de *Radio Televisión Española*.

“Las principales dificultades se han localizado en dos aspectos muy concretos: en la titularidad de los derechos de propiedad intelectual y en la selección temática”

Más allá de las lógicas dificultades técnicas que surgen en todo proyecto de esta envergadura, las principales dificultades se han localizado en dos aspectos muy concretos: en la titularidad de los derechos de propiedad intelectual y en la selección temática. Sobre el primer aspecto, se desconocía a quién pertenecían los derechos de determinadas obras dados los años transcurridos desde que se crearon. En otros casos la progresiva externalización de la producción realizada por los operadores televisivos en los últimos años ha generado

Temas	Décadas			Temas	Décadas		
	80	90	00		80	90	00
Agricultura	3	7	19	Hogar	11	12	16
Artes y cultura	10	11	14	Ocio	4	10	6
Catástrofes	4	8	15	Vida urbana	5	10	14
Guerra fría	5	13	11	Industria	12	7	17
Sociedad de consumo	5	7	25	Fiestas nacionales	7	9	4
Programas de TV controvertidos	0	4	7	Religión	5	5	11
Crimen	5	7	13	Revolución sexual	3	10	7
Democratización	4	6	26	Enfermedades y salud	9	9	19
Drogas y alcohol	4	6	8	Tecnología	7	9	20
Educación	9	10	20	Televisión sobre televisión	7	14	21
Emigración e inmigración	3	7	10	Terrorismo	4	6	13
Medio ambiente	5	7	21	Turismo y vacaciones	5	5	17
Integración europea	5	10	13	Transportes	5	4	20
Familia	4	11	26	Guerras y conflictos	2	10	19
Moda	7	10	18	Estado del bienestar	2	10	15
Alimentación	4	5	29	Trabajo	3	3	22
Revolución de género	2	3	28	Cultura juvenil	7	6	12

Tabla 1. Selección temática de TVC

discusiones sobre la propiedad de estos derechos y su explotación. Sobre la selección temática, las propias diferencias sociales y culturales de los países respectivos han dificultado la localización de materias afines o con posibilidades de comparación.

A pesar de estos obstáculos, VideoActive se ha puesto en marcha y ya constituye un activo electrónico muy relevante para docentes e investigadores del campo de la comunicación. Además, desde el consorcio se persigue la continuación del proyecto ampliándolo cuantitativa y cualitativamente, y se plantea la posibilidad de abrirlo a otros operadores y organismos interesados.

Bibliografía

Agirreazaldegui-Berriozabal, Teresa. "Claves y retos de la documentación digital en televisión". *El profesional de la información*, 2007, septiembre-octubre, v. 16, n. 5, pp. 433-442.

Besser, Howard. "Collaboration for electronic preservation". *Library trends*, 2007, verano, v. 56, n. 1, pp. 216-229.

Comisión de las Comunidades Europeas. *Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. i2010 Bibliotecas Digitales*, COM(2005) 465 final. Bruselas: Comisión Europea, 2005.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0465:FIN:ES:PDF>

De-Moragas, Miquel; Prado, Emili. *La televisión pública a l'era digital*. Barcelona: Pòrtic, 2000, ISBN 84-7306-617-0.

Gómez-Alonso, Rafael. "Investigar la historia de la televisión en España: algunos problemas documentales y metodológicos". *Área abierta*, 2004, enero, n. 7.

Hecht, Alexander; O'Dwyer, Andrew; Oomen, Johan; Scharinger, Florian. "Birth: building an interactive research and delivery network for television heritage". En: *Proceedings of Ichim digital culture and heritage*, 2004.
http://www.ichim.org/ichim04/contenu/PDF/2280_Oomen.pdf

Lloyd, Annemare. "Guarding against collective amnesia? Making significance problematic: an exploration of issues". *Library trends*, 2007, verano, v. 56, n. 1, pp. 53-65.

Prado, Emili; Fernández-Quijada, David. "The role of public service broadcasters in the era of convergence. A case study of *Televisió de Catalunya*". *Communications & strategies*, 2006, n. 62, pp. 49-69.
http://www.idate.fr/fic/revue_telech/148/CS62_PRADO_FERNANDEZ.pdf

Teruggi, Daniel. "Can we save our audio-visual heritage?". *Ariadne*, 2004, abril, n. 39.
<http://www.ariadne.ac.uk/issue39/teruggi/>

Wright, Richard. "Digital preservation of audio, video and film". *Vine: The journal of information and knowledge management systems*, 2004, v. 34, n. 2.

David Fernández-Quijada, Departament de Comunicació Audiovisual i de Publicitat, Universitat Au-

tònoma de Barcelona. Edifici I, Facultat de Ciències de la Comunicació. Campus de Bellaterra. 08193 Bellaterra (Barcelona).
david.fernandez@uab.es

Montse Fortino, Departament de Documentació, Televisió de Catalunya. c/ de la TV3, s/n. 08970 Sant Joan Despí (Barcelona).
mfortino.b@tv3.cat



Le ayudamos a crear Bibliotecas Virtuales
desde la digitalización de materiales bibliográficos hasta la asignación de metadatos y su implementación en la red, conforme a la normativa internacional

Productos para crear Bibliotecas Digitales y Virtuales

<p>DIGIBIS 5.0 Solución optimizada para la creación de Bibliotecas Digitales y la Gestión Bibliotecaria Multilingüe</p> <p>DIGIARCH 1.0 Sistema digital de descripción y gestión archivística</p> <p>Digitalización automática Con asignación dinámica de metadatos</p>		<p>OASIS PMH 2.0 Sistema integrado de instalación de diversos esquemas de metadatos:</p> <ul style="list-style-type: none"> OCML sin modificar MARC 21 EAD etc. OAI <p>Implementación de las directrices DRIVER</p>
---	--	--

- **Recolección en la Web para Entidades e Instituciones de Memoria en OAI-PMH y Dublin Core e intercambio de metadatos en METS (diferentes Profiles)**
- **Consultoría y mappings a DCMI para implementación en repositorios OAI v2.0**
- **Tecnologías abiertas para la creación, recuperación y recolección de metadatos (MARCXML, DCMI y RDF)**
- **Repositorios Institucionales para Preservación Digital a largo plazo mediante PREMIS y OAI-ISO 14721**
- **Reconocimiento Óptico de Caracteres OCR y generación dinámica de METS/AIto**
- **Servidor adicional de SRU Search/Retrieval via URL**
- **Agregadores de contenido RSS**

www.digibis.com

Evolución de los servicios de redes sociales en internet

Por Marcos Ros-Martín

Resumen: Se expone una breve evolución histórica de los servicios de redes sociales que han surgido en internet bajo el concepto de Web social, incidiendo en los ejemplos más significativos de estas aplicaciones. Se hace una aproximación al concepto de red social, a los principales ensayos realizados para su análisis y sus implicaciones dentro de los distintos tipos de capital social (vínculos fuertes y débiles), así como al impacto que los distintos servicios web pueden tener en las relaciones de los usuarios que los utilizan.

Palabras clave: Sitios de redes sociales, Redes sociales, Internet, Web 2.0.

Title: Evolution of social network services

Abstract: The paper exposes a brief history of social network services on the Web that have evolved under the concept of Web 2.0, along with the most significant examples of these platforms. The concept of social networking and the main tests used to study them are explained, as well as their implications within different kinds of social capital (weak and strong ties) and the impact that the different web services can have on the relationship among their users.

Keywords: Social network services, Social networks, Internet, Web 2.0.

Ros-Martín, Marcos. "Evolución de los servicios de redes sociales en internet". *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 552-557.

DOI: 10.3145/epi.2009.sep.10



Marcos Ros-Martín es licenciado en documentación por la Univ. Politècnica de València. Trabaja como documentalista en la Unidad Técnica de Información de Aidico-Instituto Tecnológico de la Construcción en Valencia y anteriormente fue documentalista informativo en el diario Levante-El Mercantil Valenciano. Es co-editor del blog personal El documentalista enredado.

LA IDEA INICIAL DE LAS PLATAFORMAS de redes sociales en internet parte de la teoría de los seis grados de separación.

Esta teoría fue expuesta de forma intuitiva en 1929 por el escritor **Frigyes Karinthy** en un relato corto denominado *Cadenas*. Según este autor, el número de conocidos de una persona crece exponencialmente siguiendo un número de enlaces de una cadena que serían las relaciones humanas. De este modo sólo sería necesario un pequeño número de enlaces para conectar a cualquier persona con el resto de la población.

Los intentos de demostrar esta teoría de forma científica han sido numerosos. En la década de los 50 los investigadores del *MIT* y de *IBM*, **Ithiel de Sola Pool** y **Manfred Mochén** respectivamente, trataron de encontrar el número de pasos necesarios para que toda la red humana estuviese interconectada.

En 1967 el sociólogo de la *Harvard University* **Stanley Milgram** se propuso demostrar la teoría mediante un experimento basado en el envío de cartas postales. Teniendo presente que una red social comprende a un conjunto de personas con un patrón de interacciones entre ellas, **Milgram** diseñó un experimento en el que una serie de individuos dispersos geográficamente (primero en Omaha, Nebraska y posteriormente en Wichita, Kansas) enviaron una carta a un compañero suyo en Boston. La condición era que tenían que conocer personalmente a las personas a las que enviaban las postales. Además, uno de los objetivos consistía en que las cartas llegasen lo más pronto posible. Una vez éstas comenzaron a llegar a su destino, **Milgram** trazó su recorrido y la red de contactos, llegando a la conclusión de que, en promedio, eran necesarios seis pasos para interconectar cualquier persona dentro de Estados Unidos.

A pesar de no poder demostrar la teoría de forma matemática, se han enunciado numerosos conceptos a partir de ella como el popular *Número de Erdős*¹ o, más recientemente, en *Six degrees: the science of a connected age* del sociólogo **Duncan Watts**. Este profesor de la *Columbia University* expone el experimento que llevó a cabo en 2001 cuando intentó repetir la comprobación de **Milgram**, aunque en esta ocasión el medio utilizado fue el correo electrónico. **Watts** envió un email en el que pedía a los destinatarios que lo reenviasen a sus contactos para comprobar cuánto tardaba en volver al primer emisor. Finalmente el mensaje fue reenviado a 48.000 personas

“Según Watts, la media de intermediarios entre dos personas cualesquiera a escala mundial es de seis”

de 157 países, lo que llevó al investigador a aventurar que la media de intermediarios entre dos personas cualesquiera a escala mundial era de seis (Raghavan, 2002).

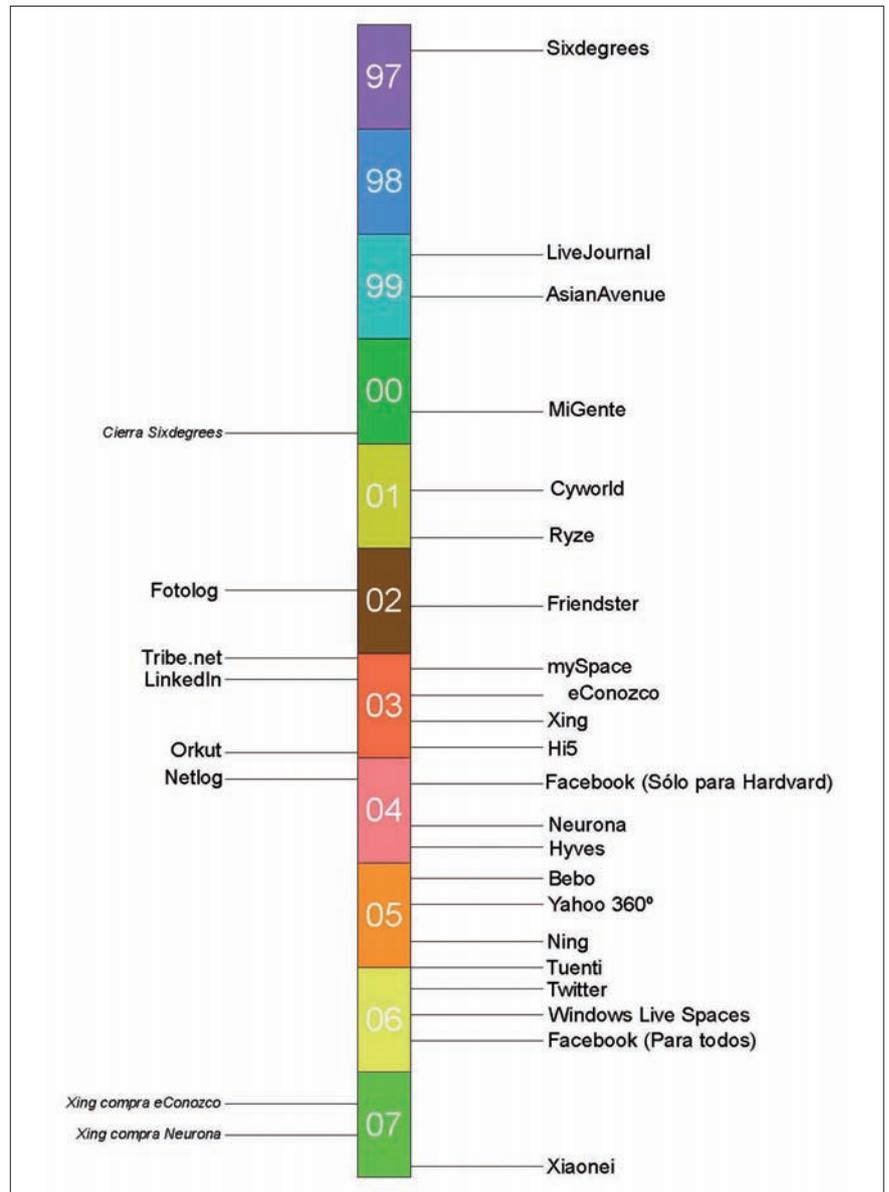
1. El capital social

Los modelos jerárquicos de las redes sociales son asumidos por Watts, Dodds y Newman (2002), que consideran que los individuos pertenecen a grupos que a su vez se encuentran embebidos por grupos jerárquicos más grandes. De esta forma, se define grupo como una colección de individuos que se encuentran unidos por una serie de características sociales asociadas. Por ejemplo un científico pertenecerá a un laboratorio de investigación, que a su vez formará parte de un departamento, que a su vez se encontrará dentro de la estructura de una universidad o de un centro tecnológico.

Obviamente un individuo puede estar situado dentro de varios grupos no necesariamente relacionados entre sí, lo que conduce a la consideración del capital social como el conjunto de recursos que tiene una persona gracias a las relaciones que mantiene con otras. Este capital social puede ser real o virtual (online) dependiendo de los ámbitos en los que la persona se mueva; además permite a un individuo obtener recursos de otros miembros de las redes sociales a las que pertenece. Estos recursos pueden ser de todo tipo como información, relaciones personales o la capacidad de organizar acciones conjuntas con otros (Ellison; Steinfield; Lampe, 2007).

Se distinguen dos tipos de capital social: el constituido por vínculos débiles mediante los que los individuos intercambian información pero sin componente emocional, y el capital social en el que además se intercambian lazos afectivos.

El debate sobre si internet aumenta o disminuye el capital social de sus usuarios ha sido intenso tanto en la sociología como en



Cronología de los servicios de redes sociales en internet más populares (adaptado de Boyd; Ellison, 2007)

los medios de comunicación. Nie (2001) argumentaba que el uso de internet disminuía el contacto cara a cara con los contactos sociales, lo que conllevaba la reducción de su capital social. Este razonamiento ha recibido muchas críticas (Bargh, 2004) e incluso se ha llegado a la conclusión contraria: que las relaciones mediante la Web tienen muchos efectos positivos en las interacciones con la comunidad, desarrollo personal y capital social.

Asimismo recientes investigaciones han remarcado la importancia de la creación de relaciones personales en internet, porque a pesar de tratarse de la creación de capital

social débil, éste puede convertirse en fuerte mediante servicios de distribución de fotografías y otros materiales. Incluso se ha llegado a sugerir que existe la posibilidad de que las redes sociales en internet configuren un nuevo tipo de capital social puesto que aumenta la unión de estas relaciones, y las alarga en el tiempo aunque el contacto físico sea mínimo.

2. Breve historia de las redes sociales en internet

Hay distintas vías para llevar a cabo investigaciones sobre el establecimiento de redes sociales en internet². Por ejemplo, se pueden



Redes sociales

trazar las relaciones sociales dentro de la Web analizando los medios utilizados para la comunicación. Así, disponemos de trabajos sobre análisis de redes mediante la utilización del correo electrónico (Tyler; Tang, 2003; Arenas; Díaz-Guilera 2009) o incluso se han llegado a elaborar productos comerciales basándose en las relaciones que se establecían dentro de la Red en forma de enlaces (Brin; Page, 1998)³. Por otro lado, también se puede ser más específico y se pueden llevar a cabo análisis sobre una plataforma determinada de publicación web para el análisis de redes como pueden ser los blogs (Merelo; Tricas, 2005).

El ámbito de estudio de las redes sociales en internet puede ser muy amplio, por lo que creemos necesario presentar algunas definiciones (ver el recuadro).

Aunque en cada una de las definiciones se resaltan algunas particularidades, como los servicios que ofrecen al usuario (foros o mensajería privada), o la plataforma sobre la que se implementan, todas inciden en que su utilidad fundamental es el establecimiento de relaciones personales y el intercambio de información.

Así podríamos definir un *servicio de red social en internet*⁵ como una plataforma web cuyo fin es la creación de comunidades en línea mediante la representación de las conexiones personales que los usuarios disponen los unos de los otros. En estos servicios se puede compartir información mediante la utilización de servicios agregados de mensajería personal, *microblogging*, publicación de fotografías, formación de grupos de interés, etc.

“Un servicio de red social en internet es una plataforma web cuyo fin es la creación de comunidades en línea”

2.1. Antecedentes (1997-2003)

Aunque puedan parecer un fenómeno reciente, surgieron antes del nacimiento de la Web 2.0. Se puede establecer la fecha de su aparición en 1997, momento en el que se inaugura la web *SixDegrees.com*.

Las principales características de este sitio web eran la creación

de perfiles, de listados de amigos, la posibilidad de envío de mensajes a amigos (funciones que ya permitían otras comunidades virtuales), y a partir de 1998, la navegación en las listas de amigos por parte de terceros (esta función sí que era una novedad). Desgraciadamente la falta de un modelo de negocio viable le abocó a su desaparición, y durante la crisis bursátil tecnológica de principios del siglo XXI tuvo que cerrar.

Entre 1997 y 2001 innumerables aplicaciones comenzaron a articular la posibilidad de crear perfiles y de listar amigos agregados a los mismos. *AsianAve* (1997, anteriormente *AsianAvenue*), *BlackPlanet* (1999) y *MiGente* (2000)⁶ permitían crear perfiles personales, profesionales o de contactos y se podían añadir amigos sin necesidad de aprobación por parte del agregado. Por otro lado, servicios como *CyWorld*⁷ (2001) o *LunarStorm*⁸ (2000) se rediseñaron para empezar a agregar funciones propias de red social como listas de amigos, libros de visita y páginas personales.

Más tarde aparecieron webs como *Ryze.com*⁹ (2001), *Tribe.net*¹⁰ (2003), *LinkedIn*¹¹ (2003) o

Algunas definiciones de red social

- A social network service focuses on building online communities of people who share interests and/or activities, or who are interested in exploring the interests and activities of others. Most social network services are web based and provide a variety of ways for users to interact, such as e-mail and instant messaging services (*Wikipedia*).
- Social network sites (SNSs) such as *Friendster*, *CyWorld*, and *MySpace* allow individuals to present themselves, articulate their social networks, and establish or maintain connections with others. These sites can be oriented towards work-related contexts (e. g., *LinkedIn.com*), romantic relationship initiation (the original goal of *Friendster.com*), connecting those with shared interests such as music or politics (e. g., *MySpace.com*), or the college student population (the original incarnation of *Facebook.com*). Participants may use the sites to interact with people they already know offline or to meet new people (**Ellison**, 2007).
- The internet extended social networks by new ways of communication like email, chat or forums. These techniques make it easier for people to stay in contact and to maintain their relationships. Profiling oneself changed completely. The anonymity of the internet makes it possible for users to decide which information about their personality they want to share with others, for example on their own home page or in chat rooms. Users even can create one or more fake identities and play someone else, which is an attraction for many users (**Ackerman**, 2008).
- We define social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site (**Boyd**, 2007).
- Social networking sites are the latest online communication tool that allows users to create a public or semi-public profile, create and view their own as well as other users' online social networks, and interact with people in their networks (**Subrahmanyam**, 2008).
- Social networking services gather information on users' social contacts, construct a large interconnected social network, and reveal to users how they are connected to others in the network. The premise of these businesses is that individuals might be only a few steps removed from a desirable business or social partner, but not realize it. The services allow their users to get to know one's friends of friends and hence expand their own social circle (**Adamic**, 2005).

Friendster (2002), cuya concepción inicial fue no competir el uno con el otro y, de hecho, en un primer momento se apoyaron.

Sin embargo su desarrollo fue dispar. Por un lado, *Ryze.com* ha seguido una evolución muy positiva gracias a su repercusión en medios impresos norteamericanos. *Tribe.net* creció buscando nichos específicos de negocio, aunque en la actualidad está atravesando dificultades tras pasar a ser de pago. *LinkedIn* ha logrado la mayor popularidad de las tres, constituyéndose en una de las referencias del sector de las redes sociales profesionales a escala global y gracias al lanzamiento de versiones en distintos idiomas.

Finalmente, el caso de *Friendster*¹² es particular puesto que las decisiones adoptadas por sus responsables provocaron su abandono masivo por parte de los usuarios. En un principio fue concebida como un complemento a *Ryze.com* y se contemplaba como competen-

cia de la web de contactos *Match.com*, basándose en la premisa de que la gente que se quiere conocer realmente debía poseer gustos similares, por lo que la incorporación a la red vía "amigo de un amigo" podía ser más viable que aquellos individuos que no tienen ninguna relación ya creada. En un principio *Friendster* captó la atención de un sector muy específico de la población norteamericana y su base de *early adopters* alcanzaría la cifra de 300.000 inscritos.

“Aunque puedan parecer un fenómeno reciente, las redes sociales surgieron antes del nacimiento de la Web 2.0”

Sin embargo, según fue aumentando su popularidad, se encontró con dificultades técnicas y sociales.

Desde el punto de vista técnico, no disponía de un alojamiento lo suficientemente potente para manejar su rápido crecimiento. Por otro lado, su difusión en los medios de comunicación tradicionales desequilibró la comunidad ya creada y los usuarios pronto se encontraron, más allá de sus amistades más cercanas, a otras que no lo eran tanto, y para empeorar las cosas, *Friendster* adoptó una serie de medidas que restringían su uso, lo que destruyó su base de usuarios.

Concretamente, tal y como ya se ha señalado, el diseño inicial no permitía la visión de perfiles que estuviesen a una distancia superior a cuatro grados de separación. Sin embargo, los usuarios comenzaron a ir más allá de sus conocidos agregando a personas que no conocían, para ampliar sus círculos sociales. La inclusión por parte de los administradores de una opción para conocer los perfiles más populares permitió una mayor incidencia en este hecho, mientras que la creación

de perfiles falsos (de personajes populares como actores, cantantes o presentadores) también empeoró la situación. Finalmente, *Friendster* decidió la eliminación de los perfiles falsos, lo que hacía pensar que la empresa no compartía los intereses de los usuarios y condujo a un movimiento contestatario que provocó el paulatino abandono de sus inscritos, que buscaron otras redes sociales.

2.2. Momento actual

Con la explosión de la Web 2.0, los sitios web dedicados a compartir contenidos multimedia comenzaron a implementar funciones de las redes sociales. Dentro de este tipo podrían incluirse servicios como *Flickr* (intercambio de fotos), *Last.FM* (hábitos de escucha de música) o *YouTube* (intercambio de vídeos). Sin embargo, esto no fue un impedimento para el lanzamiento de nuevas redes sociales o el intento de reforzar su presencia en otros ámbitos geográficos.

En cualquier caso, una de las principales particularidades a la hora de lanzar una aplicación de este tipo es su imprevisibilidad para obtener una respuesta del público hacia el que se ha orientado. Por ejemplo *Orkut*¹³, adquirida por *Google* en 2003, estaba dirigida a los ciudadanos estadounidenses, aunque hoy disfruta de una enorme popularidad en Brasil convirtiéndose casi en su red social nacional. Por otro lado *MySpace* (2003) creció gracias al apoyo que recibió por parte de las bandas de música *indie* tras la decepción de los usuarios con el cambio de la política de *Friendster*. El efecto arrastre que estas bandas provocaron tras la migración de fans se vio potenciado gracias a la posibilidad que *MySpace* ofrecía de personalizar las páginas personales.

El despegue de *MySpace*¹⁴ se produjo en 2004 tras la incorporación de adolescentes que a su vez

invitaban al resto de sus amigos a unirse tras la sugerencia de miembros de mayor edad. El hecho de que *MySpace* adaptase su política de uso al detectar y aceptar a menores, supuso un incremento en el número de usuarios y, al mismo tiempo, provocó que los grupos de adolescentes y de los fans de bandas de música *indie* que hasta entonces no se interrelacionaban entre ellos, lo hiciesen mediante sus páginas en *MySpace*. En cualquier caso, no fue hasta que los medios de comunicación comenzaron a publicitar esta red social cuando se produjeron incorporaciones masivas desde todos los sustratos sociales. El éxito de esta red era tal que fue adquirida por la empresa *News Corporation* en julio de 2005.

No debe sorprendernos la adquisición de sitios de redes sociales entre empresas dedicadas a medios de comunicación o tecnológicas. De hecho, el caso español es típico en la ampliación de bases de usuarios de redes sociales mediante la adquisición de otras. La red alemana *Xing*¹⁵ (2006) compró dos de las redes sociales hispanohablantes más populares, *Neurona* (2004) y *eConozco* (2007) durante 2007, buscando de esta forma su desembarco en otras zonas geográficas. Esta fórmula no tiene por qué reproducirse siempre, ya que de hecho la estrategia de la estadounidense *LinkedIn* pasa por la inclusión de una versión en castellano de su sitio web facilitando la interacción entre usuarios angloparlantes e hispanos.

Finalmente, cabe destacar el caso de *Facebook*¹⁶ (2004). Fue diseñada en un primer momento para uso privado específico en la *Harvard University*, y tan sólo se podía acceder a ella si se disponía de un correo electrónico del centro universitario. Posteriormente fue ampliándose hacia otras universidades, empresas y, finalmente, a cualquiera que dispusiese de un correo electrónico (2006). Por supuesto

que esto no supone que los nuevos usuarios puedan acceder a las redes privadas ya establecidas dentro de *Facebook* y, por otro lado, tampoco pueden acceder al perfil de otro usuario hasta que éste no lo admita como amigo u opte dejar su perfil abierto y visible para todos.

2.3. Popularidad y utilidad

Aunque aparentemente pueda parecer que el éxito de una red social se mida por su crecimiento rápido a toda la sociedad en general, lo cierto es que han surgido otros tipos de redes más específicas enfocadas hacia un sector concreto del mercado. En cualquier caso esas plataformas proporcionan la ampliación de relaciones más allá del contacto físico, e instalan (o permiten instalar) en ellas otros tipos de aplicaciones para atraer y retener a los usuarios.

No hay datos fiables sobre su grado de aceptación entre los internautas¹⁷, pero su crecimiento y popularidad ha obligado a muchas compañías y administraciones públicas a restringir su uso dentro de la jornada laboral debido a que la mayoría de redes sociales ofrecen más oportunidades de ocio que aplicaciones productivas para la institución.

Por otro lado, parece que fomentan una tendencia hacia la creación de identidades digitales de los internautas dentro de la Web, así como la creación de comunidades virtuales erigidas no tanto alrededor de un tema, como centradas en individuos.

Hasta ahora las redes sociales han tenido éxito casi exclusivamente como diversión y entretenimiento, pero bien podrían usarse para realizar trabajos colaborativos entre profesionales y entre organizaciones.

Notas

1. El número de *Erdős* es un modo de describir la distancia colaborativa, fundamentalmente en

lo relativo a publicaciones matemáticas entre un autor cualquiera y el prolífico **Paul Erdős**. http://es.wikipedia.org/wiki/N%C3%BAmero_de_Erd%C5%91s

2. A pesar de que se utilicen de forma indistinta, los términos internet y Web (de World Wide Web) tienen significados diferentes. Internet es un conjunto descentralizado de redes de comunicación interconectadas, que utilizan la familia de protocolos TCP/IP, garantizando que las redes físicas heterogéneas que la componen funcionen como una red lógica única, de alcance mundial. La Web es un conjunto de protocolos que hace posible de forma sencilla la consulta remota de archivos de hipertexto.

3. Uno de los productos más interesantes para el análisis de relaciones mediante sitios web es *TouchGraph*, que permite visualizar resultados de buscadores y servicios como *Google*, *Facebook* o *Amazon*.
<http://www.touchgraph.com/navigator.html>

4. *Wikipedia*. Social Network Service.
http://en.wikipedia.org/wiki/Social_network_service

5. Se puede consultar un amplio listado de redes sociales en internet en la *Wikipedia*
http://en.wikipedia.org/wiki/List_of_social_networking_websites

6. Todas ellas pertenecientes a la misma empresa, *Community Connect Inc.*, y cada una enfocada a distintos grupos étnicos estadounidenses:

AsianAvenue
<http://en.wikipedia.org/wiki/AsianAvenue>

BlackPlanet
<http://en.wikipedia.org/wiki/BlackPlanet>

MiGente
<http://en.wikipedia.org/wiki/MiGente>

7. *CyWorld*
<http://en.wikipedia.org/wiki/CyWorld>

8. Sitio web sueco nacido en 1996 bajo la denominación *StajlPlejs*, se rebautizó como *LunarStorm* en 2000.
<http://en.wikipedia.org/wiki/LunarStorm>

9. *Ryze*
<http://en.wikipedia.org/wiki/Ryze>

10. *Tribe.net*
<http://en.wikipedia.org/wiki/Tribe.net>

11. *LinkedIn*
<http://en.wikipedia.org/wiki/LinkedIn>

12. *Friendster*
<http://en.wikipedia.org/wiki/Friendster>

13. *Orkut*
<http://en.wikipedia.org/wiki/Orkut>

14. *MySpace*
<http://en.wikipedia.org/wiki/MySpace>

15. *Xing*
<http://en.wikipedia.org/wiki/XING>

16. *Facebook*
<http://en.wikipedia.org/wiki/Facebook.com>

17. Recientemente *ComScore* ha publicado un análisis sobre los usos de los servicios de redes sociales en Europa.
<http://www.comscore.com/press/release.asp?press=2774>

Bibliografía

Ackermann, Markus et al. "HelloWorld: an open source, distributed and secure social network". En: *Icpdas* 2008.
http://www.w3.org/2008/09/msnws/papers/HelloWorld_paper.pdf

Adamic, Lada; Adar, Eytan. "How to search a social network". *Social networks*, 2005, v. 27, n. 3, pp. 187-203.
<http://www.cond.org/socsearch.pdf>

Arenas, Alex; Díaz-Guilera, Albert. "El correo electrónico y la formación de comunidades en organizaciones". *El profesional de la información*, 2009, enero-febrero, v. 18, n. 1, pp. 27-33.

Bargh, John A.; McKenna, Katelyn. "The internet and social life". *Annual review of psychology*, 2004, v. 55, n. 1, pp. 573-590.

Boyd, Danah M.; Ellison, Nicole. "Social network sites: definition, history, and scholarship". *Journal of computer-mediated communication*, 2007, v. 13, n. 1, article 11.
<http://jcmc.indiana.edu/vol13/issue1/boyd.elli.html>

Brin, Sergey; Page, Lawrence. "The anatomy of a large-scale hypertextual Web search engine". *Computer networks and ISDN systems*, 1998, n. 30, pp. 107-117.
<http://infolab.stanford.edu/pub/papers/google.pdf>

Dwyer, Catherine; Hiltz, Starr Roxanne; Passerini, Katia. "Trust and privacy concern within social networking sites: a comparison of Facebook and MySpace". En: *Proceedings of the Thirtieth Americas conference on information systems*, Keystone, Colorado August 09-12-2007.
<http://csis.pace.edu/~dwyer/research/DwyerAM-CIS2007.pdf>

Ellison, Nicole B.; Steinfield, Charles; Lampe, Cliff. "The benefits of Facebook "friends": social capital and college students' use of online social network sites". *Journal of computer-mediated communication*, 2007, v. 12, n. 4, pp. 1143-1168.
<http://jcmc.indiana.edu/vol12/issue4/ellison.html>

Freire, Juan. "Redes sociales: ¿modelos organizativos o servicios digitales?". *El profesional de la información*, 2008, v. 17, n. 6, noviembre-diciembre, pp. 585-588.
<http://www.elprofesionaldeinformacion.com/contenidos/2008/noviembre/01.html>

Govani, Tabreez; Pashley, Harriet. *Student awareness of the privacy implications when using Facebook*, 2005.
<http://lorrie.cranor.org/courses/fa05/tubzhl.pdf>

Gross, Ralph; Acquisti, Alessandro. "Information revelation and privacy in online social networks (The Facebook case)". En: *Pre-proceedings version. ACM Workshop on privacy in the electronic society (WPES)*, 2005.
<http://www.heinz.cmu.edu/~acquisti/papers/privacy-facebook-gross-acquisti.pdf>

Kautz, Henry; Selman, Bart; Shah, Mehul. "Referral Web: combining social networks and collaborative filtering". *Communications of the ACM*, March 1997, v. 40, n. 3, pp. 63-65.

Merelo, Juan J.; Tricas, Fernando. "Métrica de la blogosfera. Algunas medidas y relaciones en la

blogosfera hispana". *Telos: Cuadernos de comunicación e innovación*, 2005, n. 65, pp. 101-104.

Nie, Norman H. "Sociability, interpersonal relations and the internet: reconciling conflicting findings". *American Behavioural Scientist*, 2001, v. 45, n. 3, pp. 402-435.

O'Reilly, Tim. *What is Web 2.0. Design patterns and business models for the next generation of software*, 30/9/2005.
<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

Pekárek, Martin; Leenes, Ronald. "Privacy and social network sites: follow the money!". En: *W3C Workshop on the future of social networking*. Barcelona, January 15-16, 2009.
<http://www.w3.org/2008/09/msnws/papers/tilt.pdf>

Peset, Fernanda; Ferrer-Sapena, Antonia; Baiget, Tomás. "Evolución social y networking en la comunidad biblio-documental". *El profesional de la información*, 2008, noviembre-diciembre, v. 17, n. 6, pp. 627-635.

Rabkin, Ariel. "Personal knowledge questions for fallback authentication: security questions in the era of Facebook". En: *Symposium on usable privacy and security 2008*, July 23-25, Pittsburgh.
<http://cups.cs.cmu.edu/soups/2008/proceedings/p13Rabkin.pdf>

Raghavan, Prabhakar. "Social networks: from the Web to the enterprise". *IEEE Internet Computing*. January-February 2002, v. 6, n. 1, pp. 91-94.

Scherp, Ansgar et al. "Leveraging Web 2.0 communities in professional organisations". En: *W3C Workshop on the future of social networking*, Barcelona, Spain, 15-16 January, 2009.
<http://www.w3.org/2008/09/msnws/papers/ScherpEtAl-LeveragingWeb2Communities.pdf>

Subrahmanyam, Kaveri; Reich, Stephanie M.; Waechter, Natalia; Espinoza, Guadalupe. "Online and offline social networks: use of social networking sites by emerging adults". *Journal of applied developmental psychology*, 2008, v. 29, n. 6, pp. 420-433.

Tyler, Joshua R.; Tang, John C. "When can I expect an email response? A study of rhythms in email usage". En: *Proceedings of the Eight European Conference on Computer-Supported Cooperative Work*, September 14-18th 2003, Helsinki, Finland.

Watts, Duncan J.; Dodds, Peter Sheridan; Newman, Mark E. J. "Identity and search in social networks". *Science*, 2002, v. 296, n. 5571, pp. 1302-1305.
http://cdg.columbia.edu/uploads/papers/watts2002_identitySearch.pdf

Watts, Duncan J. *Six degrees: the science of a connected age*. New York: W. W. Norton & Company, 2003.

Marcos Ros-Martín. *Instituto Tecnológico de la Construcción (Aidico)*, Paterna (Valencia).
marcos.ros@aidico.es
marcos@documentalistaenredado.net

Espacios (casi) infinitos

Los sistemas de archivo y almacenaje de Eun van más allá de los convencionales del mercado, aportando soluciones no sólo para el ahorro de espacio, sino para la creación del mismo.

Sistemas de vanguardia que multiplican el espacio útil mediante una inteligente aplicación de la tecnología y el diseño de estructuras ergonómicas.

Sistemas que mediante la mejora de la gestión de la documentación suponen un claro ahorro de tiempos y espacios, y por lo tanto de costes.

INTELIGENTE

estanterías eun s.a.

B° Salbatore, 17 - Apdo. 82
20200 Beasain (Gipuzkoa) • Spain
T.: +34 902 114 907
F.: +34 902 114 908
E.: comercial@eun.es



eun madrid s.a.

C/ Del Yunque, 27
Pol. Ind. Santa Ana
28522 Rivas Vaciamadrid (Madrid) • Spain
T.: +34 902 170 599
F.: +34 91 301 13 07
E.: eunmadrid@eun.es

eun méxico s.a. de c.v.

Río Tiber No. Desp. 501
Col. Cuauhtémoc, c.p. 06500
México, D.F.
T.: 5208 8896
F.: 5514 2547
E.: comercial@eun.com.mx



programa de bibliotecas
EL ESPACIO MÁS INTELIGENTE



estanterías móviles
EL ESPACIO EN MOVIMIENTO



estanterías móviles motorizadas
EL ESPACIO EN UN CLICK



estanterías móviles con control electrónico
LA TECNOLOGÍA APLICADA AL ESPACIO



Innovación en Sistemas de Archivo y Almacenaje



Agregación de índices bibliométricos para evaluar la producción científica de los investigadores

Por Francisco Herrera, Enrique Herrera-Viedma, Sergio Alonso y Francisco-Javier Cabrerizo



Francisco Herrera es catedrático del departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada. Sus líneas de investigación abarcan la computación con palabras y toma de decisiones, minería de datos, sistemas difusos, algoritmos evolutivos y bibliometría. Es miembro del comité editorial de numerosas revistas y ha publicado más de 150 artículos científicos. Presenta un índice $h = 38$ según WoS.



Enrique Herrera-Viedma es catedrático de Ciencias de la Computación e Inteligencia Artificial y Vicedecano de Investigación en la facultad de Comunicación y Documentación de la Univ. de Granada. Sus líneas de investigación incluyen la computación con palabras y toma de decisiones, recuperación de información web y bibliometría. Ha publicado más de 70 artículos científicos y presenta un índice $h = 23$ según WoS.



Sergio Alonso es profesor ayudante del departamento de Lenguajes y Sistemas Informáticos de la Universidad de Granada. Ha publicado diversos artículos científicos en revistas y en congresos internacionales y nacionales. Sus líneas de investigación incluyen la toma de decisión en grupo, modelos de consenso, modelado lingüístico difuso, bibliotecas digitales, evaluación de la calidad web y bibliometría.



Francisco-Javier Cabrerizo es profesor ayudante del departamento de Ingeniería del Software y Sistemas Informáticos de la Universidad Nacional de Educación a Distancia. Ha publicado diversos artículos científicos en revistas y en congresos. Sus líneas de investigación incluyen la toma de decisión en grupo, modelos de consenso, bibliotecas digitales, evaluación de la calidad web y bibliometría.

Resumen: Se muestra la importancia de la agregación de índices bibliométricos como método para evaluar la calidad científica de los investigadores, ya que permite crear nuevos índices mucho más significativos que combinan las ventajas de los índices agregados.

Palabras clave: Agregación de índices, Evaluación de investigadores, Índice h , Índice g , Índice hg .

Title: Aggregation of bibliometric indices to evaluate the scientific production of researchers

Abstract: In this paper we show that the aggregation of bibliometric indices is an appropriate way to evaluate scientific production of researchers because it allows us to develop much more precise new bibliometric indices that combine the advantages of each index that is aggregated.

Keywords: Aggregation of indices, Evaluation of researchers, h -index, g -index, hg -index.

Herrera, Francisco; Herrera-Viedma, Enrique; Alonso, Sergio; Cabrerizo, Francisco-Javier. "Agregación de índices bibliométricos para evaluar la producción científica de los investigadores". *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 559-561

DOI: 10.3145/epi.2009.sep.11

Introducción

LA EVALUACIÓN DE LA PRODUCCIÓN CIENTÍFICA de los investigadores es una tarea

fundamental para decidir sobre la aprobación de proyectos de investigación, contratación o promoción de investigadores, etc.

El uso de índices bibliométricos ha atraído una creciente atención, debido a los beneficios de obtener un criterio de evaluación que se re-

suma en un único valor numérico (Grupo SCImago, 2006).

Existen diferentes índices que permiten medir tanto la producción científica de los investigadores como el impacto de la misma. El índice h de Hirsch (2005) ha alcanzado una elevada popularidad y ha sido aplicado a gran variedad de áreas. Su idea es muy simple: un investigador tiene un índice h si h de sus N_p artículos tienen al menos h citas cada uno, y los otros ($N_p - h$) artículos no tienen más de h citas cada uno.

Para intentar superar sus limitaciones, numerosos autores han propuesto novedosos índices basados en éste. Uno de los más interesantes es el índice g (Egghe, 2006): un investigador tiene un índice g si g es el rango más alto tal que sus g trabajos más citados tienen juntos al menos g^2 citas. Esto también significa que sus $(g + 1)$ trabajos más citados tienen menos de $(g + 1)^2$ citas.

Una lista completa de publicaciones relacionadas con el índice h , sus variantes y material adicional sobre este tema se puede encontrar en Egghe (2009), Alonso; Cabrerizo; Herrera-Viedma; Herrera (2009b) y en este sitio web:

<http://sci2s.ugr.es/hindex>

Cada índice identifica aspectos diferentes de la producción científica de un investigador, con lo que la información que proporcionan se complementa muy bien (Borrmann; Mutz; Daniel, 2008). Siguiendo la idea de los metabuscadores de internet, que combinan los resultados de buscadores para enriquecer las respuestas a los usuarios, algunos autores están investigando la combinación de índices bibliométricos mediante operadores de agregación adecuados para crear metaíndices bibliométricos. Estos combinan las ventajas de los índices agregados permitiendo dar una visión más certera de la calidad científica de los investigadores. Al-

gunos ejemplos de metaíndices son: hg -index (Alonso; Cabrerizo; Herrera-Viedma; Herrera, 2009a), q^2 -index (Cabrerizo; Alonso; Herrera-Viedma; Herrera, 2009), y R -index (Jin; Liang; Rousseau; Egghe, 2007).

En este trabajo se muestra cómo la agregación de índices bibliométricos que miden distintos aspectos de la producción científica de un investigador puede proporcionar una evaluación más rica de la misma que si se usa exclusivamente un único índice. Para ello se presenta un ejemplo basado en el metaíndice hg -index.

El metaíndice hg

Permite combinar los diferentes aspectos evaluados por los índices h y g :

- el índice h representa el núcleo de publicaciones (h -core) más citadas del investigador;
- el índice g combina la cantidad de artículos más citados de un autor con la intensidad de citas de los mismos.

El hg -index de un investigador se calcula como la media geométrica de sus índices h y g :

$$hg = \sqrt{h \cdot g}$$

Atendiendo a las propiedades de la media geométrica y de los pro-

prios índices h y g , el índice hg nos proporciona generalmente un valor más cercano a h que a g , incrementa la granularidad en la clasificación, y nos da una perspectiva equilibrada entre el impacto de la mayoría de los mejores trabajos de un investigador y los más altamente citados.

Caso de estudio: bibliometría

En esta sección analizamos el comportamiento del índice hg con respecto a los índices h y g en un caso de estudio en el ámbito de la bibliometría. Para ello empleamos un ejemplo (Egghe, 2006) donde se comparan algunos investigadores del campo de la bibliometría usando los índices h y g y añadimos el índice hg (tabla 1).

En la tabla 1 podemos comprobar que cada uno de los índices mide aspectos diferentes de la producción científica de los investigadores. El índice h se centra en localizar el núcleo de artículos más citados. El índice g permite identificar y valorar los investigadores que, a pesar de tener un núcleo de publicaciones relativamente pequeño, son muy citadas. Por último, el índice hg pondera ambos aspectos de la producción de cada autor obteniendo una valoración más equilibrada de sus publicaciones, y proporcionando mayor granularidad en la clasificación de investigadores que los índices h y g por separado.

	Índice h		Índice g		Índice hg
Garfield	27	Garfield	59	Garfield	39,91
Narin	27	Narin	40	Narin	32,86
Braun	25	Small	39	Braun	30,82
Van Raan	19	Braun	38	Small	26,50
Glänzel	18	Schubert	30	Schubert	23,24
Moed	18	Glänzel	27	Van Raan	22,65
Schubert	18	Martin	27	Glänzel	22,05
Small	18	Moed	27	Moed	22,05
Martin	16	Van Raan	27	Martin	20,78
Egghe	13	Ingwersen	26	Ingwersen	18,38
Ingwersen	13	White	25	White	17,32
Leydesdorff	13	Egghe	19	Egghe	15,72
Rousseau	13	Leydesdorff	19	Leydesdorff	15,72
White	12	Rousseau	15	Rousseau	13,96

Tabla 1. Investigadores del área de bibliometría ordenados según sus índices h , g y hg

“La agregación de índices bibliométricos que miden distintos aspectos de la producción científica de un investigador proporciona una evaluación más rica que un único índice”

Si prestamos atención al índice *h*, comprobamos que **Van Raan** (índice *h* = 19) está por delante de **Small** (18) en la clasificación. Por el contrario, el índice *g* de este último (39) es bastante mayor que el del anterior (27). Sería más lógico que la evaluación de la producción científica de **Small** fuera mayor que la de **Van Raan**, ya que tienen un índice *h* similar, pero el impacto de los trabajos de **Small** es bastante mayor.

Si sólo tenemos en cuenta el índice *g*, observamos que **Small** (39) está por delante de **Braun** (38) en la clasificación. Sin embargo, observamos que el índice *h* de **Braun** (25) es bastante mayor que el de **Small** (18). Sería lógico que la evaluación de la producción científica de **Braun** fuera mayor que la de **Small**, ya que tienen un índice *g* similar, pero el índice *h* de **Braun** es bastante mayor que el de **Small**.

El problema es que los índices *h* y *g* evalúan dimensiones diferentes de la producción científica de los investigadores. Sin embargo, el índice *hg* distingue mejor entre los investigadores ya que tiene en cuenta la información proporcionada por ambos.

La afirmación de que los índices *h* y *g* evalúan dimensiones diferentes de la producción científica de un investigador se puede demostrar por la baja correlación existente entre ellos. Para medirla, como no está claro si los valores de ambos siguen una distribución normal, hemos calculado los coeficientes de correlación de *Spearman* ρ (tabla 2).

ρ	Índice <i>h</i>	Índice <i>g</i>	Índice <i>hg</i>
Índice <i>h</i>	1,000	0,881	0,922
Índice <i>g</i>	–	1,000	0,985
Índice <i>hg</i>	–	–	1,000

Tabla 2. Coeficientes de correlación de *Spearman*

La baja correlación entre los índices *h* y *g* se da porque el índice *h* puede ser bajo y el *g* muy alto debido a que unos pocos artículos hayan recibido muchas citas. Sin embargo, la correlación entre los índices *hg* y *h* y entre los índices *hg* y *g* es bastante alta.

“El índice *hg* es homogéneo y robusto, tiene un buen comportamiento y permite obtener una buena evaluación de la producción científica de los investigadores”

Teniendo en cuenta los resultados obtenidos en el ejemplo estudiado y la alta correlación con los índices *h* y *g*, podemos afirmar que el índice *hg* es homogéneo y robusto, que tiene un buen comportamiento y que permite obtener una buena evaluación de la producción científica de los investigadores. Comprobamos así cómo la agregación de diferentes fuentes de información proporciona mejores resultados que si las utilizamos por separado.

Bibliografía

- Alonso, Sergio; Cabrerizo, Francisco-Javier; Herrera-Viedma, Enrique; Herrera, Francisco. “*hg*-index: a new index to characterize the scientific output of researchers based on the *h*- and *g*-indices”. *Scientometrics*, 2009a, in press. DOI: 10.1007/s11192-009-0047-5.
- Alonso, Sergio; Cabrerizo, Francisco-Javier; Herrera-Viedma, Enrique; Herrera, Francisco. “*h*-index: A review focused in its variants, computation and standardization for different scientific fields”. *Journal of informetrics*, 2009b, in press. v. 3, n. 4, pp. 273-289. DOI: 10.1016/j.joi.2009.04.001.

Bornmann, Lutz; Mutz, Rüdiger; Daniel, Hans-Dieter. “Are there better indices for evaluation purposes than the *h*-index? A comparison of nine different variants of the *h*-index using data from biomedicine”. *Journal of the American Society for Information Science and Technology*, 2008, v. 59, n. 5, pp. 830-837.

Cabrerizo, Francisco-Javier; Alonso, Sergio; Herrera-Viedma, Enrique; Herrera, Francisco. “*q*²-Index: quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core”. *Journal of informetrics*, 2009, in press. DOI: 10.1016/j.joi.2009.06.005.

Egghe, Leo. “Theory and practice of the *g*-index”. *Scientometrics*, 2006, v. 69, n. 1, pp. 131-152.

Egghe, Leo. “The Hirsch-index and related impact measures”. *Annual review of information science and technology*, v. 4, 2009, in press.

Grupo SCImago. “El índice *h* de Hirsch: aportaciones a un debate”. *El profesional de la información*, 2006, v. 15, n. 4, pp. 304-306. <http://www.elprofesionaldeinformacion.com/contenidos/2006/julio/8.pdf>

Grupo SCImago. “El índice *h* de Hirsch: su aplicación a algunos de los científicos españoles más destacados”. *El profesional de la información*, 2007, v. 16, n. 1, pp. 47-49. <http://www.elprofesionaldeinformacion.com/contenidos/2007/enero/05.pdf>

Hirsch, Jorge. “An index to quantify an individual’s research output”. En: *Proceedings of the National Academy of Sciences of the United States of America*, 2005, pp. 16.569-16.572.

Jin, Bihui; Liang, Liming; Rousseau, Ronald; Egghe, Leo. “The *R*- and *AR*- indices: complementing the *h*-index”. *Chinese science bulletin*, 2007, v. 52, n. 6, pp. 855-863.

Rousseau, Ronald. “New developments related to the Hirsch index”. *Science focus*, 2006, v. 1, n. 4, pp. 23-25.

Francisco Herrera, Enrique Herrera-Viedma. Departamento de Ciencias de la Computación e Inteligencia Artificial. ETS de Ingenierías Informática y de Telecomunicación, Universidad de Granada. herrera@decsai.ugr.es viedma@decsai.ugr.es

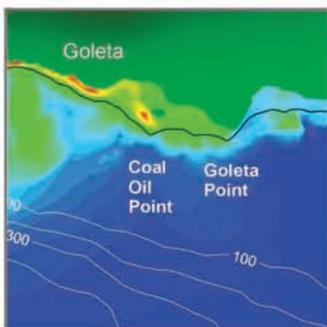
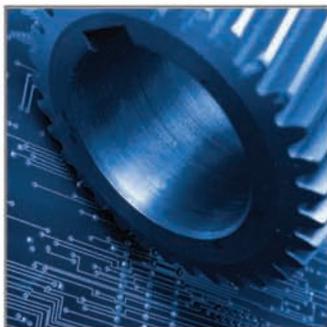
Sergio Alonso. Departamento de Lenguajes y Sistemas Informáticos. ETS de Ingeniería Informática, Universidad de Granada. zerjioi@ugr.es

Francisco-Javier Cabrerizo. Departamento de Ingeniería del Software y Sistemas Informáticos, Universidad Nacional de Educación a Distancia. cabrerizo@issi.uned.es



CSA Illustrata™

Para descubrir información oculta, **empiece aquí...**



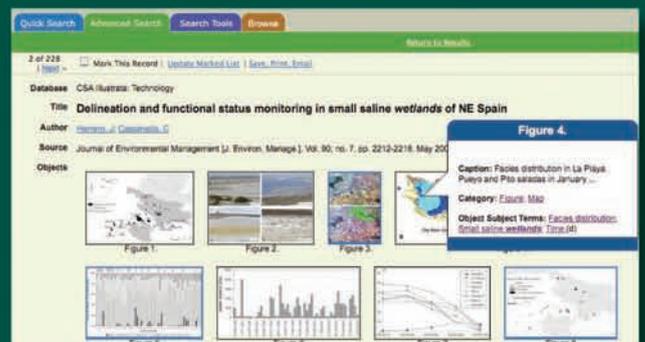
- Indexación completa de las tablas, ilustraciones, imágenes y todo tipo de objetos gráficos.
- Identifica fácilmente artículos difíciles de encontrar.
- Permite al investigador seleccionar los resultados más relevantes gracias a la indexación completa de tablas y gráficos (deep indexing).
- Más precisión, relevancia y eficacia en la recuperación de información.
- Ayuda a la biblioteca a rentabilizar sus inversiones en recursos con texto completo.

CSA Illustrata™: Natural Sciences

Más de 4.700.000 objetos indizados de más de 3.400 revistas revisadas en el área de las Ciencias Naturales.

CSA Illustrata™: Technology

Más de 4.400.000 objetos indizados de más de 2.700 revistas revisadas en el área de la tecnología.



Para más información contacte con la oficina de ProQuest en Madrid:
editor@spain.proquest.com con la referencia **AD 834 09**.

www.proquest.co.uk/go/csaiillustrata

OSTI-DOE, la fuente de información sobre energía. Entrevista a Karen Spence

Por Fernanda Peset y Tomàs Baiget

Resumen: Entrevista a Karen Spence, subdirectora de Sistemas de Información de la Office of Scientific and Technical Information (OSTI), Department of Energy, de los Estados Unidos, que sirve para introducirnos en las importantes bases de datos, repositorios y proyectos de cooperación en información científica y técnica (especialmente sobre energía y medio ambiente) que realiza dicha Oficina.

Palabras clave: Información científica y técnica, Energía, Medio ambiente, Bases de datos, Web, Repositorio, Cooperación, Department of Energy, DOE, Office of Scientific and Technical Information, OSTI.



Karen Spence (centro) entrevistada por Tomàs Baiget y Fernanda Peset

Title: OSTI-DOE, the source of energy information. Interview with Karen Spence

Abstract: An interview with Karen Spence, assistant director of Information Systems, Office of Scientific and Technical Information (OSTI), United States Department of Energy, describes the major OSTI databases, repositories and scientific and technical information co-operation projects (especially on energy and environment).

Keywords: Scientific and technical information, Energy, Environment, Databases, Web, Repository, Co-operation, Department of Energy, DOE, Office of Scientific and Technical Information, OSTI.

Peset, Fernanda; Baiget, Tomàs. "OSTI-DOE, la fuente de información sobre energía. Entrevista a Karen Spence". *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 563-566.

DOI: 10.3145/epi.2009.sep.12

LA OFFICE OF SCIENTIFIC AND TECHNICAL INFORMATION (OSTI), del US Department of Energy, es una de las instituciones clásicas y pioneras de la documentación (en 2007 celebró su 60º aniversario), que por los años 50 y 60 desarrolló teorías y prácticas que sirvieron de modelo a las primeras generaciones de profesionales de bibliotecas especializadas que empezaban a distinguirse de los bibliotecarios, asentando los fundamentos de la Documentación o Information science.

Más tarde, en los 80 y 90 fue famosa por su base de datos sobre

energía, aunque en muchos países no pudimos usarla debido a determinados pactos de colaboración con agencias corresponsales (en España la *Junta de Energía Nuclear*, hoy *Centro de Investigaciones Energé-*

ticas Medioambientales y Tecnológicas, Ciemat), que la consideraron como propia, no abierta al público –cosas típicas de las políticas de otras épocas, que también sucedió durante unos años con la base de



Instalaciones de la OSTI en Oak Ridge, Tennessee

datos *Agris*, de la *FAO*, cuyo representante en España era el *INIA*—.

Hemos aprovechado la visita a España de **Karen Spence**, subdirectora de Sistemas de Información, de la *OSTI*, para entrevistarla y además para ayudar a difundir unas bases de datos importantes y poco conocidas:

¿Qué es y qué hace la OSTI?

– Es una agencia del *Departamento de Energía del Gobierno de los EUA* que siempre ha tenido una gran vocación en el tema de la documentación y de las bases de datos. Es claro que la Energía (y todos sus temas relacionados) es una materia muy importante, pero además tradicionalmente la *OSTI* ha estado bastante por delante de otras agencias. Tenemos un lema famoso: “El saber se contagia: nuestra misión es que todo el mundo lo pille”. Si el intercambio de conocimientos se acelera, el progreso científico también se acelera y las décadas se comprimen a años, los años a meses, y los meses a días...

Hemos visto varias referencias al “bien público” en documentos de la OSTI. ¿Qué importancia dáis a este concepto?

– Como es sabido, un bien público es uno que no disminuye en valor con el uso, y (casi) no tiene coste marginal para los usuarios después de la primera utilización.

Un bien público científico es un caso especial, pues su valor *aumenta* con el uso. Por consiguiente, para aprovechar al máximo el uso de la información generada por la investigación, es imprescindible que sea un bien público.

En la OSTI propugnáis el open access, pero luego muchas veces son los técnicos los que retienen la información.

– Sí, y esto se acepta para períodos de tiempo que permitan a los investigadores explorar suficiente-

mente las consecuencias y posibilidades de sus descubrimientos. Lo mismo que ocurre con las pequeñas empresas, los grupos pequeños de investigación pueden ser más lentos para seguir y evaluar las consecuencias inmediatas de su investigación, por lo que dicha auto-protección es bastante aceptable.

De todas maneras, como dice **Ángel Borrego**, de la *Universidad de Barcelona*, se produce el síndrome Jekyll-Hyde: los científicos se comportan de forma diferente según sean lectores o autores.

¿Planes para el futuro?

– Técnicamente ahora mismo la *OSTI* tiene un peso muy importante entre las demás agencias investigadoras de los EUA sobre todo por nuestra aportación a las plataformas de difusión, pero estamos siguiendo de cerca los pasos de otros, como por ejemplo los *National Institutes of Health (NIH)*, y su mandato a los investigadores que subvenciona para que hagan públicos sus resultados. Nos parece muy buena la sugerencia de que los autores firmen su cesión del copyright en estos términos: “La revista reconoce que el autor retiene el derecho a proporcionar una copia del manuscrito final a los *NIH* cuando la

revista acepte publicar el artículo, o posteriormente, para el archivo público en *PubMed Central* tan pronto como sea posible después de la publicación en la revista”.

¿Se notará la crisis económica?

– Esperamos que no demasiado. El despliegue tecnológico de la *OSTI* para el acceso a la información siempre se ha desarrollado con una utilización lo más juiciosa posible de unos fondos más bien limitados. Su presupuesto anual es de 9 millones US\$, que representa sólo el 0,1% del presupuesto para I+D del *DOE*.

Productos y servicios de información

DOE Information bridge

DOE Information bridge (1991) ofrece libre acceso a 170.000 informes en texto completo, desde 1991, producidos por el *DOE* y sus contratistas, sobre física, química, materiales, biología, ciencias ambientales, energía, ingeniería, informática, ciencias de la información y energías renovables. A la derecha de la captura de pantalla se distingue un apartado en el que se puede opinar y discutir sobre cada informe.

<http://www.osti.gov/bridge>



<http://www.osti.gov/energycitations>

ses de datos (sobre las que se hacen búsquedas federadas o simultáneas), y provee enlaces a 3.000 sociedades profesionales.

OSTI quiere hacer evolucionar EPN de manera que de ser una fuente básica de información pase a ser un servicio de apoyo al investigador en el sentido más amplio, en todas sus necesidades de información profesional.

Por ejemplo podría utilizarse para:

- dar soporte a actividades de información paralelas como la creación, almacenamiento, conservación y archivo de documentos;
- construcción automática de

Energy citations database

Energy Citations Database (1943-) contiene 2,3 millones de referencias bibliográficas de informes, ponencias, artículos, libros, tesis, y patentes de interés para el DOE, y sus agencias predecesoras [la Energy Research & Development Administration (ERDA) y la Atomic Energy Commission (AEC)].

Colaboración entre agencias del Gobierno

Science.gov

Es una coalición voluntaria de 12 agencias federales, aunque entre ellas OSTI ocupa un lugar preeminente.

Ofrece acceso a 40 bases de datos y 1.950 sitios web científicos, con un total de 200 millones de páginas.

Realmente Science.gov es una mina de información muy poco conocida en Europa, por lo que recomendamos a nuestros lectores que la tengan presente cuando tengan que buscar información científico-técnica, especialmente si tiene que estar relacionada con las administraciones públicas.



<http://www.science.gov/>

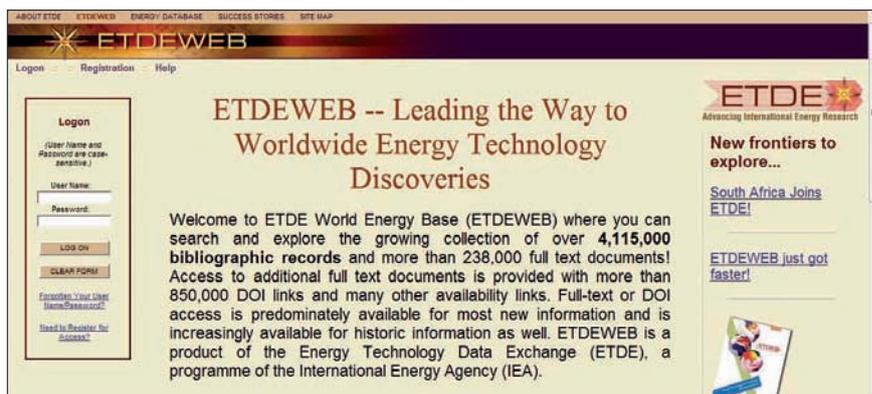
Colaboración internacional E-prints Network

Iniciado en enero de 2000, E-prints Network (EPN) se ha convertido en uno de los principales productos de OSTI.

Proporciona acceso a 26.000 sitios web, con 1 millón de e-prints a texto completo, y 3 millones de documentos adicionales de 52 ba-

estrategias de búsqueda basadas en búsquedas previas que hayan dado buenos resultados;

- uso de software para gestión del conocimiento sobre la colección de documentos;
- localización de expertos;
- búsqueda de puestos de trabajo, becas, subvenciones;
- y así sucesivamente.



<http://www.etde.org/>

ETDE Web

El proyecto *Energy Technology Data Exchange* es parte de la *International Energy Agency (IEA)*, y la *OSTI* constituye su motor gestionándolo.

Con 4,2 millones de referencias (1974-), es la base de datos más grande del mundo sobre energía. “*EtdeWeb* es claramente el sistema de información de referencia sobre energía, medio ambiente y temas relacionados”, proclamaba la *Comisión Europea* en mayo de 2005.

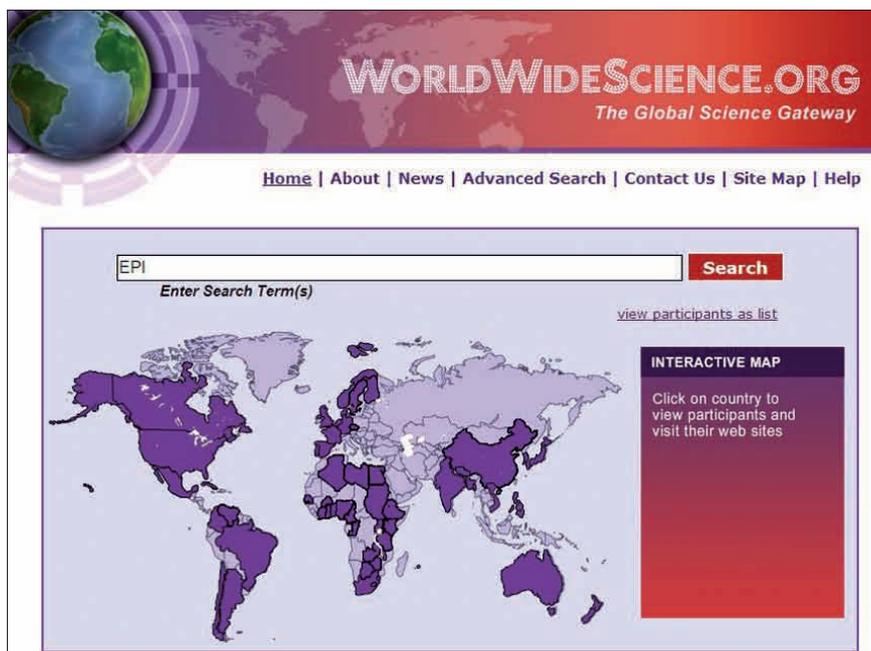
EtdeWeb tiene integrada la antigua base de datos *DOE Energy* (distribuida por *Dialog* con el nº 103).

De 238.000 de las citas se puede obtener directamente el correspondiente texto completo en pdf. Además se incluyen 850.000 DOIs que permiten también localizar el pdf en editoriales de pago.

El acceso está restringido a los países miembros de la *IEA* colaboradores de *ETDE*. España es miembro fundador y los españoles podemos usar este servicio después de rellenar un formulario online, y recibir la aprobación del *Ciemat*. En Latinoamérica los miembros de *ETDE* son Brasil y México, aunque otros países también pueden acceder a la base de datos.

Global Science Gateway

Como se recordará, la declaración de intenciones para crear este



<http://worldwidescience.org/>

portal mundial de la ciencia fue firmada el 21 de enero de 2007 por **Raymond Orbach**, subsecretario de Ciencia del *DOE*, y **Lynne Brindley**, directora ejecutiva de la *British Library*.

La *OSTI* se encarga de su operación.

Actualmente el servicio se llama *World Wide Science*, y con la adición de Corea (junio 2008) y China (octubre 2008), da acceso a 32 repositorios (entre ellos *E-LIS*) de 44 países, que totalizan 200 millones de documentos.

Karen Spence, Assistant Director

for Information Systems, Office of Scientific and Technical Information, US Department of Energy.

Tel.: +1-865 574 02 95; fax: 865-576 93 57

spencek@osti.gov

<http://www.osti.gov>

<http://www.science.gov/workshop/spencek.html>

Fernanda Peset, Univ. Politècnica de Valencia, Depto. de Comunicación Audiovisual, Documentación e Historia del Arte (*Dcadha*), Camino de Vera, s/n. 46022 Valencia. mpesetm@upv.es

Tomàs Baiget, EPI SCP, Apartado 32280, 08080 Barcelona.

baiget@sarenet.es

Suscripciones

Renovar (o comenzar) la suscripción a “El profesional de la información” es ágil y sencillo.

Usted puede gestionar online su suscripción conectándose a esta web:

<http://www.elprofesionaldelainformacion.com/suscripciones.html>

Si lo desea puede comunicar con nosotros dirigiéndose a:

suscripciones@elprofesionaldelainformacion.com

o al teléfono: +34-609 352 954

Manual de indización: teoría y práctica (Isidoro Gil-Leiva)

Por Adolfo Alonso-Arroyo

Alonso-Arroyo, Adolfo. "Manual de indización: teoría y práctica (Isidoro Gil-Leiva)". (Reseña de libro). *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 567-568.

DOI: 10.3145/epi.2009.sep.13



Isidoro Gil-Leiva es diplomado en biblioteconomía y documentación, licenciado en geografía e historia y doctor desde 1997 con una tesis sobre indización automática. Es profesor titular en la Universidad de Murcia. En 1999 publica el libro "La automatización de la indización de documentos" y como editor ha preparado las I Jornadas de Tratamiento y Recuperación de Información (2002). También ha publicado numerosos artículos sobre indización en revistas nacionales e internacionales. Ha impartido conferencias y cursos sobre indización y ha realizado estancias docentes y de investigación en Costa Rica, Uruguay y Brasil. Durante 2008 y 2009 ha visitado diferentes archivos y bibliotecas nacionales en el marco de un proyecto de investigación sobre políticas de indización en Latinoamérica.
<http://webs.um.es/isgil>

SIGUIENDO EN LA LÍNEA DE SU ANTERIOR TRABAJO, La automatización de la indización de documentos (Trea, 1999), el autor ha concebido este manual como un material básico para profesionales de la información, investigadores, profesores y alumnos, así como para aquellas

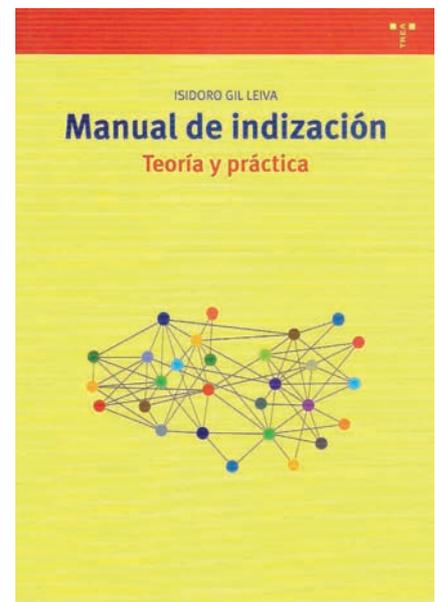
personas que se están iniciando en este campo del saber.

En su redacción se ha utilizado un cuidadoso estilo, con un lenguaje sencillo pero riguroso, bien estructurado, lejos de tecnicismos y complicadas explicaciones teóricas.

El manual está dividido en 6 capítulos bien diferenciados pero a su vez entrelazados entre sí, con múltiples ejemplos intercalados en el texto que aclaran la teoría en cada momento y lo convierten en un manual muy dinámico. Así mismo, con el fin de no desviar al lector del tema central, se introducen alertas gráficas para reforzar y resaltar los aspectos más significativos.

En el primer capítulo titulado "El proceso cognitivo de indización" se explica cómo el indizador necesita de un proceso cognitivo basado en la percepción sensorial (memoria) de la información y en la organización de la comunicación mediante la información para identificar y seleccionar palabras clave, términos de indización o la asigna-

"Cabe resaltar el epígrafe dedicado a las políticas de indización llevadas a cabo por algunas instituciones en sus bases de datos documentales"



Gil-Leiva, Isidoro. *Manual de indización: teoría y práctica.* Gijón: Trea, 2008, 432 p. (Biblioteca y administración cultural; 193).

ISBN 978-84-9704-367-0.

ción de encabezamientos de materia. Se presentan aspectos relativos a la lectura como piedra angular de la indización.

El segundo capítulo se adentra de lleno en el proceso técnico documental y en concreto en la indización. Se comienza integrando la indización dentro de este proceso para continuar diferenciando entre la indización y el índice. Se examinan la exhaustividad, precisión, corrección y consistencia como cua-

lidades que caracterizan el proceso de indización.

La distinción entre el lenguaje natural y el lenguaje controlado se desgrana en el tercer capítulo con el fin de mostrar las herramientas propias de cada uno, tales como listas de palabras clave, listas de descriptores, listas de encabezamientos de materias y tesauros, que a su vez irán acompañadas de múltiples ejemplos que permiten al lector conocer en detalle las características propias de cada una.

El siguiente capítulo completa el anterior, al presentar la indización con lenguaje natural y controlado mediante casos prácticos, utilizando para ello una completa casuística de tipos documentales. Cabe resaltar el epígrafe dedicado a las políticas de indización llevadas a cabo por algunas instituciones en sus bases de datos documentales. Termina el capítulo con un decálogo de la indización (pág. 317) en el que se recopilan los conocimientos teóricos y prácticos mostrados en esta primera

parte, necesarios para adentrarse en el próximo capítulo dedicado a la indización automática.

Analiza la automatización de la indización como una técnica interdisciplinar que necesita del aporte de la lingüística, terminología, informática, lingüística computacional y estadística, y expone las herramientas a utilizar en un sistema de indización automática, que se pueden clasificar en estadísticas, lingüísticas o mixtas. Como colofón a este apartado, el autor da a conocer los principales elementos y las características de diseño de *Sisa*, nombre que recibe el prototipo de sistema de indización semiautomática creado por él mismo.

En el último capítulo, dedicado a la evaluación de la indización, el autor comenta la evaluación intrínseca y extrínseca como dos formas diferentes de valorar la indización. En la intrínseca se toma en consideración el contenido en sí mismo; ésta puede ser cualitativa, mediante el análisis por especialistas, o cuantitativa, basada en factores de medición. En la extrínseca se valora la indización por los efectos que produce comparando sistemas de información.

El manual se complementa con una extensa bibliografía actualizada pero sin olvidar la mención a los autores clásicos. Se distribuye en

capítulos y en cada uno se recoge una parte general y una especializada para los que deseen profundizar en cada uno de los aspectos tratados.

“Es una buena guía para los estudios sobre indización y su avance hacia la indización automática”

Por todo ello, esta obra es una guía muy interesante para los estudios sobre indización y su avance hacia la indización automática. El autor muestra un amplio conocimiento de este tema, lo que unido a su habilidad redactora y a su destreza en fusionar los conocimientos teóricos y prácticos, facilita la lectura y mantiene el interés del lector pese a tratarse de un contenido técnico que en ocasiones pudiera resultar poco atrayente.

Adolfo Alonso-Arroyo. Facultat de Medicina i Odontologia, Departament d’Història de la Ciència i Documentació, Universitat de València. Avda. Blasco Ibáñez 15, 46010 Valencia. adolfo.alonso@uv.es

“Sisa es el nombre que recibe el prototipo de sistema de indización semiautomática creado por el autor”

Próximos temas centrales

Noviembre 2009	Servicios y sistemas de referencia digital
Enero 2010	Sector editorial
Marzo 2010	Publicidad y comunicación empresarial
Mayo 2010	Información bio-médica
Julio 2010	Arquitectura de la información
Septiembre 2010	Cooperación de bibliotecas en red
Noviembre 2010	Medios de comunicación en internet

Los interesados pueden remitir notas, artículos, propuestas, publicidad, comentarios, etc., sobre estos temas a: <http://recyt.fecyt.es/index.php/EPI/index>

Prácticas de las bibliotecas universitarias argentinas (Mercedes Patalano)

Por L.-Fernando Ramos-Simón

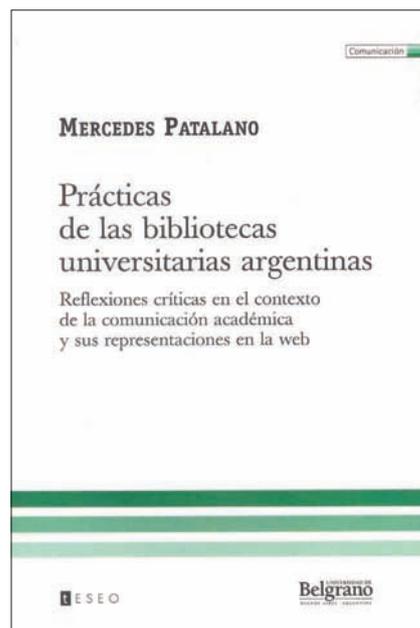
Ramos-Simón, L.-Fernando. "Prácticas de las bibliotecas universitarias argentinas. Reflexiones críticas en el contexto de la comunicación académica y sus representaciones en la Web (Mercedes Patalano)". Reseña de libro. *El profesional de la información*, 2009, septiembre-octubre, v. 18, n. 5, pp. 569-571.

DOI: 10.3145/epi.2009.sep.14



Mercedes Patalano presentando su libro en YouTube
<http://www.youtube.com/watch?v=ZFqrbFLsQ8E>

Mercedes es magíster en comunicación y cultura por la Fac. de Ciencias Sociales; y licenciada en ciencias de la información por la Fac. de Filosofía y Letras, ambas de la Univ. de Buenos Aires. Ha realizado investigaciones y publicado artículos en revistas de América Latina y España sobre el uso de las tecnologías de la información en las bibliotecas universitarias. Es directora de la Biblioteca de la Universidad de Belgrano, donde se desempeña como profesora titular sobre biblioteca digital.



Patalano, Mercedes. *Prácticas de las bibliotecas universitarias argentinas. Reflexiones críticas en el contexto de la comunicación académica y sus representaciones en la Web.* Editorial Teseo, Buenos Aires, 2008, 160 pp.

ISBN: 978 9871354191

ESTA MONOGRAFÍA CUMPLE LOS DOS OBJETIVOS que se propone en el título: por un lado, analizar el uso de las bibliotecas universitarias argentinas y, por otro, reflexionar sobre la comunicación académica en el cambio de siglo, enfoque crítico, éste,

que no se circunscribe sólo al ámbito argentino, ya que muchas de sus conclusiones son de aplicación por ejemplo en España.

Trata asuntos tales como la imagen de la biblioteca, cómo se incluye o excluye el conocimiento científico, o cómo se establece el

diálogo con los usuarios para dar respuesta a las cuestiones claves de "qué se quiere ser, qué se tiene que ofrecer y cómo se lo comunica". El panorama que nos describe no nos resulta demasiado desconocido a este lado del Atlántico.

Así, por ejemplo, la biblioteca sigue sin estar bien integrada en la universidad, el usuario privilegiado es el estudiante y se omite al resto de los integrantes de la comunidad

“El panorama que nos describe no nos resulta demasiado desconocido a este lado del Atlántico”

universitaria; también hay conclusiones más ligadas a las nuevas funciones de la biblioteca: las actividades de investigación propias de la universidad no se representan en toda su magnitud en la web y hay una baja visibilidad de la producción científica local y regional.

Datos de 123 universidades

El libro tiene su origen en una investigación académica realizada en la segunda mitad de 2005, orientada a la obtención de la maestría de la autora, en la que se estudian las prácticas de las bibliotecas a través de una muestra de 123 (86 públicas y 37 privadas) unidades bibliotecarias de las universidades del país, sobre un universo que abarca algo más de cien universidades públicas y privadas que hay en la Argentina. Según los datos que se ofrecen en otra parte del trabajo, el número total de universitarios argentinos supera el millón y medio, de los cuales 1.300.000 estudia en las universidades públicas y algo más de 200.000 en las privadas, es decir, más del 80% estudian en las universidades públicas. Quizás esta dispersión y entrecruzamiento de los datos a lo largo del estudio sea uno de los puntos más débiles y, a la vez, más meritorios del trabajo. Decimos esto porque en lugar de dividir la investigación en dos partes, una para la explotación de los datos y otra para la reflexión sobre los mismos, la autora durante todo el trabajo va disponiendo los datos a medida que discurre el discurso crítico.

En el capítulo central de la investigación se analizan tres aspectos de las bibliotecas: su imagen, la producción científica y la interacción biblioteca/usuarios.

Cómo se ven las bibliotecas a sí mismas

De los 123 sitios web analizados, sólo 45 presentan la misión de la biblioteca. Las 89 restantes se ven a sí mismas con los siguientes atributos: proveedoras de servicios (44%) y apoyo académico (26%); ya con mucha menos intensidad se presentan como depósito de libros (15%), generadora de conocimientos (9%) o dispositivo tecnológico (6%).

“La idea de servicio está presente en casi todas las bibliotecas, pero se presta una atención mínima a estudiarlos”

Participamos de la sorpresa de **Patalano** cuando subraya que si bien la idea de servicio resulta presente en prácticamente todas las bibliotecas y estudios sobre la misión de las bibliotecas, se presta una atención escasísima a estudiar tales servicios. Por citar un dato representativo, “sólo el 0,01% de los trabajos indizados en *Library and Information Science Abstracts (LISA)* se ocupa del aspecto referido al servicio de las bibliotecas”. ¿Cómo interpretar tal contradicción? La autora rebusca en declaraciones de organismos internacionales, en estudios doctrinales y en la observación cotidiana. Aparte de la disparidad y ambigüedad en el concepto de servicio, encuentra que “en la práctica cotidiana, los servicios de bibliotecas están a cargo de profesionales jóvenes que aspiran a tener mayor experiencia, para salir de la práctica del servicio ‘cara a cara’ o ‘mail a mail’ con los usuarios, y poder dedicarse a tareas de mayor prestigio y reconocimiento profesional”. De ser cierta esta apreciación de la autora, las bibliotecas deben reorientar con urgencia su praxis, pues no

cabe mayor contradicción con la misión anunciada.

“Ni las universidades ni las bibliotecas representan en la Web, en toda su magnitud, la producción científica propia”

Visibilidad de la producción científica

El segundo elemento sobre el que se centra este estudio de las prácticas de las bibliotecas universitarias es en la visibilidad de la producción científica mediante el uso de los recursos electrónicos y su difusión a través de la web. En lo primero que se detiene la autora es en el acceso a los recursos, divididos en tres apartados: bases de datos, catálogos de bibliotecas y recursos especiales. En el aspecto de la visibilidad se observa una mejoría respecto a estudios anteriores de la propia autora, aunque hace notar carencias y contradicciones en el sistema de educación superior de su país. Del análisis de las bases de datos presentes en las webs universitarias analizadas se pone de relieve un apreciable déficit, por cuanto más de la mitad no ofrecen accesos a bases de datos. De ahí deriva una importante conclusión: la escasa utilización de la información digital, así como el uso limitado y la escasa capacidad de las tecnologías de la información en Argentina. Estas consideraciones sombrías no mitigan el pesimismo cuando la autora repasa el tipo de bases de datos que son consultadas en las unidades bibliotecarias que tienen acceso a ellas, por cuanto mayoritariamente se decantan por bases de datos internacionales en inglés, muy por encima de las nacionales y regionales, en español. Esta tendencia es interpretada –al

igual que sucede en España— en razón de que se considera como elemento de mayor prestigio y calidad toda la investigación que se publica en revistas especializadas de origen extranjero.

En cuanto a la presencia en la web del catálogo en línea de la biblioteca, la proporción mejora hasta alcanzar el 80% de la muestra, pero es muy bajo el porcentaje (2,4) de bibliotecas que ofrecen el acceso al catálogo de tesis, dato que concuerda con el escaso interés que se da a la producción científica propia, como escaso es también el acceso a los catálogos colectivos nacionales. Finalmente, en el apartado recursos especiales, la autora agrupa un amplio y diverso conjunto de colecciones en línea (p. e. *SciELO*, *Latindex*, *Portal Mincyt*...) alrededor de los cuales profundiza en las carencias de las bibliotecas universitarias en su labor de dar visibilidad a la producción científica de la región latinoamericana, para concluir con la evidencia de que “las universidades argentinas no representan en la web su producción científica en toda su magnitud y similar situación se observa en los sitios web de las bibliotecas universitarias” y más adelante se insiste en que se centran en la ciencia producida fuera de la región.

“La biblioteca está orientada a los estudiantes, con escasa inclusión de profesores e investigadores”

Interacción biblioteca/usuarios

El último apartado de este capítulo central está dedicado a la interacción biblioteca/usuarios, al objeto de analizar el “contrato de lectura”, es decir, la relación que se establece entre un soporte y su lectura.

El análisis de dicha relación lo realiza **Patalano** a través de las webs de las bibliotecas de 12 facultades de la *Universidad de Buenos Aires* que participan en el estudio. Las principales conclusiones que se extraen aluden a la fuerte imagen institucional de la biblioteca, así como a la ausencia de individualización de responsabilidades en la gestión de la misma. En el imaginario construido sobre el usuario de los servicios de la biblioteca aparece el estudiante como destinatario central y principal, con muy escasa presencia de profesores e investigadores. En sentido inverso, esta ausencia es interpretada como que la biblioteca

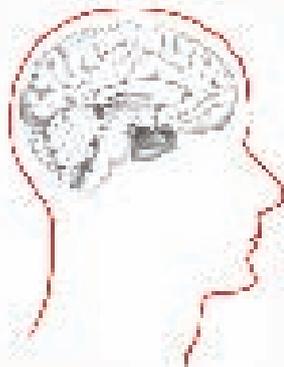
está orientada a los estudiantes con escasa inclusión del resto de la comunidad universitaria.

En un último capítulo de la obra, antes de las conclusiones, se abordan los problemas de las bibliotecas en la última década, centrándose en particular en la comunicación académica y en las tecnologías de la información en donde la autora valora el panorama actual desde su atalaya de experta bibliotecaria y concienzuda estudiosa de este campo de conocimiento que son las Ciencias de la Información y Documentación. Las principales reflexiones que nos deja **M. Patalano** dibujan un panorama marcado por la evolución de las tecnologías, la explosión de la producción de información, la presión de los precios y la extensión del acceso abierto. Y sobre todo ello flota la esperanza de una mejoría en la realidad bibliotecaria argentina que transita “el período más fructífero de toda su historia”.

El libro puede leerse online en:

<http://editorialteseo.com/wordpress/archives/167>

L.-Fernando Ramos-Simón, Facultad de Ciencias de la Documentación, Universidad Complutense de Madrid. ramos@ccdoc.ucm.es



2009 Análisis de tendencias en información y documentación

<http://www.thinkepi.net/anuario.html>

ANUARIO

Think
EPI

Tomás Belget Editor
Javier Guallar Coordinador

5-7 de octubre de 2009

XIV ENCUENTROS INTERNACIONALES SOBRE SISTEMAS DE INFORMACIÓN Y DOCUMENTACIÓN, IBERSID 2009
Zaragoza



Fco. Javier García Marco, Univ. de Zaragoza.
<http://www.ibersid.org>

5-7 de octubre de 2009

7º CONGRESO ARGENTINO DE ARCHIVÍSTICA. Políticas y actualización archivística
Buenos Aires
Federación de Archiveros de la República Argentina (FARA), Archivo General de la Nación
<http://www.mundoarchivistico.com>

6-7 de octubre de 2009

JORNADAS INNOVA.DOC. Gestión-e de documentos
Barcelona
Escola d'Administració Pública de Catalunya, Generalitat de Catalunya
<http://www.innovadoc.cat/cast/index.html>

14-16 de octubre de 2009

XIII JORNADAS DE INFORMACIÓN Y DOCUMENTACIÓN EN CIENCIAS DE LA SALUD
Oviedo
DG de Calidad e Innovación en los Servicios Sanitarios, Consejería de Salud y Servicios Sanitarios, Principado de Asturias.
biblioss@princast.es

15-17 de octubre de 2009

XV JORNADAS BIBLIOTECARIAS DE ANDALUCÍA. Bibliotecas: rompiendo barreras, tejiendo redes
Córdoba



Asociación Andaluza de Bibliotecarios, Ollerías 45-47, 29012 Málaga
aab@aab.es
http://www.aab.es/pdfs/jba15/triptico_15_jba.pdf
<http://www.aab.es>

20-22 de octubre de 2009

11 ILDS CONF - STRATEGIC ALLIANCES AND PARTNERSHIPS IN INTERLENDING AND DOCUMENT SUPPLY
Hannover, Alemania
<http://www.ilds2009.eu/en/home.html>

22-23 de octubre de 2009

INTL SYMPOSIUM ON METHODOLOGY AND RESEARCH METHODS IN LIBRARY AND INFORMATION SCIENCE
Ankara
Ankara University, Faculty of Letters, Dept of Information and Records Management
http://bilgibelge.humanity.ankara.edu.tr/goal_eng.html

23-24 de octubre de 2009

IV CONGRESO NACIONAL DE BIBLIOTECAS MÓVILES
León
Roberto Soto, aclebim@yahoo.es
<http://www.bibliobuses.com>

27-29 de octubre de 2009

XV REUNIÓN INTERAMERICANA DE BIBLIOTECARIOS Y DOCUMENTALISTAS Y ESPECIALISTAS EN INFORMACIÓN AGRÍCOLA (RIBDA 2009). In-



novación tecnológica en el acceso abierto a la información agrícola
Lima
Biblioteca Nacional del Perú; Asociación Interamericana de Bibliotecarios y Documentalistas Agrícolas (AIBDA)
<http://www.aibdaperu.org/>

28-30 de octubre de 2009

II CONF INTL BRECHA DIGITAL E INCLUSIÓN SOCIAL
Madrid
Inst Agustín Millares de Documentación y Gestión de la Información, Univ Carlos III de Madrid
millares-gestion@UC3M.ES
<http://www.uc3m.es/portal/page/portal/>

29-30 de octubre de 2009

2º INTL UDC CONF CLASSIFICATION AT A CROSSROADS: multiple directions to usability
La Haya
Koninklijke Bibliotheek, UDC Consortium
<http://www.udcc.org/seminar2009/index.htm>

4-6 de noviembre de 2009

IV ENCUENTRO DE CATALOGACIÓN Y METADATOS
México DF
Centro Universitario de Invest. Bibliotecológicas (CUIB) e Inst de Invest. Bibliográficas, Univ Nacional Autónoma de México (UNAM); Biblioteca Nacional de México.
Juan Manuel Robles-Correa
jmrobles@cuib.unam.mx
http://cuib.unam.mx/Conv_IV_ECyM_09.pdf

4-6 de noviembre de 2009

INTL CONF ON INTELLIGENT NETWORKING AND COLLABORATIVE SYSTEMS (INCoS 2009)
Barcelona
Univ Oberta de Catalunya

<http://www.uoc.edu/symposia/incos2009/presentation.html>

11-13 de noviembre de 2009

CONGRESO INTL DE CIBERPERIODISMO Y WEB 2.0

Bilbao



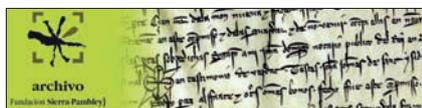
Koldo Meso. Univ del País Vasco

<http://ciberpebi.wordpress.com/>

12-13 de noviembre de 2009

ARCHIVANDO. Las nuevas tecnologías en los archivos

León



Emilia Lareo-Sola, Fundación Sierra-Pambley

archivo@sierrapambley.org

<http://www.sierrapambley.org>

18-20 de noviembre de 2009

IV ENCUENTRO IBÉRICO EDIBCIC 2009

Coimbra

edibcic2009@eventos-iuc.com

<http://eventos-iuc.com/ocs/>

18-21 de noviembre de 2009

II CONF INTL BIBLIOTECAS PARA LA VIDA. Bibliotecas y lectura.

Évora, Portugal

Centro Interdisciplinar de Historia, Culturas y Sociedades, Univ de Évora; Biblioteca Pública de Évora
<http://www.evora.net/bpe/inicial5.htm>

19-20 de noviembre de 2009

XI JORNADAS DE GESTIÓN DE LA INFORMACIÓN. Servicios polivalentes, confluencia entre profesionales de archivo, biblioteca y documentación

Madrid

Asociación Española de Documentación e Información (Sedic)
http://www.sedic.es/XIjornadas_gestinfo-convocatoria.asp

26-27 de noviembre de 2009

4TH INTL LIS-EPI MEETING. Prospec-tiva e innovación en información

Valencia

Revista El Profesional de la Información, Centro Intl. de Investigación en Estrategia y Prospectiva de la Información (Ciepi), Universitat Oberta de Catalunya (UOC), Universidad Politécnica de Valencia (UPV).

http://www.ciepi.org/lisepi/presen_09.htm



26-28 de noviembre de 2009

15TH INTL SCIENTIFIC CONF - KNOWLEDGE BASED ORGANIZATION

Sibiu, Rumanía

Land Forces Academy

<http://conference.armyacademy.ro/>

1-3 de diciembre de 2009

INTL ONLINE INFORMATION CONF & INFORMATION MANAGEMENT SOLUTIONS SPANISH MEETING POINT

Londres



Lorna Candy, *Incisive Media, Haymarket House, 28-29 Haymarket, London SW1Y 4RX*

Tel.: +44-20 7484 9700; fax: 20 7930 2238

<http://www.online-information.co.uk>

<http://www.ims-show.co.uk>

2-4 de diciembre de 2009

5TH INTL DIGITAL CURATION CONF. Moving to multi-scale science: managing complexity and diversity

Londres

Digital Curation Centre (DCC); Coalition for Networked Information (CNI)

<http://www.dcc.ac.uk/events/dcc-2009/>

15-16 de diciembre de 2009

CULTURAL HERITAGE ONLI-

NE. Empowering users: an active role for user communities

Florenca

Fondazione Rinascimento Digitale
<http://www.rinascimento-digitale.it/conference2009.phtml>

23-26 de febrero de 2010

11TH INTL ISKO CONF PARADIGMS AND CONCEPTUAL SYSTEMS IN KNOWLEDGE ORGANIZATION

Roma

<http://www.iskoi.org/rome2010/>

3-5 de marzo de 2010

3RD EUROPEAN NETWORK OF INDICATORS CONF

Paris

Benedetto Lepori, *Serv Ricerca USI-SUPSI; Centre for Organisational Research, Université della Svizzera italiana*

Tel.: +41-586 664 614

benedetto.lepori@usi.ch

<http://www.enid-europe.org/conference.html>

7-9 de abril de 2010

10^º CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS. "Políticas da informação na sociedade em rede"

Guimarães (Portugal)

Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas (BAD)

<http://www.apbad.pt/>

19-23 de abril de 2010

INFO 2010. La sociedad de la información y sus aspectos críticos

La Habana
Instituto de Información Científica y Tecnológica (Idict), Mº de Ciencia, Tecnología y Medio Ambiente, Cuba

<http://www.congreso-info.cu>

26-28 de abril de 2010

EMERGING TECHNOLOGIES IN ACADEMIC LIBRARIES

Trondheim, Norway

Norwegian University of Science

and technology Library (UBiT).
<http://www.emtacl.com>

Mayo-junio de 2010
LIBRARIES IN THE DIGITAL
AGE (LIDA)
Dubrovnik y Zadar, Croacia
http://www.ffos.hr/lida/2010_call/

14-19 de agosto de 2010
76TH IFLA WORLD LIBRARY
AND INFORMATION CON-
GRESS. Libraries: Engaging, Em-
bracing, Empowering
Brisbane, Australia
[http://www.ifla.org/IV/ifla76/index.
htm](http://www.ifla.org/IV/ifla76/index.htm)

28 de septiembre-2 de octubre de 2010
INFORMATION SEEKING IN
CONTEXT (ISIC) 2010
Murcia
Universidad de Murcia
[http://www.um.es/isic2010/index.
php](http://www.um.es/isic2010/index.php)



4th international lis-epi meeting

La información en 2015: innovación y prospectiva
26-27 de noviembre, Universidad Politécnica de Valencia

Innovación en información

Al igual que en ediciones anteriores, este encuentro reunirá a profesionales de varios países con el fin de compartir experiencias y mostrar los resultados del ejercicio profesional y de la investigación más reciente en las áreas de la Gestión de la información, la Innovación y la Comunicación científica.

Ante la gran oferta de congresos existente, se ha diseñado un programa claramente diferenciado de la mayoría, dirigido a un público objetivo minoritario pero seguramente más participativo. *LIS-EPI Meeting* no pretende ser un congreso masivo en el que se intenta atraer a todos con un programa imposible, sino un lugar de encuentro, reflexión y sobre todo discusión participativa para tratar unos pocos temas clave de la máxima actualidad, junto a un laboratorio práctico de e-books y PDAs.

<http://lisepi.ciepi.org>

Organizado por:



UNIVERSIDAD
POLITECNICA
DE VALENCIA



Universitat Oberta
de Catalunya



E-prints in Library
and Information Science
eprints.rclis.org



Patrocinado por:



ARCHIVOS RETROSPECTIVOS DE SCIENCE DIRECT

Investigación de ayer para científicos de hoy



PARA PODER AVANZAR, A VECES HAY QUE MIRAR ATRÁS

Imagine que tiene a su disposición, con un simple clic desde su escritorio, la investigación científica pionera del siglo pasado. Artículos que obtuvieron el premio Nobel, descubrimientos extraordinarios y teorías clásicas en formato electrónico – con todas las prestaciones de búsqueda y enlazados con las más importantes revistas del momento – 25.000 kms de artículos con una altura de 30 rascacielos y un peso de 4.000 toneladas al alcance de su mano.

Todo esto es lo que le ofrecen los archivos retrospectivos de ScienceDirect.

© Cas Oorthuys/Nederlands Fotoarchief

sd_bigenad0709

Para más información puede dirigirse a
www.info.sciencedirect.com/backfiles

 **ScienceDirect**[™]
makes sense.



El profesional de la información

BOLETÍN DE SUSCRIPCIÓN

Deseo recibir todos los números de la revista EPI a partir del mes de enero del año

Suscripción: Institucional Personal

Nombre: Institución:

(Los suscriptores individuales no han de escribir ningún nombre de institución, sólo indicar la dirección particular)

Departamento: NIF institucional:

Dirección:

Código postal: Ciudad: País:

Teléfono: Fax: Correo-e:

Método de pago:

Tarjeta de crédito: VISA Master Card American Express

Titular de la tarjeta:

Número de tarjeta:

Caducidad (mm/aaaa):

Cheque nominativo en euros a nombre de El profesional de la información

Transferencia bancaria a la cuenta de La Caixa 2100 0818 93 0200745544

Enviar, fotocopiado o escaneado, el resguardo de la transferencia.

Las transferencias desde fuera de España deben hacerse a:

IBAN ES95 2100 0818 9302 0074 5544

BIC/Código Swift CAIXESBBXXX

Giro postal al apartado de correos 32.280 de Barcelona

Enviar, fotocopiado o escaneado, el resguardo del giro.

Domiciliación en cuenta bancaria

Entidad: Oficina: DC: Núm:

Titular de la cuenta:

Precios 2009

Suscripción normal:
175,96 € + 4% IVA
= 183 €

Suscripción individual:
(envío sólo a domicilios particulares)
81,73 € + 4% IVA
= 85 €

Gastos de envío fuera de España:
Europa: 40 €
Américas: 60 €

Suscripción sólo online:
90 € + 4% IVA
= 93,6 €

Número suelto:
29,8 € + 4% IVA
= 31 €

Gastos de envío fuera de España:
Europa: 9 €
Américas: 15 €

Las suscripciones van por años naturales, de enero a diciembre

Enviar el boletín relleno, por correo postal o electrónico, a:

Apartado 32.280
08080 Barcelona
España

suscripciones@
elprofesionaldeinformacion.com

Teléfono de atención al suscriptor:

+34 609 352 954

Boletín para comenzar o renovar online la suscripción: <http://www.elprofesionaldeinformacion.com/suscripciones.html>

Todos los profesionales que lo deseen pueden remitir a la redacción de la revista **El profesional de la información** sus colaboraciones en forma de:

- Notas breves (experiencias, estudio de casos, análisis de productos, reseñas, etc.)

- Trabajos más amplios sobre temas de fondo para la sección "Artículos".

El texto ha de enviarse en Word o rtf. Las tablas y figuras generadas en Office deberán ir pegadas en el mismo Word. Los otros materiales gráficos (fotografías, capturas de pantalla, etc.) deben enviarse en ficheros independientes jpg con una resolución de 300 pp (unos 200 KB cada ilustración).

El tamaño ideal de un estudio para la sección "Artículos" es de 3.500 palabras. En casos excepcionales pueden publicarse artículos de mayor extensión. Los trabajos de esta sección son aprobados según el sistema tradicional "peer review": al menos dos expertos en el tema, del Consejo Asesor de la revista y/o externos, deben dar el visto bueno antes de su publicación.

Los textos deben enviarse sin formatos especiales (títulos, secciones, subsecciones, pies de página, sangrías, tabulaciones, colores, etc.).

Los trabajos para la sección "Artículos" deben incluir: a) título en castellano, b) resumen en castellano de 100-150 palabras, c) 5-10 palabras clave en castellano, d) título en inglés, e) resumen en inglés de 100-150 palabras, f) 5-10 palabras clave en inglés, g) texto completo en castellano y h) nombre de los autores, lugar de trabajo y dirección de correo electrónico. **Han de ser inéditos.**

Se valorará especialmente que los trabajos sean concisos y precisos. Se ruega a los autores que eviten una excesiva retórica.

Las citas bibliográficas en el texto se realizarán de la forma: (Apellido, año). Las referencias bibliográficas, que se limitarán a las obras citadas en el texto, han de prepararse de acuerdo con el siguiente esquema:

Información para los autores

Artículos de una publicación periódica:

Apellido, Nombre; Apellido2, Nombre2. "Título del artículo". Título de la publicación periódica, año, mes, v. [número del volumen], n. [número del ejemplar], pp. [página comienzo]-[página final].

Ponencia presentada en un congreso:

Apellido, Nombre; Apellido2, Nombre2. "Título de ponencia". En: nombre del congreso, año, pp. [página comienzo]-[página final].

Monografías e informes:

Apellido, Nombre; Apellido2, Nombre2. "Título del trabajo". Lugar de publicación: editor, fecha, ISBN [número].

Capítulo de una monografía:

Apellido, Nombre; Apellido2, Nombre2. "Título del capítulo". En: Apellido, Nombre; Apellido2, Nombre2. "Título de la monografía". Lugar de publicación: editor, fecha, pp. [página comienzo]-[página final]. ISBN [número].

Recurso en línea:

Apellido, Nombre; Apellido2, Nombre2. "Título del recurso". Dirección url iniciada en nueva línea.

Todas las contribuciones se tienen que enviar a la sección EPI de la plataforma OJS del Repositorio Español de Ciencia y Tecnología (Recyt) de la Fecyt:

<http://recyt.fecyt.es/index.php/EPI/index>

Previamente los autores deben registrarse en:

<http://recyt.fecyt.es/index.php/EPI/user/registerJournal>